Big data meets artificial intelligence

Challenges and implications for the supervision and regulation of financial services
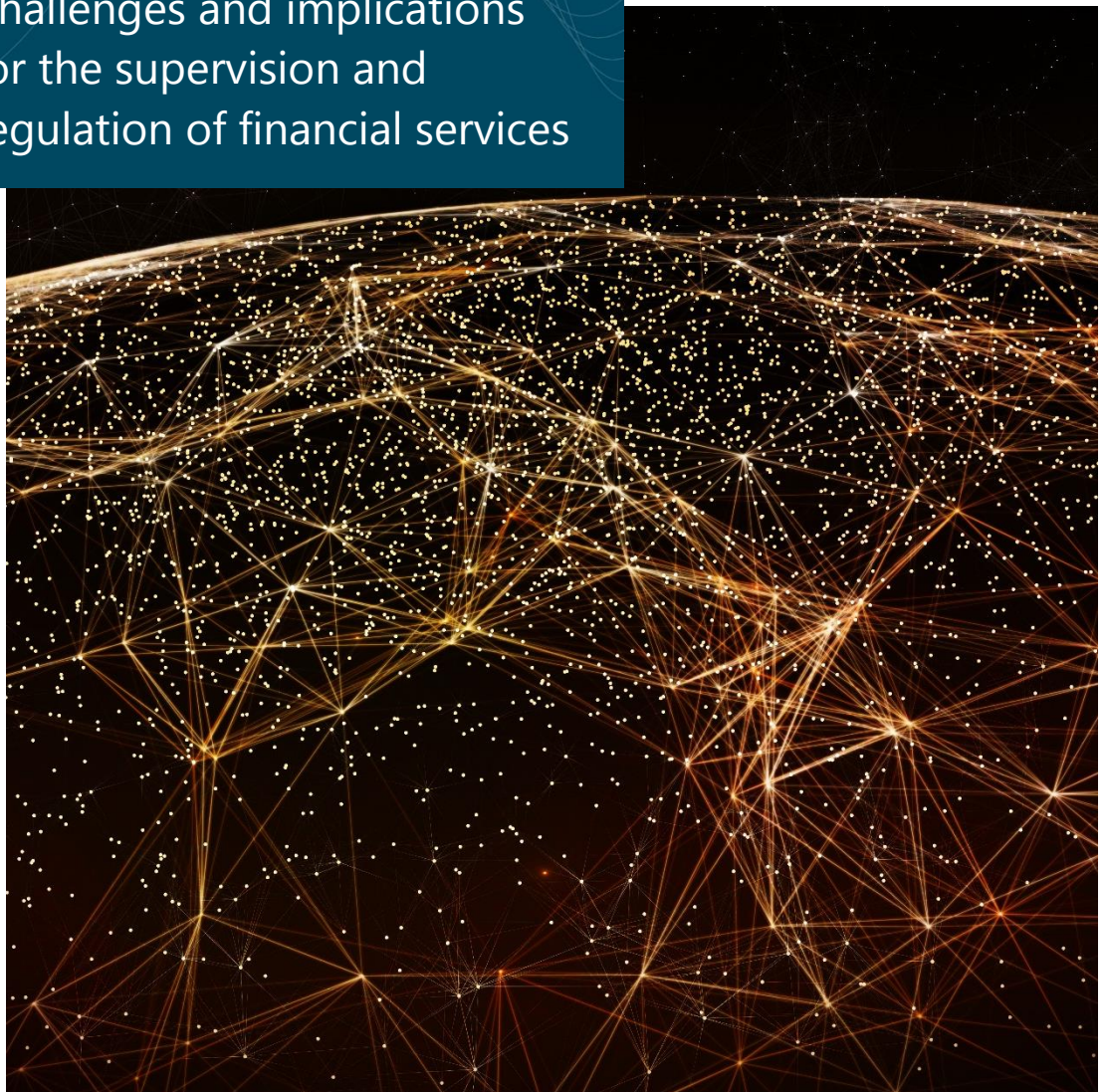
BaFin
Bundesanstalt für
Finanzdienstleistungsaufsicht

Big data meets artificial intelligence

Challenges and implications for the supervision and regulation of financial services

Disclaimer

Important:
The English translation of this report is not binding. The German version alone is authoritative in all respects.

The following report is a speculative study conducted by the Federal Financial Supervisory Authority (BaFin). It addresses the technological developments of big data analytics and artificial intelligence; elaborates scenarios outlining the potential impact these technologies might have on the financial market and on market players; and highlights the potential implications for financial supervisory authorities. BaFin commissioned this report, working closely with external partners from the scientific and the consulting community. The aim of this study is to initiate an open discussion on the impact of digitalization on established market and supervisory structures. Under no circumstances should this report be interpreted as prescript of using any particular technology or business model described therein.

The legal assessments contained in the report are not intended as a description of the administrative practices within BaFin.

Please note that although the utmost care has been taken in compiling the study, BaFin does not guarantee that the contents of the study are complete and correct. The foregoing shall not affect BaFin's liability for damage arising from injury to life, limb or health based on a negligent breach of duty on the part of BaFin or any wilful or grossly negligent breach of duty on the part of its employees or persons responsible for this study.

Foreword

Society and the business world – the financial sector in particular – are currently undergoing profound technological change. Digitalisation has reached new heights: Digital networking is increasingly prevalent and new technologies, such as the Internet of Things, are helping tackle ever more complex tasks. We have at our disposal huge, ever growing quantities of data –

Felix Hufeld

President of BaFin

think big data (BD) – that can be used ever faster and ever more efficiently – think artificial intelligence (AI) and self-learning machines. The following report describes the interaction between big data and artificial intelligence; how fundamentally the BDAI phenomenon can change the financial system; and what implications this has for supervisory and regulatory bodies. It becomes clear that BDAI is not only suited to optimizing existing structures, indeed it paves the way for completely new applications, products, services and business models – with all the inherent opportunities and risks.

To assess this topic as well as possible and from a variety of different angles, BaFin enlisted the help of external experts: Partnerschaft Deutschland, The Boston Consulting Group and the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS).

The report uses market analyses and use cases to outline potential developments as seen from the perspective of banks, insurance companies and the capital markets. However, it also makes a point of considering the consumer angle because it is the consumers who provide the data. It must be clear to everyone involved that BDAI brings with it risks as well as opportunities. These risks must be understood and actively addressed.

This also and indeed particularly applies to the supervisory and regulatory authorities and our report focuses on this aspect in the closing chapter. Faced with rapid advances in digitalization, supervisory bodies need to continually ask whether their supervisory practice is keeping apace of technological progress. The same applies to their tools and instruments of regulation. Supervisory and regulatory bodies must be technology-neutral, must embrace the principle of "same business, same risks, same rules" and all the while remain on the cutting edge – increasingly a challenge these days. This is reflected in the supervisory and regulatory key questions identified in the report. Much work is still needed to adequately answer these questions.

For such a fast-paced topic as BDAI, it is impossible to make any conclusive statements – not even in an extensive report such as this. Nevertheless, we believe that intensive discussion on BDAI is imperative and we hope to make a substantial contribution with this study. We look forward to many more discussions with representatives from the financial industry, the scientific community and the international financial regulatory authorities.

Contents

Picture credits: Cover design: iStock/spainter_vfx, Executive Board photos by Bernd Roselieb Photography

I. Summary

1.1 Introduction

Big data and artificial intelligence have triggered profound changes | Society and the business world are currently undergoing profound technological changes. Digital networking is increasingly prevalent and new technologies are helping tackle ever more complex tasks. This trend is driven in particular by the availability of large quantities of data – big data (BD) – and by the improved opportunities for using this data – artificial intelligence (AI)[1]. This report examines how these phenomena work together within the financial system and refers to them collectively as "BDAI". Applying BDAI mostly means using methods of machine learning, where algorithms give computers the ability to learn from existing data and then apply what they have learned to new data.

Successful BDAI applications are self-reinforcing and can spread rapidly | The relevance of BDAI is growing as technology, companies and consumers interact. First, current technological progress facilitates the extensive and practical use of BDAI. Second, companies are increasingly relying on data and the value they extract from it to optimise their business models and processes. Third, consumer behaviour is increasingly shaped by digital applications, which in turn boosts the generation and availability of data. The last two points in particular can have a strong, self-enhancing effect on one another. In many parts of the financial services industry, this self-reinforcing development associated with the use of BDAI is still in its early stages; however, the assumption is that the pace of this development will increase rapidly.

BDAI facilitates innovation | Combining analytics and the mass of data available helps generate new insights, which can then be used within the financial system to facilitate product and process innovations. Innovations such as these can have a disruptive impact on existing value creation processes. As a consequence, new providers can enter the market and established business processes and market structures can change.

Innovative developments require supervisory/regulatory attention at an early stage | This report addresses the potentially profound impact that BDAI technology can have on the financial services industry. We have drawn on market analyses and use cases to outline both potential developments, as seen from the perspective of banks, insurance companies and the capital markets, and the impact on consumers. Experts from the fields of science and research were involved in preparing this report.[2] This comprehensive view lays the foundation for deriving both supervisory and regulatory implications and key issues.

---

[1] Put simply, artificial intelligence can be understood as a combination of mass data, sufficient computing resources and machine learning. Methods to reproduce general human intelligence – "strong AI" – are not expected in the foreseeable future.

[2] Support for the technological assessments in this study was mainly provided by the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS).

1.2 Impact on the financial system

In examining how BDAI is used in the financial services sector, we distinguish between three main groups of providers:

- traditional companies (incumbents), in particular supervised firms, such as banks and insurance companies

- comparatively young, technology-oriented providers with specific offerings, some of which are directly supervised (fintechs/insurtechs/regtechs/legaltechs)

- and large, global technology companies (bigtechs), most of which have not been subject to supervision to date

In terms of BDAI usage, capital markets are ahead of banks and insurance companies | Banks and insurance companies are only just beginning to use BDAI. Greater use of BDAI and any knock-on effects this may have are to be expected in specific areas initially – at the customer interface, in core processes and in new business models. In the capital market, on the other hand, BDAI is already used much more frequently and extensively, as large data sets and algorithms have been used there for many years.

BDAI can foster a "winner-takes-all" market structure | The more data available to a company, the more insights it can generate. These insights can then help the company to develop more innovative products, from which it can, in turn, obtain additional data. In many cases, the data flow required for this is generated based on the principle that consumers pay with their data – a principle that can be found in the platform-based business models of many bigtechs. This self-supporting market penetration process could favour the emergence of monopoly-type BDAI providers ("winner takes all"). However, these types of offerings usually only appear to be free to the customer because, ultimately, they are financed by the evaluation of the data that the customer gives to the provider, e.g. via personalised advertising ("there is no such thing as a free lunch").

Unregulated BDAI providers could become systemically important for the financial market | Dominant BDAI providers can take data that they collect from outside the financial sector and put it to profitable use, particularly in the financial market. If these companies were to enter the financial market with their own offerings, they could quickly become systemically important. These companies could also become systemically important indirectly by making their data or infrastructures available to the financial services industry in exchange for a fee, thereby creating dependencies. For if the insights of these BDAI providers prove valuable for the financial services industry, by allowing more precise risk assessment, for example, this will lead to significant competitive pressure as virtually no competitor would want to miss out on these insights.

1.2.1 Banks

BDAI innovations foster the disaggregation of the value chain – possible separation between the customer interface and the product platform[3] | The spread of BDAI applications is intensifying the already prevailing trend toward the disaggregation of the value chain in banking. New competitors with BDAI-driven business models or BDAI applications are claiming parts of the value chain for themselves. While some new providers are entering the market at the customer interface, other providers are specialising in certain sections of the value chain on the product platform. New providers are now no longer forced to offer a complete banking product, but are free to concentrate on those parts of the value chain that offer them particularly high value added. As a result, incumbents could be pushed into providing the infrastructure services for third-party offerings. Furthermore, it is conceivable that incumbents will offer certain functions as services.

BDAI innovation could determine the outcome of competition at the customer interface | In offerings involving direct customer contact, the use of BDAI can prove a sustainable competitive advantage as it can help meet changing customer expectations. The key aspects here are, for instance, the speed of processes and decision-making, smooth interaction between provider and customer and the individualisation/personalisation of services. This is where bigtechs could leverage their experience in automating personalised services and in directly addressing customer needs in order to establish themselves at the customer interface. In fact, higher levels of competition at the customer interface could also lead to a redistribution of earnings.

Transaction data is valuable BDAI input and also attracts providers outside the financial sector | Compared with other data on customer behaviour and preferences, transaction data is of particular value for BDAI applications because it reveals what is virtually up-to-the-minute information on the actual spending behaviour and solvency of customers, as well as allowing accurate analysis. These analyses can serve, for example, as the basis for individualised bank offerings. In their role as "principal banks", incumbents find themselves well-placed to conduct such analyses: in addition to comprehensive transaction data from payments and account management, they also have an overview of the portfolios of all existing customers at their fingertips. However, payment service providers could also access this data, provided customers give their consent. Furthermore, companies outside the financial services sector – particularly bigtechs – could step up their efforts to secure access to this data, through offerings at the customer interface such as payment services, for example, in order to exploit the data obtained and thus optimise the offerings on their own platform.

BDAI innovations allow for efficiency and effectiveness gains in core processes | The use of BDAI on product platforms offers significant potential for all core banking processes, whereby the efficiency gains are typically conceivable in those areas in which there are sufficiently high numbers of similar tasks. This is primarily the case in the retail customer segment and among small and medium-sized corporate customers. Middle and back office settlement processes could be structured more efficiently by using BDAI applications to automate those processes that were previously too complex for automation. In compliance processes, too, especially in the prevention of money laundering, it is conceivable that applying BDAI would improve both efficiency and effectiveness. Furthermore, using BDAI could make risk assessment models more precise still. However, this has to be qualified by noting that extensive analytical and data-based optimisation measures have been introduced in exactly this area over the past few decades, which means that the potential for improvement might be more limited here than in the case of process automation and optimisation.

---

[3] In the following, the product platform comprises those middle and back office functions and processes that are necessary for delivering products and services to the customer interface (front office).

Sale of anonymised data analyses as a potential new business model – however, this is not set to replace traditional earnings | The monetisation of data presents financial institutions with the possibility of an additional source of income. This includes the sale of aggregated, anonymised information on certain transaction and master data. However, this is not expected to replace traditional banking earnings in the short term. Even assuming the formal consent of the customer, providers must also weigh up the value added by monetisation against any potentially negative effects on both the company's reputation and consumer trust.

1.2.2 Insurance companies

The customer interface as a generator of data and a success factor in marketing and sales | In the insurance industry, BDAI can be used in a multitude of ways, particularly in sales. Traditional claims data could be complemented by valuable (external) customer data and evaluated using BDAI methods. The resulting insights could then be used to tailor customer acquisition and cross-selling more effectively. This is of particular interest in the volume business, i.e. primarily in the retail customer business and in high-turnover segments, such as motor, home contents and residential buildings insurance. The notice periods for these insurance products are short, which is why a comparatively large number of contracts are entered into. Investments in BDAI technology to further develop the processes at the customer interface are set, as a corollary, to become increasingly important for traditional insurance companies in particular because this is where the new data is recorded – data that could complement traditional claims-based data.

New providers with expertise in data-driven business models could take over the customer interface | It is possible that new providers of insurance or insurance-related services will enter the market, and this development can already be observed to a certain extent. Bigtechs or insurtechs could try to utilise their core competences in order to take command of the customer interface through more highly personalised services and means of contact and meet changing customer expectations in this way. New providers pursuing this approach often act purely as insurance brokers. It is also conceivable that they might exploit their access to new and relevant data in order to offer their own insurance products either on their own or in cooperation with incumbents.

BDAI facilitates differential pricing as well as efficiency and effectiveness gains in core processes | If insurance companies were to make use of BDAI technology, it could change their core processes, such as product development or claims settlement, and increase both effectiveness and efficiency. Thus, new data, from wearables or telematics, for example, and new analysis methods could facilitate more differential pricing and risk assessment. Insurance companies could exploit this to develop more situational insurance products. Furthermore, new BDAI technologies could also help reduce risk costs, administrative costs and the expected cost of claims (e.g. through better fraud and loss prevention as well as loss minimisation). In particular, more highly automated processes to reduce administrative expenses in claims settlement could further simplify and speed up interactions between insurance companies and policyholders. However, depending on their risk profile and willingness to take advantage of the new offerings, policyholders could also face a number of disadvantages – in the form of modified premiums or generally higher barriers to insurance services.

BDAI-based claims prevention and the sale of anonymised data packages as new sources of income | Turning data into money – this new business model could also be used in the insurance industry although this has not yet been observed to any great extent. Insurance companies could monetise their data in two different ways. First, they could offer extended services to prevent claims, for instance, in the form of storm warnings within the context of residential building or motor insurance. By keeping in contact more frequently and providing useful information to the customer, insurance companies could increase customer loyalty. It is also possible that insurance companies will sell the data to other companies as anonymised data packages – naturally in compliance with the prohibition against non-insurance business[4] and within the bounds of data protection regulations. When dealing with policy-related data, insurers must take great care not to risk any damage to reputation or consumer trust.

1.2.3 Capital markets

More of the same, only faster and better | It is also likely that a greater use of BDAI in capital markets would lead to further gains in efficiency and effectiveness. In particular, BDAI is being increasingly introduced into more and more functions for front-office support, into the middle and back office as well as into compliance.

BDAI facilitates further automation as well as the greater use of algorithms. As a result, BDAI also allows for a rise in machine-machine interaction and faster processes in trading, sales and product development. An increase in or expansion of algorithmic (but not necessarily high-frequency) trading across all asset classes is foreseeable in the near future. This could, in turn, help exploit existing automation and digitalisation potential. Greater use of algorithms could also impact the area of advisory services, which deals with corporate mergers, for example, and tasks that have typically been performed by humans would be scalable and completed more quickly as a result.

So far, the influence of fintechs on the structure of the capital market as a whole has been limited. These companies have emerged primarily in the areas of data and analytics as well as in execution or execution technologies and in infrastructure. Fintechs have so far mainly developed in partnership with incumbents. Accordingly, they often function as an extended workbench, idea generator or specialised provider of process changes and efficiency improvements.

BDAI results in more players, more connections and thus more complexity | BDAI facilitates specialisation and the emergence of new market participants, which could contribute to a further fragmentation of the value chain. The interdependence of players so characteristic of the capital market could increase disproportionately as new market participants emerge. Therefore, BDAI could mean more players, more connections and thus more complexity.

In addition, those using BDAI benefit from easy and standardised access to data. As a result, the number of interfaces could grow and standardisation could become more widespread in the front, middle and back office. If parts of the value chain become easier to outsource, make-or-buy decisions may have to be re-evaluated in the future.

Furthermore, the role of data, analytics and process providers will become more relevant as BDAI is deployed, and, as a result of economies of scale, this group of providers may increasingly concentrate on offering their services to incumbents. Existing providers could benefit from self-enhancing data generation structures, whereby provided data is utilised to generate new data, which these providers then again make available to others.

---

[4] See section 15 of the German Insurance Supervision Act (*Versicherungsaufsichtsgesetz* – VAG)

Non-financial data such as social media content is increasingly used for predictions and valuations of illiquid assets could become more frequent | The use of BDAI could increase transparency for companies and allow them to improve their competitive positioning or develop new products. This is where the sheer variety of data used in the capital market becomes evident. Data ranges from social media content to satellite pictures. Furthermore, the financialisation of further societal and economic areas could be pursued, for example by hedging events that previously could not be insured, thus linking areas with the financial services sector that have had little connection with it up to now.

In addition, using BDAI could increase the frequency of valuations for illiquid assets. This would significantly accelerate the recording of fluctuations in market value, and these new volatilities could indeed make the marketplace more dynamic.

1.3 Success factors for BDAI applications and innovations

BDAI-readiness by setting strategic and operational IT framework conditions | BDAI applications can only be successfully implemented when data is available in sufficient quality and quantity. In order to provide this data, a suitable data platform is needed. Furthermore, BDAI applications demand new capabilities and methods. For instance, the demand for staff with a background in data science is growing, and more modern, agile working methods could be needed. And finally, as interconnectivity and data usage grow, it is vital to meet the information security challenges posed.

Consumer trust as the key to success for BDAI usage | Data-driven applications could be of value for many consumers, for instance, by enabling highly individualised products and services. Nevertheless, consumers are concerned about how their personal data is used – especially financial and health data. Both the misuse of personal data and any lack of information security can have a detrimental effect on consumer trust in the long term. This is not just a matter of consumer trust as a value-added factor, but also a matter of legal rights which are enshrined in the German constitution, such as informational self-determination.

Consumer data sovereignty enables trust in BDAI innovations | The potential of BDAI can only be exploited for financial services if it is possible to gain and maintain the trust of consumers by ensuring that their data is used as desired and in accordance with the law. Besides the technical approaches that allow for anonymised analyses, consumer data sovereignty could represent another approach. Consumers can only make a sovereign decision if they are adequately informed about the potential reach and consequences of the use of their data, if they are given reliable options for controlling how their data is used, and if they have actual freedom of choice. Providers are responsible for ensuring that these requirements are fulfilled.

1.4 Supervisory and regulatory implications

Financial stability and market supervision

Identifying and closing regulatory gaps early on | Innovations made possible by BDAI can open up the market to new providers. Their business models may not yet be adequately covered by the current regulatory framework. It is vital that such cases are identified and that the range of companies subject to supervision is extended accordingly.

Maintaining transparency and monitoring new structural relationships | Greater interconnectedness and complexity in the market can mean new risks at the interfaces between market participants. This interconnectedness can arise both indirectly, e.g. if the same models, data or platforms are used, and directly through new contractual and trade relationships made necessary in the first place by BDAI. And as these risks are no longer within the organisational structures of supervised firms, these firms may not be able to completely identify or manage these risks. The changing structure of this dynamic market and the resulting risks must therefore be evaluated and addressed from a supervisory and regulatory point of view.

Redefining and addressing systemic importance | BDAI could promote the development of new types of systemically important companies, e.g. providers of data, platforms and algorithms. Furthermore and as mentioned above, systemic importance does not necessarily arise from a legal entity, but could also result from the interactions between various market players. Traditional supervisory key figures would then be unsuitable for determining systemic importance. Likewise, established risk mitigation measures, such as capital add-ons, might not produce the intended effect. Therefore, the question arises as to whether and how the banking- and insurance-based concept of systemic importance needs to be further developed in order to also be applicable to new business models and new market structures.

Applying technological safeguards from trading to other areas | Technological safeguards that are already widespread in trading venues and in algorithmic trading are not inherently limited to these areas of application. This raises the question of whether such safeguards would be necessary and could even be usefully applied within a BDAI context, even outside of trading venues.

Firm supervision

Embedding BDAI within a proper business organisation | BDAI will create further opportunities for automating standard market processes. However, this does not imply that responsibility for the results of BDAI-supported processes is to be shifted to machines. When designing (partially) automated processes, it is therefore important to establish the right supervisory-regulatory framework and to ensure it is embedded in an effective, appropriate and proper business organisation. Appropriate documentation and effective control systems are required to ensure this. Ultimately, the responsibility for automated processes remains with the senior management of the supervised firm.

No black box excuses – explainability/traceability of models is necessary and can improve the analysis process | It is the responsibility of supervised firms to ensure that BDAI-based decisions can be explained and are understood by third-party experts. Supervisory authorities take a critical view of models that are categorised purely as black boxes. New approaches allow firms using such models to at least gain some insight into how these models work and identify the reasons behind decisions. In addition, a better understanding of models provides an opportunity to improve the analysis process – allowing, for instance, the responsible units in the supervised firm to identify statistical problems.

Continuing to develop existing governance concepts | As automation progresses, it may be necessary to extend existing governance concepts to include automated processes. For example, one could think about introducing special safeguards for certain selective and particularly risky BDAI applications, safeguards already used in other technical applications. In aviation, for instance, an additional independent system is used as a "second pair of eyes" to measure speed.

Defining supervisory requirements for the explainability and effectiveness of compliance processes | BDAI makes it easier to identify anomalies and patterns. It increases the efficiency and effectiveness of compliance processes, such as the prevention of money laundering and fraud. However, the results of the algorithms would have to be sufficiently clear to ensure that they can be understood and used by the competent authorities and law enforcement agencies. Furthermore, supervisory and regulatory authorities need to discuss whether specific minimum standards should be formulated with respect to the effectiveness of the methods used.

Preventing criminals from turning to less advanced firms | Should BDAI result in highly effective compliance processes in certain institutions, criminals may turn more to firms that are less advanced in this area. It is therefore necessary to monitor whether this will materialise.

Defining prerequisites for BDAI use in models requiring supervisory approval | Any use of BDAI in models that are subject to supervisory approval would also have to be approved by supervisory authorities accordingly on a case-by-case basis. Beyond the individual case, the question could be asked whether all BDAI methods are equally suited for use in models that require supervisory approval, or whether there are methods that should be ruled out per se. Furthermore, it is necessary to examine whether existing legal (minimum) standards for the data used and for model transparency are sufficient or whether additional requirements would be necessary. The degree of added complexity that would be appropriate or acceptable for a certain improvement in forecast quality could also be generally discussed from a supervisory point of view. In addition, in the case of dynamic BDAI models, it is necessary to examine which modifications constitute a model change in the supervisory sense, which supervised firms would then have to report and secure approval for where necessary.

Addressing increased information security risks and using BDAI to combat them | As a result of the aforementioned BDAI-related rise in complexity, there are new challenges to be met when managing information security risks. Firstly, the attack surface is growing, making monitoring more difficult; secondly, new forms of attacks are emerging. By contrast, BDAI could be used to defend companies from information security risks, for example, by analysing and detecting danger. For instance, BDAI could identify irregularities in consumer online banking behaviour and thus detect potential misuse.

Collective consumer protection

Preventing disproportionate extraction of consumer surplus[5] | BDAI provides deeper insights into customer characteristics and permits better forecasts of customer behaviour. In particular, linking financial transaction and behavioural data with data on preferences and needs can reveal information on a customer's willingness and ability to pay on a case-by-case basis. In the case of (mass) individualisation, each customer could be asked to pay the maximum price they are prepared to pay. It may be difficult for customers to detect the (hidden) price adjustment due to the (mass) individualisation of products and services. This makes it possible to extract the consumer surplus, especially for providers that have platform-based business models but do not specialise in financial services. This is particularly critical in situations where consumers urgently need a product – a loan, for instance – but do not actually have the choice of several different offers. Information asymmetries can arise where customers are unaware of the market value of the data they must provide ("paying with data") and therefore do not know the actual price of the services they are buying. Overall, consumers have to be made more aware of how their (financial) data may be used and how important or valuable it is.

Allowing for sufficient access to financial services | Selection mechanisms made possible in the first place by BDAI can disproportionately restrict the access of individual consumers to certain financial services. The situation can be particularly precarious if consumers are disadvantaged by having access to a narrower range of products but are unaware that this is caused by the personal data they have supplied.

Preventing unlawful discrimination | Algorithms could be based on features for which differentiation is prohibited by law. When using BDAI, providers have to ensure that practical applications do not lead to any discrimination of individual consumers or groups of consumers in contravention of existing legislation, both when programming the algorithms and when controlling the generated results.

Ensuring trust in the financial market – cooperation with data protection authorities could intensify | Guaranteeing personal privacy and informational self-determination is also a key requirement for building trust in the long term, particularly in the financial sector. Data protection authorities are indeed responsible for monitoring the implementation of data protection requirements; however, if data protection violations were to become more frequent, this could also have implications for BaFin. Thus, cooperation between financial and data protection authorities will become increasingly important.

Avoiding actual pressure to release data | If increasing numbers of consumers begin to feel a loss of control where financial services are concerned and there are ever fewer alternatives to providing data, the perceived or actual pressure on consumers to supply their data might intensify. One option would be for supervisory and regulatory authorities to ensure that there are also sufficient alternatives on offer in the form of conventional financial services and/or services that are economical with personal data. This is one way of avoiding any de facto pressure to release data. How exactly "economical with personal data" and "conventional" are to be defined in this context and which financial services are to be considered are matters to be discussed.

Utilising technical options for using BDAI with anonymised data | Many BDAI applications can be created with the help of sufficiently pseudonymised or anonymised data. Thanks to technical data protection measures (e.g. privacy-preserving data mining), innovations are also a possibility without further recourse to personal data and a "privacy by design" concept can further bolster consumer trust in BDAI innovations. In addition, the use of other innovative approaches may be a good way to ensure the secure handling of personal data even in the age of BDAI.

---

[5] Consumer surplus is the difference between the maximum price that a consumer is willing to pay for a product or service and the price that he/she actually has to pay on the market.

1.5 Overarching phenomena from a societal perspective

Information derived from digital user behaviour could restrict consumers' freedom to act | Digital user behaviour is often analysed by accessing personal data or data that permits certain conclusions to be drawn about individual customer characteristics. Such data often constitutes a key component of BDAI-based innovations and services. At the same time, information derived from digital user behaviour can also lead to individual behaviour being influenced and the freedom of the consumer being restricted. As an increasing amount of data is being generated and provided, it can be difficult for many consumers to assess what personal data is stored and how it is actually used. Consumers may feel forced to respond by changing their behaviour by either no longer using (online) services or by adapting their behaviour in order to optimise their digital footprint.

Avoiding regulatory heterogeneity | Heterogeneous regulatory and supervisory regulations apply in many places despite the fact that BDAI-based business models have the same impact across jurisdictions. This increases the risk of supervisory arbitrage being exploited. Considering the new market structures, it will become increasingly important to continue international dialogue to ensure a level playing field along the lines of "same business, same risks, same rules", while ensuring that the characteristics of national markets are still sufficiently taken into account, now and in the future.

II. Introduction: value creation from data with the aid of artificial intelligence

2.1 Big data and artificial intelligence as drivers of fundamental change

People and organisations are increasingly digitally connected, computer technology can understand human input with ever greater accuracy and can take on complex tasks such as language translation[6] and answering questions via virtual assistants[7]; while we are witnessing the beginnings of self-driving cars. Furthermore, computers are actually capable of exceeding the abilities of human experts in some cases.[8] These developments are driven by the wide range of possibilities of data analysis and data processing that can be summarised under the key terms "big data" (BD) and "artificial intelligence" (AI). In this context, "machine learning" (ML) has been the central methodological approach in recent years. This study takes a comprehensive look at the implications of this fundamental change for the financial sector – from the perspective both of financial service providers and of supervisory and regulatory authorities. The study examines opportunities as well as risks of an increased use of big data and artificial intelligence in the financial sector. In addition, this study identifies potential supervisory and regulatory implications as well as strategic questions that in future probably need to be addressed by the international supervisory community.

In 2001, the market research company Gartner carried out a systematic analysis of the challenges and potential arising from large volumes of data, during which it developed the '3V model' characterising big data with regard to three attributes:

"Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation."[9]

In addition to Gartner's 3Vs (volume, velocity and variety), further characteristics of big data were subsequently identified and the 3V model was extended accordingly. Two other dimensions are particularly important from a regulatory and supervisory perspective – and therefore relevant to the issues raised in this study: (data) veracity, i.e. data quality, and (data) value, i.e. the potential value of data if it is processed purposefully.[10]

Big data is a key element of artificial intelligence. In turn, the effective use of AI depends very much on the scope and quality of the available data used to train and test algorithms. Therefore, as far as possible, the two topics will be analysed and discussed together as big data and artificial intelligence (BDAI) in this study.

---

[6] Cf. Wu et al., 2016, Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. Available online: https://arxiv.org/abs/1609.08144, accessed on 19 February 2018.

[7] Well-known examples are Amazon's Alexa or Apple's Siri.

[8] For example, it has been observed that in some cases machines have provided a more accurate diagnosis of pneumonia and cancer than humans: Rajpurkar et al., 2017, CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. Available online: https://arxiv.org/abs/1711.05225, accessed on 19 February 2018. Liu et al., 2017, Detecting Cancer Metastases on Gigapixel Pathology Images. Available online: https://arxiv.org/abs/1703.02442, accessed on 19 February 2018.

[9] Gartner, 2017, Gartner IT Glossary > Big Data – From the Gartner IT Glossary: What is Big Data? Available online: https://www.gartner.com/it-glossary/big-data, accessed on 19 February 2018.

[10] Cf. IBM, 2015, Why only one of the 5 Vs of big data really matters. Available online: http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters, accessed on 19 February 2018.

The development of these methods already began during the second half of the 20th century.[11] The key element of AI is inductive learning: in contrast to traditional, rule-based approaches, not all variations of a task to be automated have to be defined ex ante. Instead, the algorithm learns in a way similar to humans, on the basis of sample and/or training data. Over the last few years, for instance, some major progress has been achieved in the area of speech recognition: neural networks are now able to transcribe human conversations just as well - if not more accurately - as human beings.[12]

The interplay between big data and artificial intelligence can be illustrated along three basic steps, at the end of which the ability of machines to act and implement based on BDAI becomes clear:
1) collecting data, 2) processing (analysing and learning from) data, and 3) taking action on the results of data processing. These steps follow a simple input-process-output logic and are made possible by key technologies that have seen significant progress over the last few years. Exemplary technologies for the input step are optical character recognition, voice and speech recognition as well as image and facial recognition. The process step involves, for example, processing information, learning based on examples and deriving recommendations and forecasts. Typical output technologies include natural language generation and implementing robotic process automation. All of these examples use machine learning methods and processes.[13]

These technologies allow for a wide range of applications which, in addition to the optimisation of models and the quality of their assessments, allow the adoption of human activities by machines. Whereas, for example, algorithms have long been used to apply risk models in banks, the interplay between documents scanned by using optical character recognition (input) and machine learning (processing) enables automatic document processing, i.e. text recognition even for poorly-structured documents (including handwritten documents).

2.2 A new momentum generated by big data and artificial intelligence

The relevance of BDAI increases through interactions between technology, business and consumers: first, the technological progress enables a broad implementation of BDAI applications in practice. Second, companies are increasingly using data to optimise their business models and processes. Third, consumer behaviour is being influenced more and more by digital applications, which in turn increases the generation and availability of data.

The last two points in particular can have a strong, self-enhancing effect on one another. This development is partly still in its infancy within the financial services sector; however, it must be assumed that it will be reinforced in future. Furthermore, central regulatory revisions are currently influencing the interplay between technology, companies and consumers. Of particular note are the revised EU directive on payment services (PSD 2)[14] and the European General Data Protection Regulation (EU GDPR). The following section outlines the interactions between technology, companies, consumers and the new regulatory environment (see Figure 1).

---

[11] For example, in the summer of 1956 as part of the Dartmouth Summer Research Project on Artificial Intelligence. Cf. J. Moor, 2006, The Dartmouth College Artificial Intelligence Conference: The Next Fifty years, AI Magazine, p. 87–91. 87 – 91.

[12] Cf. Xiong et al., 2016, Achieving Human Parity in Conversational Speech Recognition. Available online: https://arxiv.org/abs/1610.0525, accessed on 18 December 2017.

[13] Described in detail in chapter III.

[14] Payment Services Directive.

Greater data availability and
improved hardware performance

Data generation through
consumer behaviour

Marketable evaluation and
processing technologies for
data

Increased customer
benefits

Revised regulatory
environment (e.g. PSD 2
and EU GDPR)

Data use is a success factor
for value creation

Increasing
investments in BDAI

Technology   Companies   Consumers
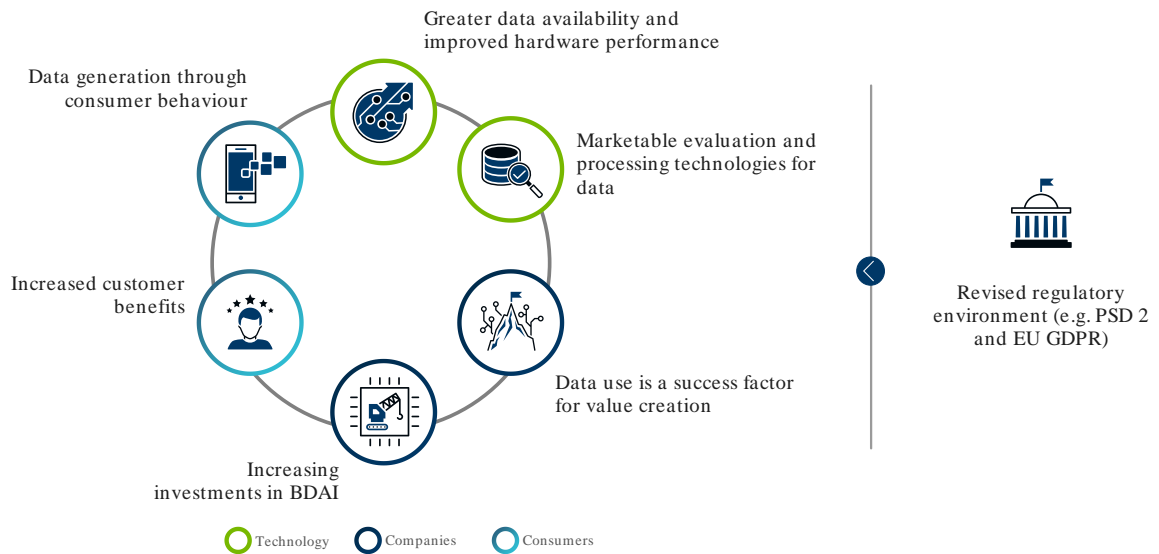
Large volumes of data are available along with powerful hardware and marketable evaluation and processing technologies.

The era of electronic data processing started back in the 1970s. At that time, the foundations were laid for the generation, storage and processing of large volumes of data. However, the mass data era only began a couple of decades later: with the expansion of the internet since the 1990s, the digital world has become part of the daily life of consumers, leading to rapid data growth. In addition, smartphones have created a massive increase in image, sound and position data since the 2000s. Today, the Internet of Things (IoT) is increasing the level of connectivity and has given rise to a growing number of measuring points via sensors over the last years.

Along with the rising volume of data, the processing power of computers has grown exponentially (see Figure 2). In addition, more cost-effective memory space is available, further improving hardware performance. Memory and computing capacity no longer have to be maintained internally but can be provided by third parties in the cloud. It is now possible to retrieve the required hardware power flexibly and selectively for specific periods. This technological development is lowering the barriers to the use of BDAI by companies.

Figure 2: Increase in the global volume of data and available computing power[15]

Global volume of data (ZB)

Historical | Forecast

163 ZB expected in 2025

Max. speed (FLOPS)
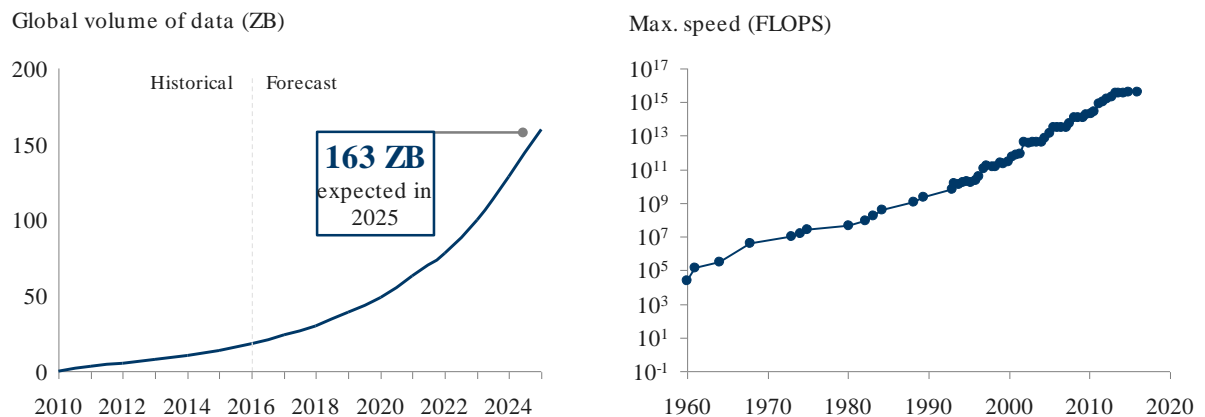
The possibilities of BDAI technologies and corresponding applications have increased massively in recent years. For example, machine vision is an elementary component of today's advanced autonomous driving[16], while automatic face recognition often delivers more accurate results than facial recognition by humans.[17] One example of the progress in automatic speech recognition would be smart loudspeakers that are controlled primarily by voice.

Technology companies have proven that the use of BDAI can deliver a significant competitive advantage. As a result, more and more capital is flowing into big data solutions and AI-related companies such as fintechs.

The availability of data and the hardware performance needed for the commercialisation of data led to a transformation of value creation in many companies.

Data-driven business models that monetise data primarily based on BDAI have made some technology companies the most highly valued companies today in terms of market capitalisation (see Figure 3). Currently, seven of the ten companies with the highest market capitalisation have a data-driven business model and are considered to be the most innovative technology firms.

---

[15] FLOPS = floating point operations per second; sources: IDC Digital Universe Study (2012), IDC DataAge 2025 Study (2017). Available online: https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf, accessed on 19 February 2018, Denning & Lewis (2017), Exponential Laws of Computing Growth. Available online: http://www0.cs.ucl.ac.uk/staff/ucacbbl/gggp/langdon_sse_8-feb-2017.pdf, accessed on 19 February 2018, Nordhaus (2007)

[16] Cf. Dubey, 2016, Stereo vision—Facing the challenges and seeing the opportunities for ADAS applications. Available online: http://www.ti.com/lit/wp/spry300/spry300.pdf, accessed on 12 December 2017.

[17] Cf. MIT, 2017, Facial Recognition Is Getting Incredibly Powerful—and Ever More Controversial. In: MIT Technology Review. Available online: https://www.technologyreview.com/the-download/608832/facial-recognition-is-getting-incredibly-powerful-and-ever-more-controversial/, accessed on 12 December 2017.

Figure 3: Proportion of companies with data-driven business models among the top 10 companies by market capitalisation[18]

[%]	Q4 2006	Q4 2011	Q4 2017
100	Toyota Motor Corporation	Walmart	JP Morgan Chase
	Industrial and Commercial Bank of China	Chevron Corporation	Johnson & Johnson
	BP	IBM	Berkshire Hathaway
	Royal Dutch Shell	Industrial and Commercial Bank of China	Alibaba Group
	Bank of America Corporation	Royal Dutch Shell	Tencent Holdings
50	PetroChina Company Limited	PetroChina Company Limited	Facebook
	Citigroup		Amazon
	General Electric Company	Exxon Mobil Corporation	Microsoft
		Alphabet[1]	Alphabet
	Exxon Mobil Corporation	Microsoft Corporation	
	Microsoft Corporation	Apple	Apple

■ Technology companies with data-driven business models

[1] At that time still "Google".

Investments in BDAI are increasing worldwide. Investments in big data solutions are likely to increase by 13 percent per year until 2021 relative to 2016.[19] Furthermore, global investment in related fintechs has increased considerably over the last few years. From 2014 to 2016, global funding in AI-related fintechs grew by approx. 60 percent to over $2 billion. The capital volume used to finance AI-related fintechs in 2016 was around seven times higher compared to 2010.[20] Hence, the high liquidity in the markets is also being attracted by BDAI solutions and concepts.

BDAI applications can create greater customer value. The use of respective services and products generates additional data, which in turn strengthens the momentum of BDAI.

Processes at the interface between companies and customers can be made much simpler and faster. Furthermore, BDAI enables data-based, targeted real-time sales control, which in turn enables tailored products and services for customers. An increasingly granular segmentation of customers as well as customisation of products can increase consumers' willingness to accept such offers.

The often quick and easy access to digital services in different industries (including through the high market penetration of smartphones) can also change consumer expectations towards financial services. Consumer behaviour is increasingly shaped by digital technology based on the processing and utilisation of consumer data. In turn, this can lead to higher availability of data in the financial context. At the same time, however, there is often a high degree of mistrust towards the use of data by companies. For example, according to certain surveys, European consumers are critical of the commercialisation of their personal data.[21] Consumer trust is therefore an important success factor in the BDAI context. This is particularly true in the financial sector, as inadequate data usage practices could undermine trust in the financial market as a whole. Overall, BDAI can increase the knowledge asymmetry between customers and providers, making it potentially harder for customers to understand whether their personal information is actually used in their interest.

---

[18] S&P capital IQ: Top 10 companies' (worldwide) market capitalisation on the following days: 29 December 2006, 30 December 2011, 29 December 2017.

[19] Investments in the form of licences for the delivery model, ongoing maintenance, on-premise subscriptions, cloud subscriptions and other services. Source: Forrester, 2016, Big Data Management Solutions Forecast, 2016 to 2021 (Global).

[20] 2015 growth in AI-related fintech funding, particularly through Palantir's $880 million round of financing. Source: BCG Fintech Control Tower.

[21] Cf. De Mooy, 2017, Center for Democracy and Technology, Datensouveränität in Zeiten von Big Data, p. 8 Available online: https://www.bertelsmann-stiftung.de/de/unsere-projekte/teilhabe-in-einer-digitalisierten-welt/projektnachrichten/datensouveraenitaet-in-zeiten-von-big-data/, accessed on 18 January 2018.

Regulatory adjustments can act as additional accelerators of changes brought about by BDAI, but also require companies to find answers to pressing issues such as data protection.[22]

With PSD 2, account information and payment initiation services, which until then had often been offered in a legally grey area, became regulated for the first time. Such service providers are now recognised as official market players. PSD 2 enables them to offer innovative services at the customer interface, since they can obtain data from established financial service providers through clearly defined interfaces, if the customer consents. The EU GDPR, on the other hand, implies fundamental changes in data protection regulations and requirements all of which also affect financial service providers. From the perspective of consumer and data protection, specifications of the EU GDPR raise new questions regarding BDAI that providers must adequately address to meet consumer expectations. This includes the explicit consent of consumers to the use of data.

Some international supervisory and regulatory authorities have already looked into the potential for change in big data and artificial intelligence in recent months. Their results are briefly summarised in chapter 7.2

2.3 Objectives and structure of this study

This study attempts to take a holistic perspective. It explicitly highlights the opportunities and risks of an increased use of BDAI in the financial sector – specifically in terms of financial stability and market supervision, corporate supervision and collective consumer protection. Three perspectives are analysed: First of all, the financial sector angle, i.e. banks and financial service providers, insurance companies and capital market players.[23] Secondly, the consumer viewpoint. And thirdly, the point of view of supervisory and regulatory authorities. The results will be used to derive potential supervisory and regulatory implications as well as strategic key questions that might need to be addressed in future by the international community of supervisors and regulators. The report does not, however, confine itself to the key questions, it also puts forward possible solutions for discussion. These are of a strategic, technical and/or regulatory nature.

A fundamental differentiating factor of this study is the strategic-economic assessment of the implications for market participants. This assessment covers banks, insurance companies and the capital market. It includes the description and assessment of possible market developments by harnessing the full potential of BDAI as well as the systematic presentation of nine specific use cases. Combining a systematic market view with specific use cases allows to identify and describe particular issues that are relevant to the supervisory and regulatory perspective in a comprehensible way. As part of the market assessment, the study looks qualitatively at the economic potential of big data and AI applications in order to provide a differentiated view of the potential penetration of BDAI technology in the market and its use by financial service providers. Hence, the study is not restricted to existing BDAI applications but also looks at the disruptive potential of BDAI. The assessment is based on extensive literature research, evaluation of current internal data analyses and the independent data analyses of industry experts.[24] In addition, market participants were consulted to assess market developments.

---

[22] See Chapter IV.

[23] In this study, we subsume – inter alia – the following under the term "capital market": securities services firms, investment funds, asset management firms and operators of trading venues (stock exchanges).

[24] The market assessments and analyses of the (technological) success factors for BDAI applications were supported by the management consulting firm The Boston Consulting Group.

The holistic approach is complemented by a detailed description of the technical and strategic implementation of BDAI. To support the technological analyses, experts from science and research were involved in preparing this study.[25] In this study, analytical methods used in the field of BDAI are presented for a broader readership in the sense of a short reference work. Innovative approaches are also taken into account; for example, an approach is presented that allows data to be used economically and evaluated while simultaneously preserving privacy. With regard to the strategic implementation of BDAI, the study also answers the question which conditions companies have to create in order to use BDAI successfully. In addition, maintaining consumer trust is discussed as another particularly important success factor for the use of big data. Furthermore, a broad IT strategy is required that ensures a functioning organisational structure and data infrastructure.

Subsequently, the study is structured in four chapters:

Chapter III ("Technological prerequisites for the use of big data and artificial intelligence") illustrates the technological complexity of BDAI for a broader readership, presents innovations in this field and addresses topics such as the comprehensibility of algorithms.

Chapter IV ("Strategic prerequisites") describes the strategic success factors for realising BDAI in the financial services industry. The presentation of conditions for the successful realisation of BDAI helps to assess the potential for change of BDAI and extends the scope of the study to additional topics that are relevant to regulatory and supervisory authorities. This chapter thus focuses on the topics of consumer trust and technological success factors such as data and cloud computing, as well as necessary competences such as employee qualifications and agile methods of working. The chapter is preceded by an analysis of BDAI from the consumer perspective (Chapter 4.1.1).

Chapter V ("Market analyses") looks at three sectors. For each of these sectors – banks, insurance and the capital market – a BDAI assessment is conducted and three use cases are presented. These use cases each relate to one of three specific areas of application of BDAI (customer interface, core processes or new business models).

Chapter VI ("Supervisory and regulatory implications") identifies specific regulatory and supervisory implications based on the preceding chapters. The identified implications are discussed from a regulatory and supervisory perspective according to three focus topics (financial stability and market supervision, supervision of firms, and collective consumer protection). In addition, key questions are also derived.

---

[25] Support for the technological assessments in this study was mainly provided by the Fraunhofer Institute for Intelligent Analysis and Information Systems.

III. Technological prerequisites for the use of big data and artificial intelligence

Summary

Machine learning (ML)

Machine learning is the term used for enabling computers to learn from data by means of suitable algorithms. This allows them to set up a model of their world and solve assigned tasks better.

ML approaches can be characterised on the basis of the learning task (e.g. classification, regression, clustering), the data types (there are special approaches e.g. for text, speech or image data) and the algorithms (technical solution for the problem).

ML processes cannot operate independently; they have to be created, applied and evaluated by specialists – so-called data scientists.

Technical background

Big data is one of the main drivers behind the success of machine learning. Advantages can be gained not only from the growing amount of data available but also and particularly from the increasing possibilities for automatically evaluating data that previously could not be analysed (such as texts, images, speech and sensor data).

New hardware solutions for the extensive parallel computing and shrinking hardware costs is another main driver behind machine learning with big data.

Algorithms

ML is highly dynamic, with many new approaches invented over the past few years. One successful approach behind many prominent applications is deep learning, i.e. the learning of complex neural networks. Deep learning can play on its advantages particularly when applied to mass data.

Classic methods often used are classification methods such as decision trees or support vector machines, regression methods such as linear regression and clustering methods such as k-means clustering.

Reinforcement learning, which can be used to control machines or intelligent agents, is another method that is currently gaining importance.

Artificial intelligence (AI)

There is currently a basic paradigm shift taking place in IT: Whereas it used to be that specialists and IT experts programmed the behaviour of complex systems, it is machine learning with big data that is currently proving superior in many areas of application. In some areas, computers already perform better than human experts.

The success of AI can still only be found in limited areas of application, however. Approaches for the general simulation of human intelligence – hard AI – are still not foreseeable.

In a nutshell, the development can be reduced to the equation: Big data + computing resources + ML = AI

New applications

Question-answering systems answer questions asked in natural language. A current example is Alexa from Amazon.

Dialogue systems engage in spoken dialogues with humans and are of interest for use in call centres, for example.

Intelligent agents carry out complex actions on behalf of humans. Outstanding performances by intelligent agents can currently be found in the games sector, e.g. Go.

New challenges

Many social requirements are being picked up on in machine learning and tackled with suitable algorithms. These include data protection, which is addressed by privacy-preserving data mining, the non-discrimination of groups of people during automatic decisions, the transparency and explainability of complex models and their results as well as data sovereignty.

There are machine learning approaches for all these challenges, yet ML alone cannot solve these issues, since they are not of a purely statistical nature – rather they pose social questions at the same time. There is no generally accepted solution available for any of the issues in particular, the field is currently the focus of numerous research activities.

3.1 From big data to artificial intelligence

There is currently a basic paradigm shift taking place in IT. Whereas the behaviour of complex systems used to be programmed jointly by domain and IT experts or at least adapted from data by means of carefully selected statistical model assumptions, machine learning is currently proving superior in many areas of application.

The idea of giving computers intelligent skills through machine learning is not new by any means; on the contrary, it arose very soon after the first computers were invented. As long ago as 1950, Alan Turing, a British mathematician and one of the thought leaders in the field of artificial intelligence, assumed that a really intelligent system as a whole would not be achieved by the detailed pre-programming of individual behaviour patterns. Instead, he imagined machines which would interact with their environment and thus become artificially intelligent:

"Our problem then is to find out how to programme these machines to play the game. At my present rate of working I produce about a thousand digits of programme a day, so that about 60 workers, working steadily through the 50 years might accomplish the job, if nothing went into the waste paper basket. Some more expeditious methods seem desirable."[26]

However, it was only the general availability of large quantities of data that led to breakthroughs in machine learning which permitted computers to perform comparably to human skills in their solving of certain tasks. This paradigm shift from programming to learning has far-reaching consequences: providing there is sufficient data available, problems that classic programming cannot solve become solvable. Nevertheless, the success of machine learning can still only be found in very limited areas of application. Approaches for the general simulation of human intelligence – hard AI – are still not foreseeable.

3.2 Machine learning and its algorithms

In general terms, machine learning (ML) describes the notion of enabling computers to learn from data and experience by means of suitable algorithms. Computers can use these to set up a model of their world and solve the tasks assigned to them better.

Quite often, a distinction is not made between the specific application of machine learning methods and the overall process of knowledge discovery which, according to one definition frequently used, is described as "the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data"[27]. This overall process usually includes a strong manual component that covers the definition of the task, selection and processing of data and the evaluation of results in particular, while the term machine learning describes the automatic part of the model preparation.

Dimensions of machine learning

Typically, a learning algorithm generates a model from data in a first phase: the learning or training phase. In a second test phase, the model is tested statistically for quality and – if possible – assessed in terms of content by the users. During the subsequent application phase the model is applied to unseen data in order to issue predictions. Typical applications of machine learning methods in the financial sector include credit ratings, the forecasting of share prices, the detection of attempted fraud in credit card transactions or the grouping of clients for sales purposes.

---

[26] Turing, 1950, Computing machinery and intelligence. Mind 49, pp. 433–460.

[27] Fayyad et al., 1996, From Data Mining to Knowledge Discovery: An Overview. Advances in Knowledge Discovery and Data Mining, In: AAAI Press, pp. 1-30.

Applications of machine learning methods can be characterised using different dimensions. The most important of these are:

■ **Learning task:** Which formal mathematical task is to be solved? Various standard learning tasks have emerged which cover most applications. The most important of these are supervised learning, unsupervised learning and reinforced learning. These will be described in more detail below.

■ **Data types**: Classical machine learning started with table-type data. Later, specialised method classes were formed for different data types, such as for the analysis of text data, images, audio/speech, networks and time series.

■ **Algorithm/model class**: A technical distinction can be made on the basis of the type of algorithm or the type of model. Typical model classes are certain types of numeric functions (e.g. linear functions, polynomials, Gaussian functions) or logic models (e.g. if-then rules and decision trees). One type of algorithm class that is currently very prominent is deep learning.

These dimensions are not independent of one another. For instance, very different learning tasks can be solved using very similar algorithms and very special algorithmic approaches that cannot be used for other types of data have been developed for special data types.

3.2.1 Supervised learning

The goal of supervised learning is to forecast a target feature (also known as a label) on the basis of explanatory features[28]. The label can be both a nominal feature (e.g. client cancels or does not cancel his contract) or a numeric feature (e.g. claim amount). The first case is known as a classification problem and the second as a regression problem. Please refer to chapter 7.1.1 for more details on the representative method of supervised machine learning.

With supervised learning, it is important to develop a general rule in order to apply the knowledge learned to new data patterns. The pure memorisation of data points previously seen is insufficient for this. If the model manages to generate predictions on the basis of unseen data that are of a similar quality to those generated with training data, it is said that the model generalises well. In this sense, data points are usually called examples, since they are examples for the correlation to be learned.

---

[28] In statistics and in machine learning different conventions have been formed: features are called variables and the target feature in particular is described as the dependent variable.

3.2.2 Unsupervised learning

Unsupervised learning uses different methods to extract intrinsic patterns from the data. In contrast to supervised learning, there is no specific question that can be coded in a concrete target feature. This makes unsupervised learning more difficult than supervised learning, since the quality of a solution is often only determined subjectively by the user and cannot be integrated objectively in the algorithms using statistical measures.

Two of the most important learning tasks in unsupervised learning are clustering and anomaly detection, for instance to guarantee data quality. Clustering aims to divide data into clusters (groups) of similar examples, i.e. two examples in the same cluster should be as similar as possible and as dissimilar as possible to examples from other clusters. One possible application for this is client segmentation on the basis of social-demographic characteristics or purchasing behaviour. One type of clustering method often used is k-means clustering, which will be explained in more detail in chapter 7.1.2.1 using examples from the financial and insurance sector.

Anomaly detection (or outlier detection) is used to detect individual data points which do not conform to the pattern of most items in a dataset. One typical example is fraud detection. Both tasks are related in that the data points which cannot be assigned to any one cluster can be regarded as anomalies. Please refer to chapter 7.1.2.2 for more details on anomaly detection.

3.2.3 Reinforcement learning

A further relevant learning task in addition to supervised and unsupervised learning is reinforcement learning[29]. This is used in situations where computer systems can carry out different actions in an environment and only occasionally receive feedback in the form of a reward, depending on how good the achieved situation is. The machine must learn which actions it should choose to maximise feedback (reward function).

Reinforcement learning is used for learning computer gaming strategies, for example, by the machine playing against itself. It can also be used for software agents, however. Thus, for example, an intelligent agent on a website can use suitable actions to try to persuade the user to do something specific (e.g. click the advertising or conclude a contract).

3.2.4 Deep learning: machine learning with big data

One currently very prominent class of learning methods, which is behind numerous breakthroughs in intelligent systems, is deep neural networks. The area of machine learning which deals with this is known as deep learning[30]. Due to the importance of deep learning in many current applications, this field will be explored in more detail here.

As far as the concept is concerned, there is no distinction between this and the classic neural networks described above. However, deep learning uses significantly more complex network architectures that can have hundreds of layers and billions of parameters to be learned rather than the two layers used by classic networks.[31] This means that these networks can be efficiently learned from data despite their extremely high degree of expression. The most important decision for the use of deep learning is thus related to the selection of the correct network architecture.

---

[29] R. Sutton and A. Barto (1998). Reinforcement Learning: an Introduction. MIT Press.

[30] Goodfellow et al., 2016, Deep Learning. MIT Press. http://deeplearningbook.org.

[31] Szegedy et al. (2015). Going Deeper with Convolutions. In: Computer Vision and Pattern Recognition (CVPR).

Deep neural networks can process all kinds of raw data which is coded as numeric vectors, including texts, images or speech. The secret behind their success is that they can learn representations independently from the raw data which make the actual task easier precisely because similar representations have a similar meaning. They automatically find structures in the given examples which provide suitable features for the actual learning task. Certain data preparation steps such as the computer-graphical detection of edges and faces as well as the linguistic detection of sounds and words are no longer required. When the model as a whole is trained, from input in the form of raw data through to output, it is also known as end-to-end machine learning. Thus it is more advanced than the classical method: With a computer program the entire behaviour must be constructed manually, with the machine learning of the 1990s it was only the features that had to be chosen with great care. Current neural networks automatically learn more and more abstract representations in their hidden layers and automatically deal with feature selection at the same time.

Neural networks are distinguished on the basis of the number and width of layers and the connections between these. With deep neural networks, the possibilities are endless in theory, and networks with new structures are being added all the time[32]. The structure of the networks and partial networks is always aligned to the learning task, as well as particularly to the type and importance of the input. Thus convolutional neural networks[33] (CNNs) were developed specifically for image processing, recurrent neural networks[34] (RNNs) for the analysis of time series data and long short-term memory networks[35] (LSTMs) for the analysis of successive sequences.

The development of deep learning lies both in advanced algorithm development and in big data: Constant growth in data quantities presents machine learning with the challenge that many of the classical learning methods can no longer be carried out efficiently on such large data quantities despite advanced developments in the fields of memory technologies and processors. Fast execution requires parallelisation over many computer nodes, such as on GPU computers (Graphics Processing Units). Deep learning is thus characterised by the fact that it is one of the few classes of learning methods that both permits very complex models (neural networks) and can be trained efficiently with big data using parallel computer infrastructures. On the other hand, it is a disadvantage of deep learning that its application only makes sense at all when a very large quantity of data is used, whereas for smaller quantities of data classical methods can demonstrate their strengths.

3.3 Evaluation and implementation of machine learning

The learning methods described generate models of different qualities, which is why it is crucial to be able to evaluate and compare the quality of a model and steer the data analysis process to the best possible result. This chapter describes the evaluation possibilities for the model quality of individual methods as well as the usual procedures in data analytics projects. The intention is particularly to demonstrate that BDAI does not provide universally usable methods which deliver optimum results without human help. Rather, various methods must be evaluated and the data often has to be prepared in a complex way.

---

[32] Van Veen, 2016, The Neural Network Zoo. Available online: http://www.asimovinstitute.org/neural-network-zoo/, accessed on 23 October 2017.

[33] LeCun et al., 1998, Gradient-based learning applied to document recognition. In: Proceedings of the IEEE, 86, pp. 2278-2324.

[34] Elman, J.L. (1990): Finding structure in time, In: Cognitive science Vol. 14, pp. 179-211.

[35] Hochreiter et al., 1997, Long short-term memory. In: Neural computation Vol. 9, pp. 1735-1780.

3.3.1 Evaluation of models

To avoid potential problems during the application of a learned model, this model must be evaluated carefully. Quality problems can occur for different reasons. The most important of these are:

- Discrepancy in the formalisation of the learning task: The criteria crucial for the business goals usually cannot be completely formalised – they are simplified or modified to allow them to be implemented efficiently in an algorithm. The models must be examined with respect to this discrepancy.

- Data selection bias: A further problem is that distortion can take place even during data collection. This can be caused by errors in the data collection and processing process (through missing data or when information which is not available at the time the model is applied in practice has been included in the training dataset) or by the world itself changing (in historic credit decisions minorities were discriminated against, so the historic dataset related to credit failures is distorted to the disadvantage of minorities). The danger here is that a data bias cannot be found from statistical tests, since any test data may also have the same bias. Instead, these problems must be found by making the model transparent, such as by examining the important attributes.

- Over-fitting and complexity: Since models are always learned from a finite amount of data, there is a risk that they cannot correctly represent the real process described by the data. As shown in Figure 4, the problem can either be that the model class of the algorithm is too simple to adequately describe the data (high bias) or that the model is so complex that it fits random deviations in the data (high variance). This latter case is referred to as over-fitting. This situation is risky, since the quality of the model for the training data seems high although in reality it is not. An important basic principle for avoiding over-fitting is that models should always be tested on new data not used for training.

Figure 4: Model complexity and generalisation ability

| Model too simple | Good model | Model too complex |
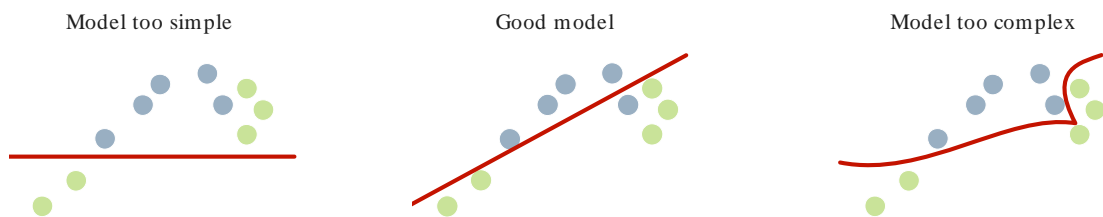
Various performance dimensions exist to evaluate models by quantifying the quality of a model in relation to given data. In the case of binary classification problems, for example, the comparison of model decision (yes/no) and true label results in a so-called contingency table.

	True label "yes"	True label "no"	Performance dimensions
Model prediction "yes"	True positives (TP)	False positives (FP)	Precision = TP/(TP+FP)
Model prediction "no"	False negatives (FN)	True negatives (TN)	
Performance dimensions	Recall = TP/(TP+FN)		Accuracy = (TP+TN)/(TP+FP+FN+TN)

This contingency table can be used to compute different performance dimensions. The dimension most often used is accuracy, which corresponds to the share of correctly classified examples. In practice, it is often the case that one label occurs significantly less often than the other, in which case performance dimensions such as precision and recall are more relevant. Many other performance dimensions exist for regression or clustering tasks as well, for example.

During the operative use of models, in particular when new models are continually being learned for new data, a one-off evaluation of model quality is insufficient. Rather, input data, model quality and the quality of the surrounding software environment as a whole must be continually monitored and tested for changes. Systems which are based on machine learning make high demands on software quality and monitoring because of their significantly higher dependence on data.[36] In terms of technology, different approaches for the validation of self-learning algorithms are conceivable:

- Automatic implementation of the above-mentioned test methods for every evolution step of the algorithm. This requires the test algorithms to be certified

- Use of a conventional reference system in order to ensure that the self-learning algorithm moves within predefined limits

Ultimately, new methods must be developed for the direct certification of a self-learning algorithm which test whether the algorithm performs correctly when any amount of new data are supplied.

---

[36] Sculley et al., 2015, Hidden Technical Debt in Machine Learning Systems. In: Advances in Neural Information Processing Systems, p. 28.

3.3.2 Data analysis as a process

Machine learning methods cannot operate independently, they must be developed and applied by specialists within the context of a greater process. Before a machine learning process can be used, the essence of the problem must be translated into a meaningful analysis task first. If the task involves cost reduction by improved fraud detection, for example, the problem must be reduced to a known learning task for machine learning methods to be used (in this case the learning of a classification model on the basis of known historic cases of fraud). In addition, important practical assumptions and limitations of the problem must be formally defined (for example: How bad is the incorrect marking of a legal transaction as fraud? Is success measured on the basis of the number of cases of fraud prevented or the amount of money saved?). In a second step, the relevant data must be provided. In the case of fraud detection, this can be transaction data, master data of the account holder and external data related to retailers, for example. Knowledge of the business application and potentially relevant data are outside the world of machine learning and must therefore be defined by a human user.

The concrete application of machine learning methods then takes place in two steps, during which the data for the learning method are provided in the form required by the concrete case on the one hand and, on the other, the learning method is parameterised and executed. Both steps are closely connected. A suitable representation of the data is used to transfer background knowledge possessed by the human expert to the learning method, which cannot unlock this knowledge alone or only with great difficulty. For instance, in the case of fraud detection, the addresses of retailers and the time stamps of transactions can be used to calculate the speed at which the user must have moved between them. An extremely high speed – the card used at noon in Frankfurt and then again at 1pm in Tokyo, for example – would be a very good indicator of a copied credit card. If the "speed" feature is provided through manual pre-processing, its use for most learning methods is trivial, whereas the detection of this pattern from the raw data is extremely difficult.

Figure 5: Illustration of the CRISP-DM process
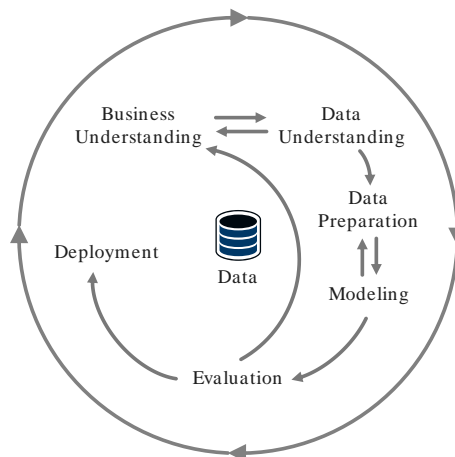
After application of the learning method, the results must be verified. This can be done purely statistically or by using qualitative factors. If possible, however, various different approaches are used, both to make the model transparent at least in part and to have it examined by the application experts. Both steps require human checks. The application in the real world can take place purely by the human (e.g. by cancelling cooperation with retailers whose names often turn up) or within an automatic system (e.g. in a bank authorisation system). Increasingly, intelligent systems are being implemented with functionalities provided by machine learning models, such as dialogue systems.

Since the 2000s at the latest, the increasing use of machine learning in companies has led to the realisation that this is a plannable method that can be standardised. One of the well-known process models which standardises the above-mentioned process is the "Cross Industrial Standard Process for Data Mining" (CRISP-DM). This is an iterative process with six phases, as shown in Figure 5.

There are other specialised test methods available especially for the financial sector. Different specialist test methods are used for regulatory credit rating models, for example, which can be adopted without change:

- Test case generation: Systematic generation of test cases which do not occur in the training data but could occur in a changed market situation for example, including monitoring of the plausibility of the model predictions

- Representativeness: Proof on the basis of certain assumptions about the data that the model is representative for the productive data

- Stress tests: Systematic changing of the current data in order to check the effects of external shocks (e.g. deterioration of property prices)

- Plausibility checks: Evaluation of whether the forecasts or model parameters are plausible. With linear models, individual parameters can be checked for their operating direction With more complex models, it can be checked whether the operating direction of the overall score shows the expected behaviour, e.g. whether in the case of car insurance the forecast of accident probability declines with the no claims bonus period.

The situation is significantly more complex where intelligent agents (bots) are used. This is the term used to describe technologies which replace more complex human activities, such as dialogue systems or automatic share dealing through to the development of strategies. The specific technical measures that are feasible and make sense depend largely on the respective application case and the connected risk profile.

3.4 Machine learning-based systems and applications

There are already numerous applications which not only use methods of machine learning but also contribute brand new skills and thus create added value. Some of these applications are explained in more detail below.

3.4.1 Automated document analysis

In traditional data processing, texts are not legible due to their lack of structure unless they are predefined forms. Yet at the same time, documents are the central data basis for many companies. Text mining can be used to classify and group texts as well as to automatically extract information and make it useful. More recent text mining applications automatically summarise texts or can answer questions about them, which in turn generates further possible uses.

Typical text mining applications in the financial sector are the automatic extraction of important information from descriptions of damage in order to process insurance claims, the analysis of typical reasons for customer complaints or, in future, possibly even fully automated customer communication. The digital development of documents for the purpose of process automation is the object of several use cases in this study.

Text mining differs from other areas of machine learning in terms of content, particularly in the way data are prepared and made available to the algorithms. If the text is only available as a hard copy or scanned image, analysis may have to be preceded by an optical character recognition (OCR) step, for which standard software is available.

The methods then used for classification again correspond to the standard methods. Due to the fact that text data are often available in large quantities, deep learning methods in particular have been very successful. One basic technology is "word embeddings"[37], which map words to highly dimensional numeric vectors so that words with a similar meaning are mapped to similar vectors. Such approaches open the use of numeric learning methods for text mining. Text mining approaches are very much dependent on the respective language. On top of this, different amounts of text data are available in different languages and the research communities differ in terms of size. This has meant that high-quality solutions have so far only been available for the most important languages, particularly English and Chinese, which results in a competitive disadvantage when other languages need to be processed.

Driven by these developments, there are already some software solutions available that have been developed using artificial intelligence technologies that structure any type of document form, make them searchable, check them, extract information from them or even generate new documents. Apart from digitisation efforts, companies are also pressing ahead with the integration and development of such solutions to meet regulatory requirements.

3.4.2 Speech recognition

Automatic speech recognition transforms audio signals into text, thus making the spoken word available for further analyses and services. Speech recognition is based on machine learning models and is traditionally made up of three components: An acoustic model describes what language phonemes sound like, a dictionary contains information related to pronunciation and a language model describes which sequences of words (sentences and other expressions) are most probable. In the meantime, however, the end-to-end models described above are also being used increasingly for this task. These map the entire recognition process in one complex deep neural network.

One possible application of speech recognition in the banking and insurance sector is in call centres. Speech recognition can be used for different purposes[38]:

- Detection of emotions, stress, wishes and expectations in the customer's voice. The customer experience can be improved by a suitable reaction

- Monitoring and coaching of call centre agents to improve service quality

- Reduction of operative costs through automation

- Identification of up-selling and cross-selling possibilities during a conversation by monitoring the course of the conversation

- Automatic recording in text form for purposes of documentation and further processing

---

[37] Mikolov et al., 2013, Distributed Representations of Words and Phrases and their Compositionality. In: Advances in Neural Information Processing Systems 26.

[38] http://searchcrm.techtarget.com/report/Top-five-benefits-of-speech-analytics-for-the-call-center, accessed on 20 February 2018.

Speech recognition must mainly be regarded as a technology which enables the system to communicate at a basic level through language. Dialogue systems take this technology further by enabling computers to hold conversations. To achieve this, the generation of a suitable answer is required in addition to the pure recognition of what the partner in conversation is saying.

3.4.3 Question answering

Question-answering systems can answer questions in natural language, for example when a customer asks for their bank balance, the current figure in euros can be given as the answer. Current systems can answer more or less complex questions relatively accurately. Complex questions such as why questions or logical questions which require world knowledge are still beyond current possibilities, however. Question-answering technology became known through the appearance of the Watson system[39] on the quiz show Jeopardy.

The answers usually come from knowledge bases or text documents. Question-answering systems have to understand the contents of a question, map the understood content of a query to their answer base and select the most plausible answer from those possible. The answering strategies are learned from known question-and-answer pairs using deep learning and natural language processing in order to achieve the greatest possible flexibility.

Yet for what feels like completely natural interaction with humans, an intelligent computer program must be able to engage in complete dialogues. Unlike the pure question-answering situation, the system must be able to follow the dialogue status and work towards the required goal, which makes the challenges significantly more difficult.

3.4.4 Intelligent agents

If the automation of a computer program is so far advanced that there is a high degree of autonomy in the decision, the term software agents or bots is used to describe this. A software agent or intelligent agent is a computer program which interacts in a complex environment, makes autonomous decisions on the behalf of the user and pursues a goal. An agent for the shares market, for example, could make autonomous investment decisions or deal with other routine tasks.

The benefit of an agent obviously depends on the complexity of its world understanding, the intelligence with which it makes decisions and its flexibility in adapting to changing situations. While the use of agents in operative applications is still very limited, new developments in computer games provide an indication of the prospects. Examples such as AlphaGo from Google,[40] which beat Lee Sedol, the human "Go" master, or the genetic fuzzy system from Psibernetix[41] that beat air force pilots in a simulated air battle in 2016, are indicators of what is already possible in terms of technology. A type of behaviour distinct from pure rote learning could be observed both with AlphaGo and the fuzzy system from Psibernetix.

---

[39] Ferrucci et al. (2010). Building Watson: An Overview of the DeepQA Project. In: AI Magazine, 31, pp. 59-79.

[40] Silver et al., 2016, Mastering the game of Go with deep neural networks and tree search. In: Nature, 529, pp. 484–489.

[41] Ernest et al., 2016, Genetic Fuzzy based Artificial Intelligence for Unmanned Combat Aerial. In: Journal of Defense Management, p. 6.

A transition to intelligent agents can also be seen in the combination with robotic process automation (RPA). The use of machine learning in RPA pledges to replace or supplement humans in three of their core abilities: communication with the customer – through chatbots, for example – extraction of relevant data and decision-making. The new possibilities of RPA are based on services which can understand the information relevant for the process and make it useful – a task for text mining in particular. In complex documents, embedded illustrations and tables are also relevant, alongside the text as the main data type. In the case of extraction of information, pretrained models for text comprehension or image analysis can be used for the automatic processing of damage claims to insurance companies, allowing the relevant information to be extracted for the next processing steps, for example.

3.5 Technical approaches as a solution to the social demands from big data analytics

The approaches to BDAI presented in the previous sections represent the status quo of the practical use of such methods. As machine learning methods are being used more comprehensively in applications which are increasingly critical, the legal and ethical requirements for such use as well as public notice of problematic cases also increase. The following sections present selected research approaches which take particular account of the extent to which the models can be further developed and supplemented under regulatory, ethical and legal aspects. Current research in the field of machine learning focuses particularly on intelligibility, data protection, data sovereignty and the prevention of discrimination in automatic decisions.

3.5.1 Intelligibility of models

The challenge that methods of machine learning in general and deep learning in particular face is that they are unintelligible to the user in different ways compared with simpler applications. They are seen as black boxes, i.e. the user has no direct possibility of finding out why or how the algorithm has made its decision and thus produced the result.[42] This is mainly due to the fact that the input and output values of neural networks are linked in a very non-linear and complex way.

The importance of the intelligibility of the results compared with other requirements, e.g. the statistically determined quality of the model, is not a question of technology. Rather, it must be decided in the individual case. A better understanding of models provides an opportunity to improve the analysis process – for instance, the users are then in a position to recognise problems such as overfitting and data bias. Better intelligibility can raise later acceptance of the results, particularly when legal or safety-related applications are involved. Intelligibility affects several levels: it helps the data scientist to adjust and improve models; it helps the user to evaluate the relevance of the results; and it helps decision-makers and regulatory bodies to evaluate the ability of learning systems to be certified and approved for critical applications.

3.5.2 Transparency and explainability

As far as the intelligibility of AI systems[43] is concerned, a distinction is made between transparency and explainability. Transparency describes the behaviour of the system being completely comprehensible. However, this requirement can almost never be fulfilled since many models are necessarily complex. In contrast, explainability describes being able to list the major factors of influence for a concrete individual decision. This is much easier to fulfil in terms of technology.

---

[42] Beuth 2017, Die rätselhafte Gedankenwelt eines Computers, Die Zeit. Available online: http://www.zeit.de/digital/internet/2017-03/kuenstliche-intelligenz-black-box-transparenz-fraunhofer-hhi-darpa, accessed on 20 February 2018.

[43] Doshi-Velez et al., 2017, Accountability of AI under the Law: The Role of Explanation. Available online: https://arxiv.org/abs/1711.01134, accessed on 20 February 2018.

There are two main approaches for the creation of transparent systems:

1. On the one hand, an attempt can be made to generate comprehensible approximations for any models. This approximation should be characterised by being able to reproduce the behaviour of the original model with as little deviation as possible, despite being able to be depicted in a way that humans can understand. The TREPAN algorithm[44] is one of the first such approaches.

2. On the other hand, there are many approaches for the generation of directly comprehensible models from data, such as decision trees[45] or subgroups[46]. Such models are not always ideal for the learning task, yet if transparency is more important than the quality of the result, these methods should be preferred.

The generation of explanations is currently a highly active field of research in machine learning[47]. One of the most famous approaches is the LIME algorithm[48], which uses simpler methods to form a local explanation model for the individual case to be explained and similar data points. It makes the availability of features necessary that permit interpretation by humans. For this reason it can often be essential during the reproduction of results of the methods to abstract from the data to a level comprehensible to humans. As an example: Sections of images can be summarised in superpixels (connected clusters of pixels) so that their content can be made accessible to humans as new features. Other approaches are prototype-based and deliver representative individual cases to explain decisions[49].

Overall, new approaches could provide at least some insight into how the models work and the reasons for decisions, even with very complex models, thereby preventing models from being seen purely as black boxes.

3.5.3 Privacy-preserving data mining

Protection of both personal data and company data forms the scope of action for all data use. Alongside legal and organisational measures which are designed to ensure sufficient data protection, suitable technical measures and IT infrastructures exist to simplify and support this process. For one, the principle of privacy-by-design exists for technical systems, i.e. the requirement to design the system in such a way that no information that is problematic in terms of data protection can be processed at all. As a concrete example, it is conceivable not to allow any plain text input in data input masks for customer data, since checking the contents in terms of data privacy during ongoing operation would involve enormous resources. This approach becomes problematic through the fact that the combination of relatively simple information can often lead to the identification of individual persons. It has been demonstrated, for instance, that 87 percent of the US population are already uniquely identified through postcode, gender and date of birth[50].

---

[44] Craven et al., 1996, Extracting tree-structured representations of trained networks. In: Advances in Neural Processing Systems, volume 8, pp. 24-30.

[45] Quinlan 1993, C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, Inc.

[46] Grosskreutz et al., 2012, An Enhanced Relevance Criterion for More Concise Supervised Pattern Discovery. In: Proceedings of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2012), Beijing, China.

[47] Montavon et al., 2017, Methods for Interpreting and Understanding Deep Neural Networks. In: Digital Signal Processing 73, pp. 1-15.

[48] Ribeiro et al., 2016, "Why should I trust you?": Explaining the Predictions of Any Classifier. Proc. of the 22nd International Conference of Knowledge Discovery and Data Mining, p. 1135-1144.

[49] Been et al., 2016, Examples are not Enough, Learn to Criticize! Criticism for Interpretability. In: Neural Information Processing Systems 2016.

[50] Sweeney, 2000, Simple Demographics Often Identify People Uniquely. In: Carnegie Mellon University, Data Privacy Working Paper 3.

Driven by increasing requirements and challenges in data protection, techniques are also being researched under the umbrella term "privacy-preserving data mining"[51][52] to integrate data protection requirements directly in data analysis. In other words, this field of research provides answers to questions such as "Which data and patterns can be published without reservations?" and "How can a certain question be analysed in such a way that no sensitive information is disclosed?"

The main idea behind privacy-preserving data mining is that identifying features – such as the name and address of an insured person – are not usually of interest for the analysis itself. On the contrary, in order to achieve generally valid knowledge, the patterns found in the data must be more general than would be necessary for the identification of an individual. The situation is further complicated by the fact that combining certain features could possibly lead to conclusions about the individual. In particular, care must be taken to ensure that analysis results do not allow conclusions to be drawn about sensitive information such as illnesses or the ethnicity of individuals.

Two of the main approaches of privacy-preserving data mining are anonymisation and secure distributed computation. Anonymisation attempts to suppress critical information as data are accessed, which makes the approach very generally applicable but means the risk of a significant loss of quality for the results. A standard approach to anonymisation is k-anonymity. Here, the data are processed in such a way that there is no query that applies to fewer than k database entries. Another approach which ensures privacy by randomising data analysis is differential privacy[53]. Secure distributed analysis aims to avoid information leaks during execution of data analysis methods on the entire data. Furthermore, the analysis of distributed data using cryptographic methods is carried out in such a way that the participants do not receive any information other than the required result.

All in all it must be stated that, based on currently available scientific knowledge, the possibility of carrying out privacy-preserving data mining must always be proven on a case-by-case basis. There is no automatic approach which guarantees data privacy generally and automatically for all types of analyses. Nevertheless, literature includes numerous approaches which can be considered for the analysis of sensitive data.

3.5.4 Infrastructures for data sovereignty

Companies in the financial sector not only have to check their own data processing and guarantee data protection, they also have to control data exchange and the joint cultivation of sometimes-sensitive data within the business ecosystem. Since data exchange is not only necessary for running a business but also promotes innovations and new business models, data is increasingly being used as a strategic resource for business success and thus becoming more valuable. The more valuable data becomes, however, the greater the desire to protect and control it – which in turn is not consistent with data exchange.

For this reason, data sovereignty is increasingly becoming a key skill for the financial sector.[54] It is defined as the "ability of a natural or juristic person for exclusive self-determination with regard to the commodity data"[55]. This means that the owner of the data must decide, control and check in a sovereign way what happens to his data, who receives it and what it is used for. In addition to economic and legal general conditions, IT methods in particular are required to create the prerequisites for data sovereignty.

---

[51] Grosskreutz et al., 2010, Privacy-preserving data-mining. In: Informatik-Spektrum, 33, p. 380-383.

[52] Aggarwal et al, 2008, Privacy-Preserving Data Mining: Models and Algorithms. In: Springer.

[53] Dwork, 2006, Differential Privacy. 33rd International Colloquium on Automata, Languages and Programming, (ICALP 2006). Springer, pp. 1–12.

[54] Chapter Fostering customer and consumer trust documents data sovereignty from the perspective of consumers

[55] Otto, 2016, Digitale Souveränität: Beitrag des Industrial Data Space. Fraunhofer.

One reference project is the joint initiative launched by business, politics and research at the end of 2014, "Industrial Data Space". It aims to establish the development and use of industrial data space at national, European and international levels. Industrial data space can be seen as an alternative and supplementary architectural draft to existing concepts which are either characterised by data being managed completely centrally and correspondingly monopolistically, or by the fact that every single exchange of data must be negotiated separately.

Industrial data space provides comprehensive mechanisms that protect the confidentiality, integrity and availability of the data during exchange. The federal structure of the industrial data space guarantees the data owner's complete sovereignty over their data. If required, the data can remain with its owner (decentralised) and is only exchanged within the context of agreed usage and under strict protective measures. Centralised data management, which usually goes hand in hand with a loss of control over one's own data, thus becomes optional. This allows even critical data of strategic and economic value to be exchanged safely and sovereignly and used jointly in the business ecosystem.

The decentralised architecture of the industrial data space, certification of participants, clearly defined rules and use of innovative safety technologies to check data flow and data use aim to create trust and allow companies to remain masters of their own data. Industrial data space thus not only increases the readiness of companies in general to exchange data, it promotes the safe exchange in particular of sensitive data, which was seldom shared before, thus making it available for data-driven innovations for the first time.

3.5.5 Non-discriminating data analysis

There are alarming examples that underline the limitations of current systems: Google classified dark-skinned humans as gorillas on some photos[56] and Microsoft had to withdraw its chatbot Tay from Twitter after just a few hours because it had learned racist behaviour from other users[57]. The examples illustrate that bias in the base data has to be dealt with and that AI systems have to behave ethically.

The result of a system based on machine learning can be unethical and discriminatory for several reasons: either because developers or users have taught the machine this behaviour consciously or unconsciously or not prevented it – as was the case with the chatbot Tay – or because the system has learned and adopted prejudices present in the data. Thus, for example, it was demonstrated in a case of word correlations gained automatically from large quantities of text taken from the internet that men are often mentioned in the context of programming and women in the context of housework in texts related to professions[58]. In the first case unethical behaviour is integrated in the analysis process, in the second case the ethical prejudices or stereotypes implicitly present in the data were only exposed through the analysis process.

---

[56] Zhang 2015, Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software, In: Forbes. Available online: https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/, accessed on 20 February 2018.

[57] Hunte, 2016, Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter, In: The Guardian. Available online: https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter, accessed on 20 February 2018.

[58] Bolikbasi et al. (2016). Man is to Computer Programme as Woman is to Homemaker? Debiasing Word Embeddings. In: Neural Information Processing Systems 29.

Accordingly, methods of non-discriminating data analysis can pursue different technical approaches[59]. The data can be cleared of critical patterns before analysis – by ensuring gender, skin colour and age are well distributed and represented in the data for face recognition, for example – or the methods themselves can contain ethical restrictions with regard to the permitted solutions. Another possibility is to revise the outputs by removing critical patterns, for example, and to manually improve forecasts for disadvantaged groups.

The technical challenge in each case is to transform the ethical/legal definition of discrimination into a mathematical one so that it can be monitored by algorithm and prevented. If the objective is to avoid discrimination in insurance premiums that are based on gender, the general condition that the same number of positive decisions are taken for men and women can be added. In contrast, the pure deletion of the gender feature will not be sufficient, since in many cases gender can be deduced with sufficient accuracy through correlating features such as certain professions.

In conclusion, it can be stated that there is no currently accepted standard for non-discriminating data analysis, although there are very many approaches and research activities focussing on this issue. The problems and possible approaches to a solution can thus be assumed as known for the definition and implementation of analysis projects.

---

[59] Custers et al., 2013, Discrimination and Privacy in the Information Society – Data Mining and Profiling in Large Databases. Springer.

IV. Strategic prerequisites

4.1 Fostering customer and consumer trust[60]

4.1.1 The customer perspective

BDAI can be harnessed to develop process and product innovations and therefore to create individualised products and services for the customer. This requires sufficiently large and valid data sets – customers are often requested to provide this data themselves. The consumer is not simply the customer for BDAI products but indeed also an important supplier of data and it is not always entirely clear to the consumers whether and how their data is used. It is most likely that the asymmetries of power and information already discernible between consumers and companies will continue to grow. In this respect at least consumers appear to be the potentially weaker, structurally inferior market players.

If these asymmetries continue to grow, collective consumer interests may be indirectly affected – something that might undermine trust in financial institutions and insurance companies and, as a corollary, the integrity of the financial market. It is therefore important to take genuine collective consumer interests into account when implementing BDAI.

Two conflicting consumer interests

To put it simply, consumer interest in BDAI-based products and services can be divided into two categories. On the one hand, the consumer is interested in the direct and specific value added of a BDAI product. On the other hand, consumers are interested in preserving their privacy, their right to self-determination, their freedom to make their own decisions and their autonomy[61] – both in and outside of the financial market.

These two consumer interests could become conflicting interests: The value of privacy, self-determination, freedom of decision-making and autonomy is not immediately obvious and appears to be of secondary concern for many consumers. By contrast, the product benefits are immediately clear. Therefore, consumers in the European legal setting tend to grant the required consent for the use of their data very quickly – often due to a lack of alternatives. However, if data is then used or passed on in an excessive manner and in a way the users would not have expected when they granted their consent, their interest in privacy, data protection, self-determination and autonomy can again rise to the fore. Even where data protection requirements are adhered to, any data usage that is perceived by consumers as excessive could potentially erode trust.[62]

---

[60] The terms "consumers" and "customers" are used synonymously in this chapter. "Customers" refers both to individual customers as well as freelancers and small businesses.

[61] See Hoffmann-Riem, The power of digital corporations, 1 October 2017. Available online: https://www.law-school.de/fileadmin/content/law-school.de/de/units/unit_affil_riem/pdf/418_Die_Macht_digitaler_Konzerne.pdf, pp. 2 ff., last accessed 12 April 2018.

[62] Schäfer 2018, Facebook Sturz erinnert an die Lehman-Pleite, Süddeutsche Zeitung. Available online: http://www.sueddeutsche.de/digital/datenschutz-facebooks-sturz-erinnert-an-die-lehman-pleite, accessed on 20 April 2018.

Consumers have difficulties foreseeing the consequences of sharing their data

Companies are particularly interested in consumers' financial and economic data because it reveals a person's economic core (income, financial status, spending behaviour, contractual obligations, state of health, etc.) as well as allowing companies to analyse personalities and habits. This data can then be used in the interests of the consumers, for instance to provide more tailored and individualised offers. However, it can also be used to fully exploit what consumers are willing to pay as explained below. Consumers therefore find it difficult to estimate the long-term consequences of sharing data for BDAI applications. Indeed, it is often difficult for them to weigh up the long-term consequences of sharing data with the benefits the product offers in the short term. In such cases, there is an asymmetry of information between a data supplier and a data user regarding the actual consequences of data sharing as well as the value of data.

Companies with data-based business models active in sectors other than the financial industry (e.g. bigtechs) may indeed be very interested in offering very reasonably priced financial products by means of cross-subsidisation in order to gain access to valuable financial transaction and behaviour data of consumers. These companies could then monetise this information and use it to boost earnings in their core business.

Companies could exploit the consumers' willingness to pay

Companies can use BDAI applications to link financial data with other data they have at their disposal to gain hitherto impossible, deep insights into consumer habits and personality structures. This knowledge, in some cases very private knowledge, can also be used to the detriment of consumer interests. From an economic point of view, it is in the interest of the consumers' privacy that providers do not gain access to information regarding their willingness and ability to pay. This is still the case even when consumers feel they have "nothing to hide". Otherwise, the danger of consumers buying overpriced products is real because companies using BDAI are able to distil in-depth knowledge of the consumers' willingness to pay for broad consumer groups – either because they have access to the data themselves or by buying the information they need. They can then harness this knowledge to increase earnings or optimise risk. This is not merely about paying higher prices for better or more tailored services, for example paying higher insurance premiums to hedge higher risks. It is also about how companies can potentially extract the consumer surplus[63]. BDAI applications can indeed make it easier to design inexpensive (mass) individualised products and services. This is how individualised components can be added to standard products – at no significant cost to the company. It becomes much more difficult for consumers to compare and switch to other products. Therefore, what they pay is not the market price but the price tailored to the individual customer's willingness to pay. However, if companies remain unaware of a consumer's ability and willingness to pay, then consumers can realise economic welfare effects and pay the usual market price.

---

[63] Consumer surplus is the difference between the maximum price that a consumer is willing to pay for a product or service and the price that he/she actually has to pay on the market. By employing inexpensive (mass) individualisation, the use of BDAI may make it possible for providers of (digital) products and services to ask the maximum price that each individual customer is willing to pay. This effectively allows the providers to extract the consumer surplus. However, this assumes that the providers know how much the customer is willing to pay. Using BDAI, it is at least hypothetically possible that the necessary knowledge could be generated and used. It would, however, be very difficult for the customer to detect the (hidden) price adjustment due to the (mass) individualisation of the products and services.

Avoid discrimination and the pressure to share data

In addition to the consumers' willingness to pay, BDAI applications also make it possible to determine other personal characteristics, such as health status and emerging life changes, which could contribute to (implicit) discrimination of the consumer. For instance, algorithms can (implicitly) focus on characteristics without these being directly collected. So, for example, if financial transaction patterns reveal a change in (purchasing) behaviour, it is possible to deduce upcoming life events (e.g. separations, pregnancies) without having collected any specific data. In these cases consumers cannot make a connection between any discrimination and the data they shared and therefore cannot contest the discrimination. By using the appropriate algorithms, companies can (even inadvertently) discriminate against consumers – which is particularly controversial if consumers are then denied financial products. Deliberate discrimination would be, for instance, setting the prices so high that they are beyond the ability and willingness of specific consumers to pay. It would be problematic if customers were unable to detect and/or influence exclusion characteristics to be able to gain access to affordable financial services. This behaviour could have supervisory or regulatory implications if broad-scale use of such technologies meant that whole consumer groups have only very limited, overpriced or indeed no access at all to certain financial services.

Furthermore, broad-scale use of BDAI could practically force consumers to use digital offers and share data. This would be contrary to the interests of those consumers in particular who do not wish this type of product, who do not wish such extensive exposure of their personal data or indeed who are unable to provide any data. If non-BDAI-based products became significantly more expensive, this might mean practical exclusion from key financial products and/or discrimination. Actual pressure to share data is also not the best way to boost consumer trust in new technologies such as artificial intelligence. Whole groups of customers that are not so keen on digitalisation or who object to revealing large amounts of data could thus see their interests violated.

"Paying with data" exacerbates the problem

Using BDAI in the context of "paying with data" business models in particular means that the consumer is often the supplier of data but has limited understanding of the value of that data and what additional insights might be derived from it. In addition, consumers wrongly perceive these offers to be free although "there is no such thing as a free lunch" in a market economy! Many consumers remain unaware that they are paying with their data or they do not know the actual price they are paying. This lack of transparency fuels the above-mentioned power and information asymmetries between companies and consumers.

Correctly securing consent and creating data sovereignty

European data protection regulations basically outline that any use of personal data is subject to the consent of the user, which makes consent the critical issue in data protection. If declarations of consent are broadly phrased, they basically become authorisations with which companies can expand their frame of action – at the expense of the users.[64] In industries other than the financial sector, such broad declarations of consent are often prerequisite for buying or using a product or service, for example the rights to use news service images, provided the images were sent via this service.

---

[64] See Hoffmann-Riem, The power of digital corporations, 1 October 2017. Available online: https://www.law-school.de/fileadmin/content/law-school.de/de/units/unit_affil_riem/pdf/418_Die_Macht_digitaler_Konzerne.pdf, p. 12, last accessed 12 April 2018.

Consent to the use of data and the conditions of consent are therefore of central importance in safeguarding consumer interests. Providers must bear in mind that, when giving their consent, consumers are less aware of data protection and more conscious of the spontaneous benefits of the product. This often means that consumers blindly grant their consent for reasons of convenience and speed. Market players should not exploit this weakness. In any case, it is indeed critical that consumers only agree to extensive data usage because they do not want to read what are usually complicated declarations of consent because they feel the social pressure or because they see no alternative if they want to use the product at all.

4.1.2   Relevance of consumer trust in the context of financial services

Data-driven applications can offer great benefits to consumers. At the same time, consumers are sensitised to the use of data by companies.

As described above, BDAI can increase the individual benefits of products and services to customers. Consumer and behavioural data can be used, for example, to provide personalised offers. It should be noted that some consumers provide their data for a specific purpose in order to take advantage of certain services and the perceived benefit thereof. Customers are becoming more and more used to the additional benefits resulting from such services, such as the personalised form of address. It is possible that, in the future, customers will have similar expectations of offers from financial institutions and insurance companies.

Products and services based on BDAI are, by definition, dependent on sufficient data availability. To use certain services or products (for example private health insurance), providing personal data is unavoidable. Other applications, on the other hand, are less reliant on data that is tied to a specific person. In these cases, sufficiently anonymised or pseudonymous data could be used. Appropriate methods such as privacy-preserving data mining are described in chapter 3.5 of this study.[65] In addition, providers can also use innovative technologies, including distributed ledger or block chain technologies, to better protect personal data when using BDAI innovations.

Surveys show that consumers have concerns about companies using their personal data. According to a Eurobarometer survey, only 15 percent of European consumers feel they have complete control over the information they provide online.[66] 67 percent of those surveyed are concerned about not having complete control over the information they provide online. Furthermore, 52 percent of those surveyed[67] disagreed with the statement, "You don't mind providing personal information in return for free services online (e.g. free email address)."[68] Almost 70 percent of those surveyed also claim to be concerned about their information being used for a purpose other than the one for which it was collected.[69]

---

[65] See Marnau, 2016, Anonymisierung, Pseudonymisierung und Transparenz für Big Data. In: Datenschutz und Datensicherheit, p. 428–433 (also for a discussion of the methods used in the context of the EU GDPR).

[66] European Commission, 2015, Factsheet Data Protection Eurobarometer. Available online: http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_eurobarometer_240615_en.pdf, accessed on 22 January 2018.

[67] 29 percent: "Agree"; 15 percent: "Not applicable"; 4 percent: "Don't know".
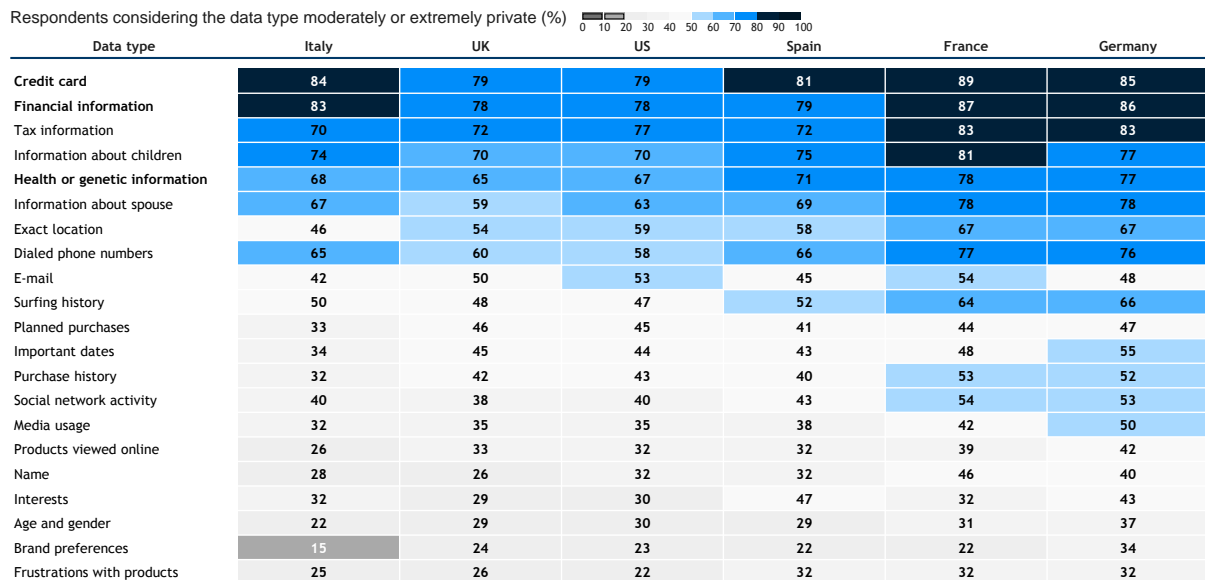
[68] European Commission, 2015, Factsheet Data Protection Eurobarometer. Available online: http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_eurobarometer_240615_en.pdf, accessed on 22 January 2018.

[69] Ibid.

However, there is a discrepancy between many consumers' actual behaviour and the concerns voiced in surveys regarding the use of data by online services. For instance, the number of consumers that make use of such services is growing constantly.[70] How this discrepancy can be explained and why personal data (e.g. in social networks) is provided on a large scale despite general concerns is the subject of scientific and social debate.[71] Due to the particular sensitivity of financial data, it is critical that careful study is made of the extent to which the behavioural patterns observed in online services and in social networks can be applied to the financial context.

Overall, it can be seen that consumers are now at least aware of how companies use their data. Consumers show a particularly high degree of sensitivity when it comes to the use of financial and health data. Both in the European Union's major economies and in the US, these types of data are considered particularly private (see Figure 6). Financial service providers must take this particular sensitivity into account when designing BDAI offerings. Any possible misuse of data – the customers' subjective perception is sufficient in that regard – could lead to a loss of trust.

Figure 6: An international comparison of sensitivity regarding data types[72]

Respondents considering the data type moderately or extremely private (%)

0 10 20 30 40 50 60 70 80 90 100

Data type	Italy	UK	US	Spain	France	Germany
Credit card	84	79	79	81	89	85
Financial information	83	78	78	79	87	86
Tax information	70	72	77	72	83	83
Information about children	74	70	70	75	81	77
Health or genetic information	68	65	67	71	78	77
Information about spouse	67	59	63	69	78	78
Exact location	46	54	59	58	67	67
Dialed phone numbers	65	60	58	66	77	76
E-mail	42	50	53	45	54	48
Surfing history	50	48	47	52	64	66
Planned purchases	33	46	45	41	44	47
Important dates	34	45	44	43	48	55
Purchase history	32	42	43	40	53	52
Social network activity	40	38	40	43	54	53
Media usage	32	35	35	38	42	50
Products viewed online	26	33	32	32	39	42
Name	28	26	32	32	46	40
Interests	32	29	30	47	32	43
Age and gender	22	29	30	29	31	37
Brand preferences	15	24	23	22	22	34
Frustrations with products	25	26	22	32	32	32

Question: "How private do you consider the following types of data?" Answer on a scale of 1 to 5, where 1 means "Not at all private" and 5 means "Extremely private".

---

[70] The number of daily active Facebook users grew between Q1 2011 and Q4 2017 from 372 million to 1.4 billion (+277 percent). Source: Statista, 2018, Number of daily active Facebook users worldwide as of 4th quarter 2017 (in millions). Available online: https://www.statista.com/statistics/346167/facebook-global-dau/, accessed on 16 February 2018.

[71] A variety of psychological, sociological, situational and economic factors may be considered. See Kokolakis, 2015, Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. In: Computers & Security. Available online: https://www.researchgate.net/publication/280244291_Privacy_attitudes_and_privacy_behaviour_A_review_of_current_research_on _the_privacy_paradox_phenomenon, accessed on: 19 February 2018. Among other things, the desire for recognition and participation in the (virtual) community, habituation and habit effects and the desire for convenience and speed are mentioned. Also, the essential benefits of privacy and the potential consequences of data provision may be displaced, forgotten, or unknown due to informational asymmetries. Eventually, a perceived absence of freedom of choice could play a role.

[72] BCG Big Data and Trust Consumer Survey 2015 (survey of over 8,000 consumers in Italy, the UK, the USA, Spain, France and Germany). Source: Rose et al., 2016, Bridging the Trust Gap: Data Misuse and Stewardship by the Numbers. Available online: https://www.bcg.com/de-de/publications/2016/big-data-technology-digital-bridging-trust-gap-data-misuse.aspx, accessed on 21 February 2018.

Established financial institutions enjoy a high level of customer and consumer trust. Data misuse can therefore have negative effects on customer relationships and, depending on the extent of the misuse, even on trust in the financial services sector as a whole.

A comparison shows that the financial services sector enjoys high trust levels among consumers when it comes to the treatment of personal data (see Figure 7). 56 percent of European consumers trust banks and financial institutions to protect their personal information.

Figure 7: Trust in the protection of personal information by public authorities and private companies[73]
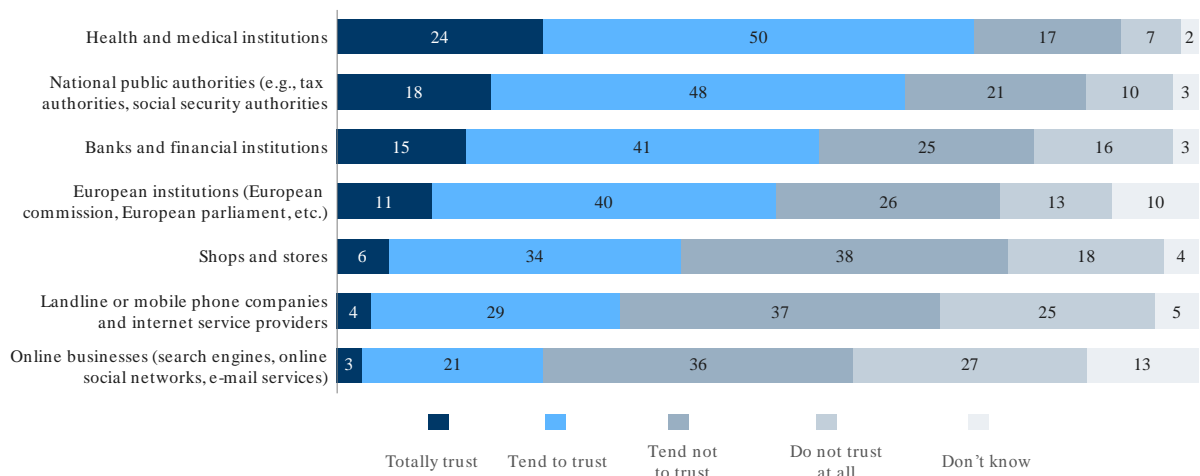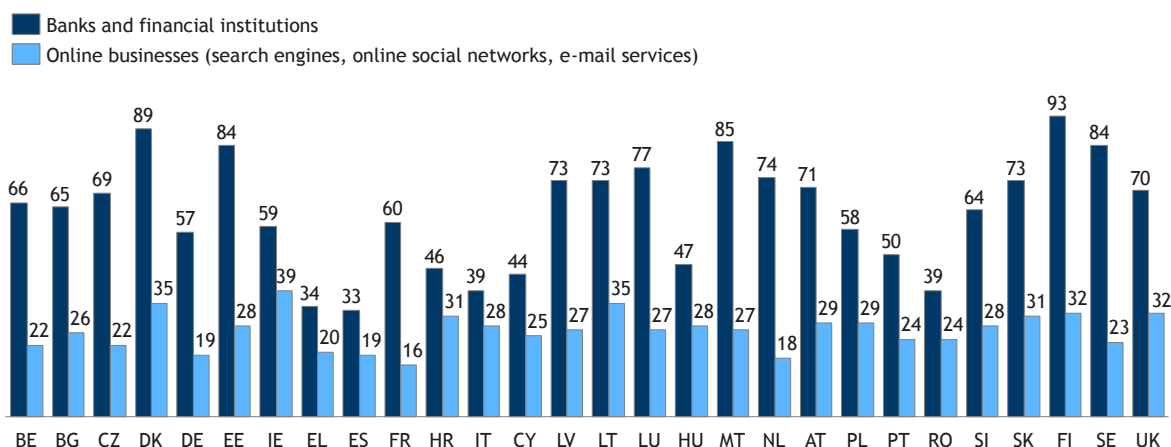

Survey question: "Different authorities (government departments, local authorities, agencies) and private companies collect and store personal information about you. To what extent do you trust the following authorities and private companies to protect your personal information?"

A cross-country comparison gives a more differentiated picture (see Figure 8). Compared to the average across the EU 28, it becomes obvious that those surveyed in Germany, France, the Netherlands and Spain particularly lack the confidence that online businesses will look after their personal data.[74]

---

[73] European Commission, 2015, Special Eurobarometer 431.

[74] A survey by market research institute GfK, on behalf of the Association of German Banks, came to a similar conclusion. See Association of German Banks, 2017, Bank der Zukunft: Die Kunden auf dem Weg zur Digitalisierung mitnehmen! Available online: https://bankenverband.de/newsroom/meinungsumfragen/bank-der-zukunft-die-kunden-auf-dem-weg-zur-digitalisierung-mitnehmen/, accessed on 15 November 2018.

Figure 8: A comparison of trust in financial service providers and online businesses[75]

■ Banks and financial institutions
■ Online businesses (search engines, online social networks, e-mail services)

	BE	BG	CZ	DK	DE	EE	IE	EL	ES	FR	HR	IT	CY	LV	LT	LU	HU	MT	NL	AT	PL	PT	RO	SI	SK	FI	SE	UK
Banks	66	65	69	89	57	84	59	34	33	60	46	39	44	73	73	77	47	85	74	71	58	50	39	64	73	93	84	70
Online	22	26	22	35	19	28	39	20	19	16	31	28	25	27	35	27	28	27	18	29	29	24	24	28	31	32	23	32

Survey question: "Different authorities (government departments, local authorities, agencies) and private companies collect and store personal information about you. To what extent do you trust the following authorities and private companies to protect your personal information?"

Customer and consumer trust in financial institutions and insurance companies has grown historically. Due to the high level of consumer sensitivity with regard to personal financial data, trust in financial institutions has always been a central element for business development (e.g. banking secrecy). The protection of personal data was and is a central management task and obligation in this context. The importance of data protection in a company is underlined by the fact that most companies are required by law to appoint a data protection officer who, in turn, has special privileges, such as special protection against dismissal. The regulation and supervision of financial services has also boosted public confidence in the market and in the business practices of financial institutions.

In principle, a distinction can be made between two risks with regard to customer and consumer trust. On the one hand, data misuse from the consumer's point of view – in other words, a perceived misuse of data that, depending on the situation, could be considered a violation of privacy – can damage consumer trust in the use of data. On the other hand, trust can also be damaged by data theft or breaches, in other words insufficient information security or a lack thereof. This can include incidents in which unauthorised persons gain access to consumers' personal data – either by means of a deliberate attack from the outside or through mistakes made by the company (see Chapter 4.2).

Perceived misuse of data does not specifically refer to a violation of data use agreements with the financial services provider, nor is it particularly relevant whether the misuse of data really causes economic damage to the consumer. What is important is that the consumer perceives – even if it is only subjectively – that the service provider is using the data in a way that does not comply with the purpose for which it was collected as understood by the consumer.[76]

Furthermore, data theft and breaches present a risk to customer relationships as they may shake customers' trust in the security of their data. As the abnormal churn rate shows,[77] negative effects on customer relationships can be particularly high in the financial services and health sectors. (See Figure 9).
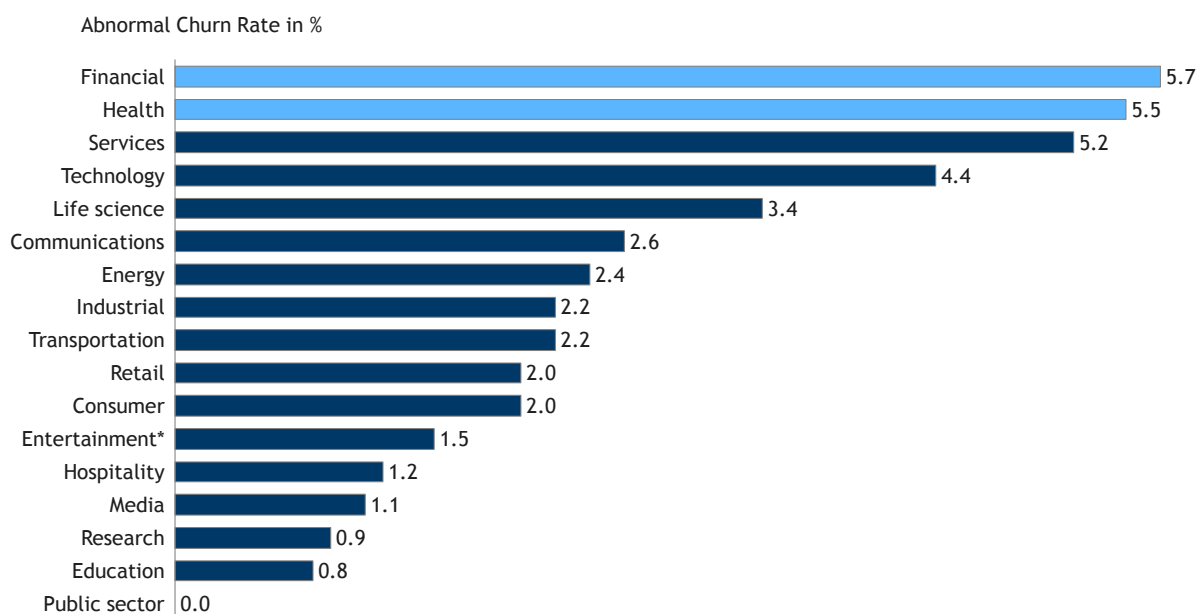
---

[75] European Commission, 2015, Special Eurobarometer 431.

[76] See Rose et al., 2016, Bridging the Trust Gap: Data Misuse and Stewardship by the Numbers. Available online: https://www.bcg.com/de-de/publications/2016/big-data-technology-digital-bridging-trust-gap-data-misuse.aspx, accessed on 21 February 2018.

[77] Loss of customers due to a data theft or breach can be measured by an abnormal churn rate. This reflects the increase in expected customer losses compared to the scenario without a data theft/breach.

Abnormal Churn Rate in %

Sector	Rate
Financial	5.7
Health	5.5
Services	5.2
Technology	4.4
Life science	3.4
Communications	2.6
Energy	2.4
Industrial	2.2
Transportation	2.2
Retail	2.0
Consumer	2.0
Entertainment*	1.5
Hospitality	1.2
Media	1.1
Research	0.9
Education	0.8
Public sector	0.0

The EU GDPR is intended to strengthen data and consumer protection in the new data era.

The aim of the EU's General Data Protection Regulation (EU GDPR) is to strengthen consumer rights and data protection. It came into force in May 2018. The EU GDPR introduces some fundamental changes to data protection regulation and a large number of new requirements for companies. Depending on the extent to which companies need to adapt, the EU GDPR may also have wide-ranging implications for companies' business and IT practices. The new regulation is directly applicable in every EU member state. The central aims of the EU GDPR include consistent EU-wide handling of personal data, strong privacy protection with clearly defined responsibilities in the event of violations and more rights and self-determination for consumers.

One central element of the EU GDPR intended to strengthen consumer rights is the requirement of explicit consent for the use of personal data. This consent is a legitimacy criterion and is subject to certain conditions[79]. In accordance with recital 32 of the EU GDPR, "Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data [...]".[80] Furthermore, consent is not freely given if the consumer "[...] has no genuine or free choice or is unable to refuse or withdraw consent without detriment."[81] According to Article 12 of the EU GDPR, information and communications from companies to consumers, including with regard to collecting personal data, should be "[...] in a concise, transparent, intelligible and easily accessible form, using clear and plain language [...]".[82] In addition, the principle of transparency should apply. This requires that "[...] additionally, where appropriate, visualisation be used. [...] This is of particular relevance in situations where the proliferation of actors and the technological complexity

---

[78] Ponemon Institute, 2017, 2017 Cost of Data Breach Study, p. 20. Available online: https://www.ponemon.org/library/2017-cost-of-data-breach-study-united-states, accessed on 29 November 2017. * Data from 2016 only.

[79] Art. 6 EU GDPR, Lawfulness of processing and Art. 7 EU GDPR, Conditions for consent.

[80] Recital 32 EU GDPR, Consent.

[81] Recital 42 EU GDPR, Burden of proof and requirements of consent.

[82] Art. 12 EU GDPR, Transparent information, communication and modalities for the exercise of the rights of the data subject.

of practice make it difficult for data subjects to know and understand whether, by whom and for what purpose personal data relating to them are being collected, such as in the case of online advertising."[83]

One additional condition for the explicit consent for the use of data arises from Article 7, Paragraph 4, where "package deals" are addressed with regard to data collection. Given the principle of free consent,[84] "[w]hen assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract."

4.1.3 Increase in complexity in the context of BDAI

BDAI applications increase the complexity of data usage, which can make it difficult for consumers to understand.

The increasing complexity and variety of data usage could make it more difficult for consumers to fully grasp how their data is being used. The disaggregation of value chains and outsourcing to third-party providers creates unclear data networks that consumers may find difficult to understand (see Chapter 4.1). This presents providers with new challenges when it comes to informing consumers about the use of data in a suitable way. If providers do not succeed in addressing this challenge, the knowledge asymmetry between consumers and the providers of corresponding offers who actually process the data could continue to increase.

As digitalisation progresses, it will also become increasingly difficult for consumers to understand what data about them is actually available and might potentially be used. Individuals can easily lose track of the entire data history of their digital consumer activity and how this data is linked with data that was not originally provided to the data processing company.

The strategic relevance of customer and consumer trust depends on the position and orientation of companies.

Proactively addressing these challenges can help maintain and even strengthen customer and consumer trust. Based on the market assessment in chapter V, three scenarios can be distinguished, each implying a specific approach with regard to customer and consumer trust.

1. Traditional financial service providers evolve using BDAI

Traditional financial service providers who already enjoy comparatively high levels of customer and consumer trust must safeguard against losing this trust-based advantage when evolving using BDAI (especially at the customer interface). Compared with providers who enjoy lesser levels of customer trust, any possible damage to a company's image stemming from data misuse could have more serious effects. Therefore, in addition to ensuring conformity with the EU GDPR, new innovative solutions for maintaining and increasing the levels of trust in data usage should be examined. Thereby, traditional financial service providers have the opportunity to leverage their trust advantage as a competitive edge. Existing high levels of trust could mean that consumers tend to be more willing to agree to the use of their personal data and information.

---

[83] Recital 58 EU GDPR, Principle of transparency.

[84] Recital 43 EU GDPR, Free consent: "Consent is presumed not to be freely given if it does not allow separate consent to be given to different personal data processing operations despite it being appropriate in the individual case, or if the performance of a contract, including the provision of a service, is dependent on the consent despite such consent not being necessary for such performance."

2. Traditional financial service providers moving towards a more data-driven BDAI business model (e.g. creation of a platform bank)

Existing levels of customer and consumer trust offer a significant opportunity to traditional financial service providers who are aiming for a primarily data-driven BDAI business model. There is, however, also an increased risk that this trust could be damaged if data is misused. The risk of a perceived misuse of data, for example, could increase as the complexity of data usage increases and the transparency thereof for the consumer decreases. Depending on both the significance of the financial service providers in the market and the frequency of possible data incidents, effects resulting from data misuse may also adversely affect the reputation of the entire sector.

3. Bigtechs offer services associated with traditional financial service providers

Should bigtechs start to offer financial services, they must examine how they can compensate for the fact that traditional financial service providers enjoy higher levels of trust. According to surveys, consumers – particularly in Germany – show limited willingness to carry out their banking with bigtech companies.[85] Furthermore, these new providers would have to fulfil the same supervisory and regulatory requirements as incumbents. However, as with established financial institutions, meeting these requirements could strengthen customer and consumer trust – provided the principle of corresponding limited-purpose data usage is established.

Regardless of their strategic focus, all providers within these three scenarios face the challenge of building sufficient customer and consumer trust in their use of data. A proactive approach to the EU GDPR which, in the interest of customers, goes beyond the regulated minimum standards and avoids legally grey areas, could prove a competitive advantage. Going forward, deliberately providing more data-economical or conventional offers besides data-intensive ones, could likewise result in a competitive advantage. In order to build trust in their data usage, companies could also voluntarily pledge to store or process personal data only for as long as is necessary to provide the product and then to delete the data completely.

4.1.4 Conclusion: ensure consumer sovereignty to build trust

New and established banks, insurance companies and other financial service providers need to ask themselves first and foremost which trust-building or trust-sustaining measures they should take considering BDAI and data usage as well as the fact that the regulatory and competitive environment has not yet been fully established. This is not only about trust as a value-adding factor but also about legal rights such as informational self-determination. Such rights are constitutionally protected in the Federal Republic of Germany[86] and also shape the normative order which influences legal relationships between private parties.[87] The EU GDPR also makes it clear that providers using BDAI are required to create adequate framework conditions for the use of BDAI governing how they relate to their contractual partners. The following paragraph outlines the data sovereignty approach by way of illustration.

---

[85] According to a survey carried out by the Association of German Banks, only 3 percent of those asked would be willing to switch their banking business to companies such as Google, Amazon or Facebook if they started offering banking services. Source: Association of German Banks, 2017, Bank der Zukunft: Die Kunden auf dem Weg zur Digitalisierung mitnehmen! Available online: https://bankenverband.de/newsroom/meinungsumfragen/bank-der-zukunft-die-kunden-auf-dem-weg-zur-digitalisierung-mitnehmen/, accessed on 17 November 2017.

[86] German Federal Constitutional Court (15 January 1958) - 1 BvR 400/57; BVerfGE.

[87] BeckOK DatenschutzR/Brink BDSG pp. 141f.

Data sovereignty can be understood as the individuals' capacity to maintain transparency and control over the possession, use and deletion of their personal data.[88] The aim of this approach is to allow consumers to exercise their potential individual responsibility in a sovereign manner. However, consumers can only make their own, well-informed decisions if they have transparent, clear and easy-to-understand information on BDAI data usage. It should also be ensured that consumers can revoke their consent to the use of their data for a specific purpose, that consumers are also informed of indirect consequences and that they have a genuine and free choice. Providers are responsible for ensuring these prerequisites. Thus, consumers can make free, informed, conscious and sustainable decisions regarding the use of BDAI and in future they will not have to rely solely on self-restrictions that companies impose on themselves when using consumer data.

Providing consumers with appropriate information about the use of data.

Privacy policies, particularly in their current widespread form in online services, are often not helpful in providing consumers with appropriate information. In reality, they are rarely ever read. For example, according to a survey conducted by The Boston Consulting Group (BCG), two-thirds of consumers say that they do not read privacy policies because they are too long and too complex. More than half of consumers surveyed in selected EU countries and in the US also say that data protection provisions contain too much "legalese" and are therefore difficult to follow.[89] A special edition of Eurobarometer on data protection also came to the conclusion that two-thirds of those surveyed (67 percent) consider privacy policies to be too long and 38 percent consider them to be too unclear and complex.[90] Providing privacy policies that are easy for the consumer to understand and adapted to the specific decision context could also help to address this problem. One possibility might be an abridged explanation, written in language geared towards the particular target group that would enable consumers to think through the consequences of providing data for that particular purpose. Financial service providers could develop their own solutions and thus help to ensure that consumers are better informed.

For example, in its report "Smart Data – Smart Privacy?", the FZI Research Center for Information Technology suggests that the results of the data protection impact assessment (see Art. 35 EU GDPR) should also be made available to consumers in simplified form to give them a basis for making decisions on the provision of data. Furthermore, the FZI remarks that transparency could be improved if an intuitive traffic light system were introduced that highlights the risks inherent in data usage.[91]

---

[88] The Bertelsmann Foundation, for example, presented a comprehensive policy framework for its data sovereignty concept in the context of big data. The framework integrates both the individual and the company and collective responsibility to ensure that control is maintained over the use of data. See Bertelsmann Foundation, 2017, Datensouveränität in Zeiten von Big Data, pp. 24–32. Available online: https://www.bertelsmann-stiftung.de/de/unsere-projekte/teilhabe-in-einer-digitalisierten-welt/projektnachrichten/datensouveraenitaet-in-zeiten-von-big-data/, accessed on 13 November 2017.

[89] See Rose et al., 2016, Bridging the Trust Gap: Data Misuse and Stewardship by the Numbers. Available online: https://www.bcg.com/de-de/publications/2016/big-data-technology-digital-bridging-trust-gap-data-misuse.aspx, accessed on 21 February 2018.

[90] European Commission, 2015, Special Eurobarometer 431.

[91] FZI Research Center for Information Technology, 2015, Smart Data – Smart Privacy? Impulse für eine interdisziplinär rechtlich-technische Evaluation, p. 13 f. Available online: https://www.digitale-technologien.de/DT/Redaktion/DE/Downloads/Publikation/SmartData_Thesenpapier_smart_Privacy.pdf?__blob=publicationFile&v=7, accessed on 8 January 2018.

Enable consumers to control the use of their data.

Germany's Advisory Council for Consumer Affairs also dealt with the topic of sovereignty in the context of digitalisation.[92] In addition to suggestions such as a one-page privacy policy, it also proposed creating a consumer-centred data portal. A portal of that kind could give consumers more control over the use (scope and content) of their individual data by the various providers. The aim is to allow consumers to centrally delete and change their data as well as to manage the access rights to their data centrally by using a single portal.

This is the approach taken for instance by VERIMI, a cooperation of various private entities in Germany.[93] [94] VERIMI aims to provide convenient, secure and independent handling of identities and data on the Internet[95]. The platform seeks to allow consumers to manage their digital identities and data securely for reuse by companies and public authorities on the Internet. For companies, the platform also intends to offer legal certainty within the context of the EU GDPR and to "securely place offers on local and global markets, independent of international platform providers".[96] [97]

4.2 Success factors concerning IT strategy

In addition to the success factor "consumer trust" described in chapter 4.1, there are other, more technological requirements for the successful use of BDAI. Financial service providers are already required to define a sustainable IT strategy, derived from the business strategy that defines both targets and the measures for achieving them[98]. In the BDAI context, IT strategies will, however, take on an even more important role once the changes in the market described in the next chapters come into effect. This is because BDAI is expected to have an impact on the data and IT architecture, (IT) organisational setup and processes, skills (at employee and management level) and working methods, information security and outsourcing. These technological success factors will be explained in more detail in the sections to follow.

---

[92] See Advisory Council for Consumer Affairs, 2017, Digitale Souveränität – Gutachten des Sachverständigenrats für Verbraucherfragen. Available online: http://www.svr-verbraucherfragen.de/wp-content/uploads/Gutachten_Digitale_Souver%C3%A4nit%C3%A4t_.pdf, accessed on 14 November 2017.

[93] Among other companies, the alliance includes Allianz, Axel Springer, Bundesdruckerei, Daimler, Deutsche Bank, Lufthansa and Deutsche Telekom.

[94] An additional platform is the "Log-in Alliance". Three media and Internet companies have come together to form an alliance and have developed a universal Internet-based log-in process (single sign-on). The companies' intention is to ensure that their platform is in compliance with the new European data protection provisions, which come into force in May 2018. Members of the log-in alliance include the RTL media group, ProSiebenSat.1 and United Internet (GMX, Web.de). See Berger, 2017, Medien-und Internet-Unternehmen gründen Log-in-Allianz. https://www.heise.de/newsticker/meldung/Medien-und-Internet-Unternehmen-gruenden-Log-in-Allianz-3784989.html, accessed on 12 February 2018.

[95] See VERIMI, 2017, VERIMI website. Available online: https://www.verimi.com/, accessed on 29 November 2017.

[96] Ibid.

[97] See Industrial Data Space as an approach to data sovereignty that allows data to remain at a decentralised location in the possession of its owners; see Chapter 3.5.4, Infrastructures for Data Sovereignty.

[98] For banks cf. §25a KWG, MaRisk, BAIT.

4.2.1 Data quality and scope, IT architecture and cloud computing as success factors for the use of BDAI

Data quality and scope

Data quality is intended to ensure that the input variables for a BDAI application represent a sensible and effective basis[99] for the ensuing results. To this end, processes and criteria need to be provided that can guarantee appropriate data quality for the respective use case. The specific criteria can vary depending on the use case. Typical criteria include completeness, consistency, validity and accuracy and/or timeliness.[100] Furthermore, the scope, i.e. the volume and variety of the data, can help to generate representative and meaningful BDAI models.

This is why (legal) requirements pertaining to the result of a data evaluation for compliance-related cases (e.g. in the prevention of money laundering) are always higher than those with regard to targeted customer contact (cf. use cases in Chapter 5.2). In terms of customer contact, the benefit of being able to contact more customers may outweigh the risk of contacting a few of the "wrong" customers. Hence, there are different criteria for evaluating data quality. In principle, market players have a responsibility to prevent the risk of incorrect decisions being made as a consequence of "data bias" due to insufficient data quality and scope (see Chapter 3.3).

Data management

Typically, a distinction is made between structured and unstructured data that is either generated internally or procured from external sources. Figure 10 provides an overview.

Figure 10: Examples of data types along two dimensions

	Internal data	**External data**
Structured data	• Data from core banking/insurance systems (e.g., transaction data)	• External market data (e.g. prices of financial instruments, information on emissions and takeovers)
Unstructured data	• E-mails, chat logs in the trade • Call centre calls • Images of damage cases in the insurance sector	• Publicly available data about customers such as annual reports, ad-hoc press releases • Blogs, newsfeeds, social media streams

---

[99] For example, based on Article 174(a) of Regulation (EU) No. 575/2013.

[100] See European Banking Authority, 2016, Final Draft Regulatory Technical Standards: on the specification of the assessment methodology for competent authorities regarding compliance of an institution with the requirements to use the IRB Approach in accordance with Articles 144(2), 173(3) and 180(3)(b) of Regulation (EU) No 575/2013, Article 76.

Structured data is often available in good quality and sufficient scope, e.g. transaction data in payments. However, the data stored by many financial services companies is fragmented: In many cases, there is no consolidated overview of the data, while the number of direct connections between data-generating systems, such as core banking/insurance applications and payment systems as well as data collection systems, increase the complexity. 65 percent of banks surveyed in a BCG study reported that they do not have any or have only a partially defined target data landscape[101]. Only 25 percent of banks surveyed have defined a data strategy at the corporate level.[102] Furthermore, according to another BCG study, the digital maturity level of insurance companies in Germany is on average lower than that of banks.[103]

Moreover, in recent years the volume of external and, in particular, unstructured data has grown rapidly.[104] The processability of this unstructured data (e.g. due to progress in speech recognition) also allows new use cases, such as autonomous chatbots at the customer interface.[105]

In addition to use in operative business processes, BDAI can also contribute to increasing data quality. Some large insurance companies have already started to use BDAI solutions to improve the quality of data that can be processed electronically. For example, information can be moved from unstructured sources to structured data, such as pre-sorting or classification of a mail inbox and extraction of contractual information from paper-based contracts. One additional example is the migration of insurance policies from legacy IT systems: thanks to BDAI solutions they can be transferred to new policy structures in the target system (see Chapter 5.3).

IT architecture

The use of BDAI applications is made possible through the provision of data in a data platform that is part of a market participant's IT architecture. A data platform refers to the combination of several layers,[106] shown schematically in Figure 11.

1. Integration layer to load data from various (internal and external) sources and for semantic data integration

2. Data retention layer (e.g. data lakes to store a wide range of data formats at different levels of granularity, format, quality etc.)

3. Data management layer (data dictionary, data security, data lineage, access authorisation)

4. API (application programming interface) layer, which provides the interfaces for reusing the data

In addition, individual data-processing applications may also be part of the data platform in order to guarantee high levels of security and speed (not shown).

---

[101] "Target data landscape" in this context is intended to refer to the target status for the entire IT systems used operatively and analytically to manage or store data.

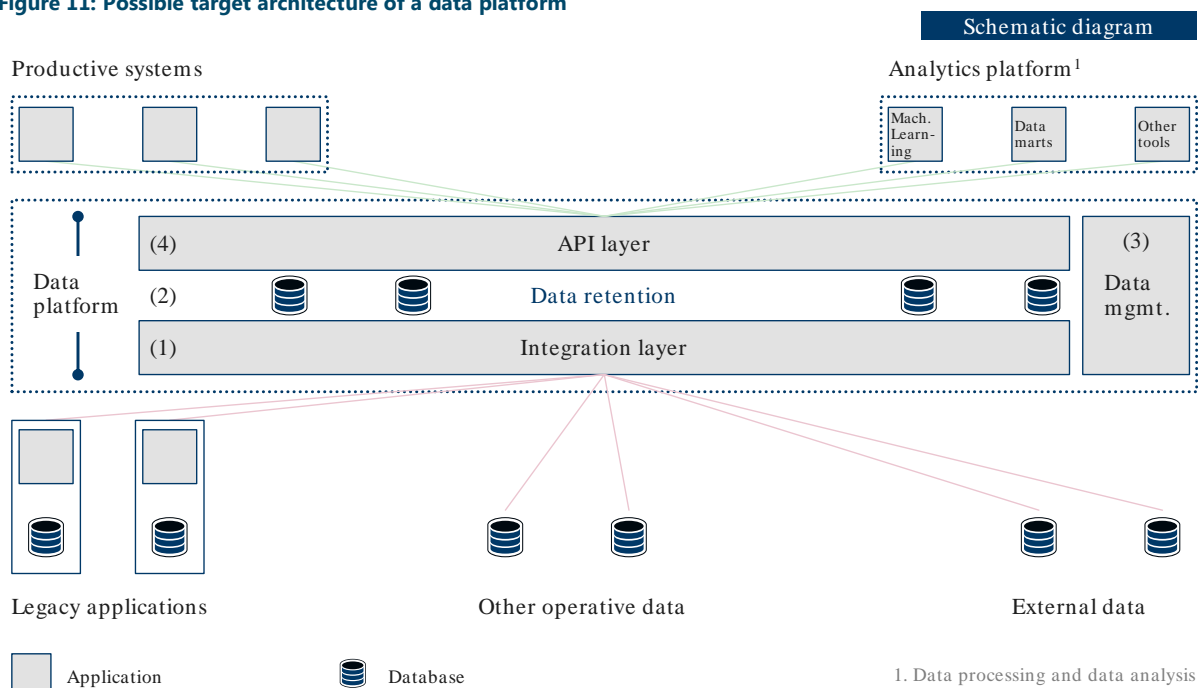[102] BCG, 2017, European IT Benchmarking in Banking.

[103] BCG, 2017, IT Benchmarking in German Insurance Companies 2017 – Digital Maturity Index.

[104] See IBM, 2013, Analytics: The real-world use of big data in financial services, p. 2. Available online: https://www-935.ibm.com/services/multimedia/Analytics_The_real_world_use_of_big_data_in_Financial_services_Mai_2013.pdf, accessed on 18 December 2017.

[105] Nadella, 2016, Remarks by Satya Nadella: Worldwide Partner Conference 2016. Available online: https://news.microsoft.com/speeches/satya-nadella-worldwide-partner-conference-2016/, accessed on 18 December 2017.

[106] See SAS, 2015, Insights from BITKOM Big Data Summit. Available online: https://blogs.sas.com/content/sasdach/2015/03/02/big-data-ist-in-deutschen-unternehmen-angekommen-erkenntnisse-vom-bitkom-big-data-summit-2015/, accessed on 18 December 2017.

Figure 11: Possible target architecture of a data platform

Schematic diagram

| Productive systems | | | Analytics platform[1] | | |

Legend: ☐ Application · ⛁ Database · 1. Data processing and data analysis

Different variants are possible when it comes to centralisation: Data platforms may, on the one hand, be developed at the level of the overall company, of individual segments or of departments. On the other hand, stand-alone BDAI applications with local data warehouses are possible (e.g. integrated, pre-configured data-platform solutions offer a combination of hardware, database software, analytics applications, etc. and thus are relatively simple to install and use). Furthermore, a distinction should be made as to whether data in the data platform is only mirrored – i.e. data will continue to be stored separately in each of the data management systems – or whether core banking/insurance systems and analytics applications access the same data sets.

Each company needs to define its own specific variant of data and IT architecture and the measures used for their implementation. In any event, these targets and measures should, as part of the IT strategy, satisfy supervisory regulatory requirements[107] and be oriented towards the business strategy of the financial service providers in question.

Use of cloud computing

Cloud computing and related services are an additional technical success factor for the use of BDAI. These allow computing and storage capacities to be used flexibly, which is relevant particularly due to the various infrastructure requirements of development and production cycles. Development cycles generally have more stringent demands, and satisfying them internally in the long term may not make economic sense. Furthermore, AI applications such as speech recognition and image, video and text analysis can be provided from the cloud.

---

[107] See Bundesanstalt für Finanzdienstleistungsaufsicht, 2017, Rundschreiben 10/2017 (BA) – Bankaufsichtliche Anforderungen an die IT (BAIT), see Note 1 and 2. Available online:
https://www.bafin.de/SharedDocs/Downloads/DE/Rundschreiben/dl_rs_1710_ba_BAIT.html, accessed on 7 February 2018.

4.2.2 Technological changes require new skills and structures as well as agile ways of working

The changes described above result, among other things, in an increasing demand for employees with mathematical and analytical abilities because the development of BDAI models requires skills in the areas of data analysis and software development (data scientists).[108] As a result, market participants need a clear personnel strategy (including requirements analysis, internal development measures and recruiting). New competitors in the bigtech world usually find themselves in a better position as they already have a large number of employees with a mathematical or analytical background and are attractive employers for potential candidates.[109] Established market participants in the financial services sector are thus locked in tough competition for talent when seeking new employees with these profiles. In Germany, for example, the STEM[110] labour gap at the end of September 2017 amounted to approximately 300,000.[111] In addition to mathematical and analytics skills, market participants need employees with the ability to identify and develop new business models in order to tap the potential offered by BDAI applications and to survive competition.

In addition, agile working methods could help address the situation of a quickly changing business environment. This pace of change will continue to increase as a result of BDAI, as described in chapter V. Agile methods could allow for faster implementation of new requirements and are likely to facilitate close collaboration between business and IT because they form joint teams from the very first phase of a project. This is particularly advantageous in the case of rapidly changing specialist requirements such as those often observed with BDAI.[112] BAIT or VAIT[113] specify certain guidelines and general conditions that are to be fulfilled by means of various methods, for example they permit the development of new functionalities in a dedicated development environment. Market participants are required to adjust their business organisation to methods such as these where necessary. New competitors such as fintechs or bigtechs tend to be in a good position in this respect because these methods are already widespread among such companies.[114]

---

[108] See IBM, 2017, The Quant Crunch: How the demand for data science skills is disrupting the job market, p. 3. Available online: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IML14576USEN, accessed on 18 December 2017.

[109] BCG analysis of indicated degrees on LinkedIn, focusing on mathematical and analytical study programmes.

[110] Science, technology, engineering, and mathematics; the German abbreviation is MINT (Mathematik, Informatik, Naturwissenschaft und Technik).

[111] Institut der Deutschen Wirtschaft Köln (IW), 2017, MINT-Herbstreport 2017 – Herausforderungen in Deutschland meistern, p. 34. Available online: https://www.iwkoeln.de/studien/gutachten/beitrag/christina-anger-sarah-berger-oliver-koppel-axel-pluennecke-mint-herbstreport-2017-368955, accessed on 18 January 2018.

[112] See BCG, 2015, Building a cutting-edge banking IT function. Available online: https://www.bcgperspectives.com/content/interviews/technology-digital-financial-institutions-ron-van-kemenade-building-cutting-edge-banking-it-function/, accessed on 18 December 2017.

[113] Versicherungsaufsichtliche Anforderungen an die IT ("VAIT"), see Bundesanstalt für Finanzdienstleistungsaufsicht, 2017, BaFin Journal December 2017, p. 17.

[114] See Kaufman, et al., 2015, The Power of People in Digital Banking Transformation. Available online: https://www.bcgperspectives.com/Images/BCG-Power-People-Digital-Banking-Transformation-Nov-2015_tcm80-199721.pdf, accessed on 18 December 2017.

New skills and methodologies are not only required at the level of individual employees – BDAI also presents the senior management of banks and insurance companies with new challenges. Market participants are required to adjust their setup and process structures to these new skills and working methods. In this context, supervisory authorities like BaFin increasingly attach importance to adequate IT skills and make it easier to appoint IT experts to the board.[115]

4.2.3 The importance of information security

Information security is gaining in importance in general, therefore systematic consideration and a clear understanding of risks are indispensable.

Information security is intended to secure the confidentiality, integrity and availability of information.[116] A lack of information security can thus have significant impact on financial services providers: If regulatory requirements (e.g. EU GDPR) are not fulfilled, fines may be imposed.[117] It may also have a negative impact on how the institutions function (e.g. due to interruption of business operations) and on consumer trust (see Chapter 4.1). In extreme cases, the economic damage may even endanger the companies' survival.

The number of information security breaches will most probably increase in the years to come. The costs to companies arising from cybercrime worldwide are estimated to rise from USD 445 billion in 2015 to USD 2.1 trillion by 2019.[118] [119] And we assume that the actual, unreported figures will be even higher.[120] In the largest economies (the USA, Japan, Germany and the UK), cyberattacks and data fraud are the most significant risks for companies according to the Global Risk Report 2017 of the World Economic Forum.[121] Accordingly, regulators and supervisory authorities are also addressing this issue.[122]

---

[115] See Bundesanstalt für Finanzdienstleistungsaufsicht, 2017, IT-Kompetenz in der Geschäftsführung. In: BaFinJournal, pp. 15 ff. Available online: https://www.bafin.de/SharedDocs/Downloads/DE/BaFinJournal/2017/bj_1712.pdf;jsessionid=677DC922154CC5480E4AA4C656A 12459.2_cid363?__blob=publicationFile&v=3, accessed on 18 December 2017.
European Banking Authority and European Securities and Markets Authority, 2017, Joint ESMA and EBA Guidelines on the assessment of the suitability of members of the management body and key function holders, p. 34. Available online: https://www.eba.europa.eu/documents/10180/1972984/Joint+ESMA+and+EBA+Guidelines+on+the+assessment+of+suitability +of+members+of+the+management+body+and+key+function+holders+%28EBA-GL-2017-12%29.pdf, accessed on 18 December 2017.

[116] BCG, 2014, Cybersecurity Meets IT Risk Management: A Corporate Immune and Defense System. Available online: https://www.bcgperspectives.com/content/articles/technology_strategy_technology_organization_cybersecurity_meets_it_risk_m anagement/, accessed on 10 January 2018.

[117] See Art. 83 EU GDPR, General conditions for imposing administrative fines.

[118] Juniper Research, 2016, Cybercrime will Cost Businesses Over $2 Trillion by 2019. Available online: https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion, accessed on 30 October 2017.

[119] See B20 Germany 2017, Taskforce Digitalization Policy Paper 2017. Available online: https://www.b20germany.org/fileadmin/user_upload/documents/B20/B20_Digitalization_Policy_Paper_2017.pdf, accessed on 30 October 2017.

[120] As a consequence of the strategic relevance of information security, there is reluctance to publicly reveal incidents involving information security.

[121] See World Economic Forum, 2017, Global Risks of Highest Concern for Doing Business. Available online: http://reports.weforum.org/global-risks-2017/global-risks-of-highest-concern-for-doing-business-2017/#, accessed on 21 November 2017.

[122] See Financial Stability Board, 2017, Stocktake of Publicly Released Cybersecurity Regulations, Guidance and Supervisory Practices. Available online: http://www.fsb.org/wp-content/uploads/P131017-2.pdf, accessed on 16 January 2018. See also Chapter 7.2.

Threats to information security come not only from the outside, in the form of hacker attacks for example, but can come from within the company, too: insufficient prevention and protection measures within the organisation, in the processes themselves and on the part of employees often pose hazards to information security.[123] As part of this study, two categories of information security risks were examined (see Figure 12):

- Data and system risks that compromise the reliability of IT systems and hardware. These also include usage errors and thus internal security risks.

- Cyber risks that include criminal activities intended to damage the functionality of a financial services provider and/or to gain a financial or non-financial benefit from these activities. These also include external risks such as hacker attacks.

Figure 12: Overview of information security risks

Data and system risks

Risk of failure or interruption
Interruptions to operations due to external factors (e.g. natural disasters) or a lack of important resources (e.g. electricity, staff)

Risk of system failure
Interruptions to operations due to compromised or faulty systems (e.g. faulty system design)

Data security risk
Risk to the authenticity, integrity and confidentiality of the data (e.g. incorrect or unintentional modification or loss of data by staff)

Risk posed by/in cooperation with third parties
Damage and/or impairments to operation due to incorrect use of data outside of a company's (own) secure corporate networks (e.g. in the context of outsourcing)

Cyber risks

Risk posed by unauthorised access
Damage due to unauthorised access without the use of malware (e.g. identity theft or phishing, employees using insecure passwords)

Hacker risk
Damage due to the use of malware (e.g. denial of service, manipulation of software and hardware)

Risk posed by/in conjunction with third parties
Damage and/or impairments to operations due to unauthorised activities by third-party systems

BDAI increases the degree of connectivity and data usage, and thus also the risks to information security. At the same time, BDAI applications can help increase information security.

In addition to the increasing risk resulting from the growing connectivity among market players and greater data usage, new risks may arise as a consequence of BDAI. One example would be adversarial attacks. Here, attackers deliberately change the input for machine-learning algorithms in order to generate the output that the attackers intended. This manipulation may be aimed, for example, at unstructured data such as images and may trigger an incorrect classification on the part of the AI system. The change made to the input (images) can, however, be so subtle that it is barely perceptible to the human eye.[124] This underlines the fact that the use of new and complex methods such as deep learning may require additional efforts to guarantee information security.

---

[123] Deutscher, et al., 2017, Building a Cyberresilient Organization, p. 3. Available online: https://www.bcgperspectives.com/content/articles/technology-digital-building-a-cyberresilient-organization/, accessed on 10 January 2018.

[124] See Kurakin et al., Adversarial Examples in the Physical World. Available online: https://arxiv.org/pdf/1607.02533.pdf, accessed on 21 February 2018.

In order to address information security risks, companies have to develop a differentiated understanding of their assets that need particular protection. Using BDAI also creates new assets that need to be protected.[125] Apart from data, this includes algorithms. These may represent an asset as well as an information security risk in the form of a potential weakness; in fact, they may be susceptible to manipulation, which could increase the information security risk particularly for analytical processes that are hard to trace (see Chapter III).

- ■ Overall, using BDAI can create and change the risks particularly quickly. It is only possible to a limited extent to address information security risks in the same manner as other operational risks. Within the context of BDAI, greater flexibility and more frequent adjustments to internal guidelines and controls are required. Furthermore, in the past, companies have generally concentrated on preventive measures. Against the backdrop of as-yet unknown risks related to BDAI usage, as well as dynamic change to risk evaluation, detection and response abilities will become more important. The same applies to the skills and abilities of employees.[126]

As a result, companies should determine a clear risk tolerance for information security. To avoid and reduce information security risks (beyond the risk tolerance), financial service providers are required to implement measures in several areas of their business model. The eight fundamental elements of cybersecurity for the financial services sector as laid down by the G7 member states could serve as building blocks (see Figure 13).[127] These can be used as the basis on which companies can develop and implement a cybersecurity strategy and an operative frame of action while also taking their own approach to risk control and risk culture into account. These elements represent individual steps in a dynamic process in which companies can systematically realign their cybersecurity strategy and their frame of action once the operative environment and the security threat change.

---

[125] European Union Agency for Network and Information Security, 2016, Big Data Threat Landscape, p. 12. Available online: https://www.enisa.europa.eu/publications/bigdata-threat-landscape, accessed on 10 October 2017.

[126] See Dreischmeier, 2017, Digital Without Cybersecurity Is a Train Wreck Waiting to Happen. Available online: https://www.linkedin.com/pulse/digital-without-cybersecurity-train-wreck-waiting-ralf-dreischmeier/, accessed on 10 January 2018.

[127] G7, 2016, G7 Fundamental Elements of Cybersecurity for the Financial Sector. Available online: https://www.ecb.europa.eu/paym/pol/shared/pdf/G7_Fundamental_Elements_Oct_2016.pdf?69e99441d6f2f131719a9cada3ca56a 5, accessed on 9 February 2018.

Figure 13: G7 fundamental elements of cybersecurity for the financial sector

	Cybersecurity and Strategy Framework		**Response**
	Governance		**Recovery**
	Risk and Control Assessment		**Information Sharing**
	Monitoring		**Continuous Learning**

Meanwhile, BDAI offers the potential for detecting information security risks more easily. Potential use cases include detecting anomalies, i.e. deviations from probable patterns (for instance when using login data), or identifying attack patterns.[128] BDAI can also contribute to a better understanding of possible cyber risks and thus to better threat intelligence.[129] Furthermore, BDAI could, through intelligent automation, significantly reduce the amount of manual labour required in cybersecurity. Overall, the quality and speed of IT defences could be increased.[130]

---

[128] See European Union Agency for Network and Information Security, 2016, Big Data Threat Landscape, p. 56–59. Available online: https://www.enisa.europa.eu/publications/bigdata-threat-landscape, accessed on 10 November 2017. Botnets refer to a group of automated malware that are run on networked computers.

[129] See Samtani et al., 2018, Exploring Emerging Hacker Assets and Key Hackers for Proactive Cyber Threat Intelligence. In: Journal of Management Information Systems.

[130] See Gartner, 2017, Top 10 Strategic Technology Trends for 2017: Adaptive Security Architecture, p. 10. Available online: https://www.gartner.com/doc/3645329/top--strategic-technology-trends, accessed on 16 January 2018.

V. Market analyses

5.1 Introduction

Analysis of the broader impact of big data and artificial intelligence.

The following sections of this chapter will examine the potential impact of BDAI on the banking and insurance industry as well as on the capital markets. In general, it is not always possible to clearly distinguish these effects from those that accompany the general digitalisation of the financial system, which is why BDAI are to be understood hereinafter in a more general sense. This includes a consideration of the elements of digitalisation that may be a precondition for the use of BDAI (e.g. digitalisation of information) as well as of those that result from the use of BDAI (e.g. digitalisation of processes).

Analysis of the market reveals a dual nature. On the one hand, BDAI allows digitalisation, automation and thus a more efficient design of processes that, to date, have been handled predominantly manually due to their heterogeneous nature. One example would be the recognition of relevant content in scanned documents and derivation of suitable actions. On the other hand, BDAI can contribute to improving valuation and prediction models – for instance for the probability of loan defaults or optimisation of the product offering in line with predicted customer requirements – thus improving effectiveness.

The analyses contained in the following sections describe the processes, products and providers currently found in the marketplace and take into consideration relevant literature and the results of accompanying conversations with experts. As with all forward-looking statements, the forecasts contained herein are subject to uncertainty.

The following representations of banks and insurance companies differ structurally from the way the capital market was analysed.

The sections pertaining to the analysis of banks and insurance companies are generally more predictive due to the fact that in both industries BDAI usage is for the most part still in its infancy. The processes involved are indeed strongly governed by rule-based and automated procedures, but banks and insurers are only now beginning to use data to a greater degree thanks to BDAI applications. By contrast, many more BDAI applications are to be found in the capital market – the result of many years of experience using algorithms and large data sets, for instance in trading with volumes in the petabyte range. In the capital market, processing the information available was a given and data was exploited very soon after emergence of the information theory, meaning that the paths taken were conceptually different to those followed by the more process-orientated banks or insurers.

Furthermore, the sections on banks and insurers follow a market structure based on customers on the one hand and on (financial) service providers on the other. Analysis of these sectors therefore focuses specifically on the customer interface, on key internal processes and on potential new business models. By contrast, the capital market is characterised by an interaction structure made up of a multitude of mostly professional players. For this reason, the analysis follows major trends.

Three major groups of providers in the financial system.

Among other things, the market observations describe potential action scenarios for players within the financial system. These players can be divided into three groups:

■ **Incumbents**, i.e. traditional companies in the industry, meaning those companies active in the market, such as banks, insurance companies and capital market players[131], which tend to have a comprehensive trading licence issued by the supervisory authorities.

■ **Fintechs/insurtechs/regtechs/legaltechs**, i.e. active and comparatively young, technology-orientated and smaller providers, which offer select functions at the customer interface or in a core process for one or several products (sometimes in cooperation with one or more incumbents).

■ **Bigtechs**, i.e. large, globally active technology companies which gain competitive advantage from the use of digital applications. They focus predominantly on online services, (IT) platforms or on making digital infrastructures available to themselves and other providers for offering products and services.[132] Some bigtechs already have a trading licence for individual financial services.[133] Such companies are not yet active or are but only to a limited degree in the financial market in Continental Europe.

---

[131] In the capital markets, the incumbents include several groups (e.g. banks, asset managers, trading centres, clearing houses, poss. also suppliers of data and analytics) and are less static in their actual composition, e.g. new hedge funds are emerging almost constantly.

[132] Cf. Bank for International Settlements, 2018, Sound Practices. Implications of fintech developments for banks and bank supervisors issued by the Basel Committee on Banking Supervision, p. 15.

[133] For example, Amazon with AmazonPay, Google with Google Pay or Alibaba with Ant Financial.

5.2 Banks

5.2.1 Introduction and status quo

In the banking sector, traditional financial institutions face new customer expectations at the same time as pressure on cost and margins as well as legacy IT systems.

Digitalisation is bringing about a change in the customer experience expected by certain customer groups in many industries. Today, large technology companies are setting the standards in terms of the speed of processes and decision-making, smooth interaction between provider and customer, intuitive user interfaces[134] and individualisation/personalisation[135] of the services and offering. These types of experience in other industries lower many customers' tolerance when providers take longer to respond to enquiries or applications. Many customers project their expectations onto products and services in the banking industry. In order to rise and meet these new standards and customer expectations, i.e. to provide an improved customer experience, many financial services providers could make greater use of BDAI applications.

The high volumes of investment in fintechs show that adjustments are already being made to meet the changed expectations of certain customer groups. For example, global investment in fintechs employing BDAI grew by 62 percent between 2014 and 2016 to a total $2.3 billion (cf. Figure 14). Moreover, even the bigtechs are already seen to be including financial services in their offering, albeit to a lesser degree. Both of these developments mean that BDAI services and applications could become more prevalent in banking.

Figure 14: Development of the global volume of investment in BDAI fintechs[136]

---
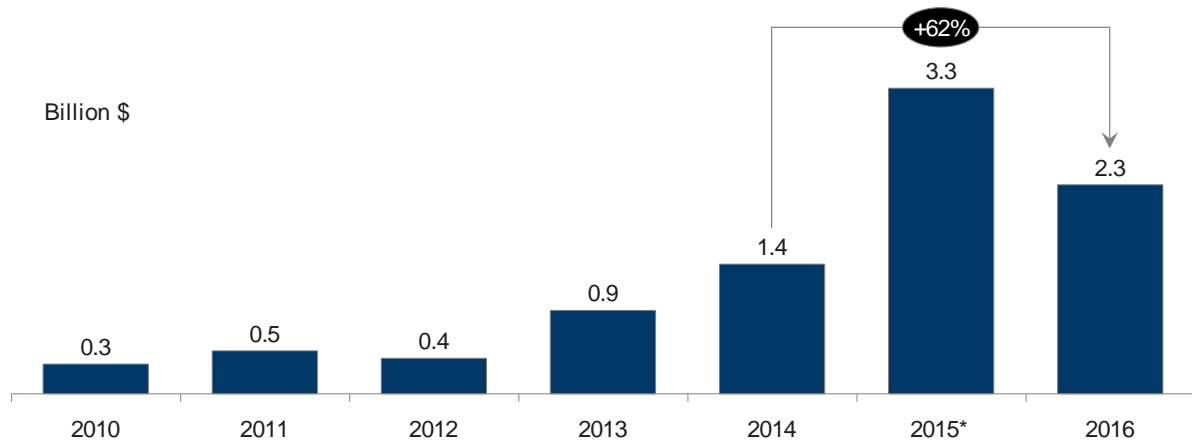
[134] Some examples are voice-activated assistants for home and mobile phones or voice control in car navigation systems. The banking sector now also offers solutions that enable customer identification or control of the mobile app via voice recognition.

[135] This concept, also known as "segment-of-one" or "mass customisation", can be considered a further refinement because it is now possible to provide each customer with a product tailored to his or her own needs.

[136] BCG FinTech Control Tower. * 2015 growth in BDAI fintech funding especially as a result of Palantir's $880 million round.

The traditional European banking industry is confronted with two challenges that could now curb widespread use of BDAI and thus a suitable response to this change in customer expectations. On the one hand, the European banking market has experienced significant cost and margin pressure since the financial crisis. Continued low interest rates, a decline in the balance sheet totals of European banks and low margins make it difficult for the institutions to generate sustainable revenues. Overall, incumbents feeling cost pressure find they have very little leeway for investment in new technologies. However, such investments could be just what is needed for incumbents to meet the changed customer expectations.

On the other hand, legacy IT – that is, an IT architecture that has grown over the years – and its sometimes functionally redundant applications might be one reason for high operating costs and actually represent an obstacle to implementing change.[137] Thus, potential innovations at the customer interface and in the technology-based optimisation of key processes can be implemented only comparatively slowly and at high cost. However, despite the great expense, such a transformation could be useful for incumbents, particularly if BDAI makes a sustainable contribution toward lowering costs. The incumbents' legacy IT and their limited ability to integrate new applications often represent a structural disadvantage compared with new providers.[138] The latter can often already boast a state-of-the-art IT architecture or can build new application landscapes without having to take any legacy IT system into account. This is why various banks are promoting the development of parallel digital banks and are starting to acquire or initiate cooperation with fintechs.[139]

The following analysis will focus on the potential impact of BDAI on the overall market, the customer interface, the core processes and new business models. In doing so, for the sake of simplification, three groups of banking products (deposits/savings, loans, payments/accounts) and three customer groups (retail customers, small corporate customers, large corporate customers) will be distinguished. In addition, the three groups of providers introduced in chapter 5.1 are also analysed: incumbents (traditional banks), bigtechs and fintechs.

---

[137] Cf. Grebe et al., 2016, Simplifying IT to Accelerate Digital Transformation. Available online: http://image-src.bcg.com/Images/BCG-Simplifying-IT-to-Accelerate-Digital-Transformation-Apr-2016_tcm9-87515.pdf, accessed on 30 January 2018.

[138] Cf. Saleh et al., 2017, Why Aren't Banks Getting More from Digital? Available online: http://image-src.bcg.com/Images/BCG-Why-Arent-Banks-Getting-More-from-Digital-Dec-2017_tcm9-179412.pdf, accessed on 30 January 2018.

[139] While in 2010 a mere six percent (across all segments) or one percent (retail banking) of investments in fintechs was made by banks (the rest by venture capital or similar), this share rose in 2016 to 49 percent (all segments) and 71 percent (retail banking); BCG FinTech Control Tower.

5.2.2 The impact of BDAI on the banking sector

One significant change in the course of disaggregation of the value chain in the banking sector can be the separation of the customer interface from the product platform[140].

The spread of BDAI applications is intensifying the already discernible trend toward disaggregation of the value chain in banking, whereby new competitors with BDAI-driven business models or BDAI applications claim sections of the value chain for themselves.[141] Two further developments foster this trend toward disaggregation: On the one hand, modified regulatory requirements (e.g. PSD 2)[142] are accelerating the creation of interfaces standardised to the highest degree[143], which in turn could intensify the break-up of the value chain. On the other hand, specific technological developments, such as the trend toward open banking[144] and thus the related broad availability of API-based applications[145], promote the technical integration of individual, as yet isolated, offerings along the value chain.[146]

The value chain can thus disaggregate both along bank business processes and products (cf. Figure 15).[147]

---

[140] In the following, the product platform comprises those middle and back office functions and processes that are necessary for delivering products and services to the customer interface (front office), but that are not visible to the customer.

[141] Several providers offer very specialised products and services, for example robo-advisors, to improve interaction with the customer.

[142] In the context of payments, PSD 2 will for instance permit account information service providers (AISPs) to query customer account information through dedicated APIs (Application Programming Interfaces, cf. footnote 145) and then use this to offer the customers new information services without the need for these AISPs to manage the customer accounts. Examples of such new services would be applications that offer customers a complete overview of several accounts (personal finance managers). These retrieve transaction data for several accounts and present them in aggregated form. (EU) 2015/2366, Payment Services Directive.

[143] Cf. Cortet et al., 2016, PSD2: The digital transformation accelerator for banks. In: Journal of Payments Strategy & Systems, Volume 10 (1). Available online: https://www.innopay.com/assets/Publications/JPSS-Spring-PSD2-digital-transformation-for-banks-Innopay.pdf, accessed on 30 January 2018.

[144] In this connection, various technological trends, such as the increasing standardisation, modularisation and encapsulation of systems and applications (for instance through the spread of APIs), should be grouped together under "Open Banking" (cf. Chapter 4.2).

[145] Application programming interfaces (API) that facilitate the simple use of complex functionalities; cf. further details in Chapter 4.2; cf. also Zachariadis et al., 2017, The API Economy and Digital Transformation in Financial Services: The Case of Open Banking. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2975199, accessed on 30 January 2018.

[146] Cf. Chapter 4.2

[147] A comparable classification can also be found in a consultation document of the Basel Committee on Banking Supervision; cf. Bank for International Settlements, 2017, Sound Practices: Implications of fintech developments for banks and bank supervisors. Basel Committee on Banking Supervision. Available online: https://www.bis.org/bcbs/publ/d415.pdf, accessed on 23 January 2018.

Figure 15: Schematic representation of the transition from an aggregated to a disaggregated value chain

Aggregated value chain

Disaggregated value chain
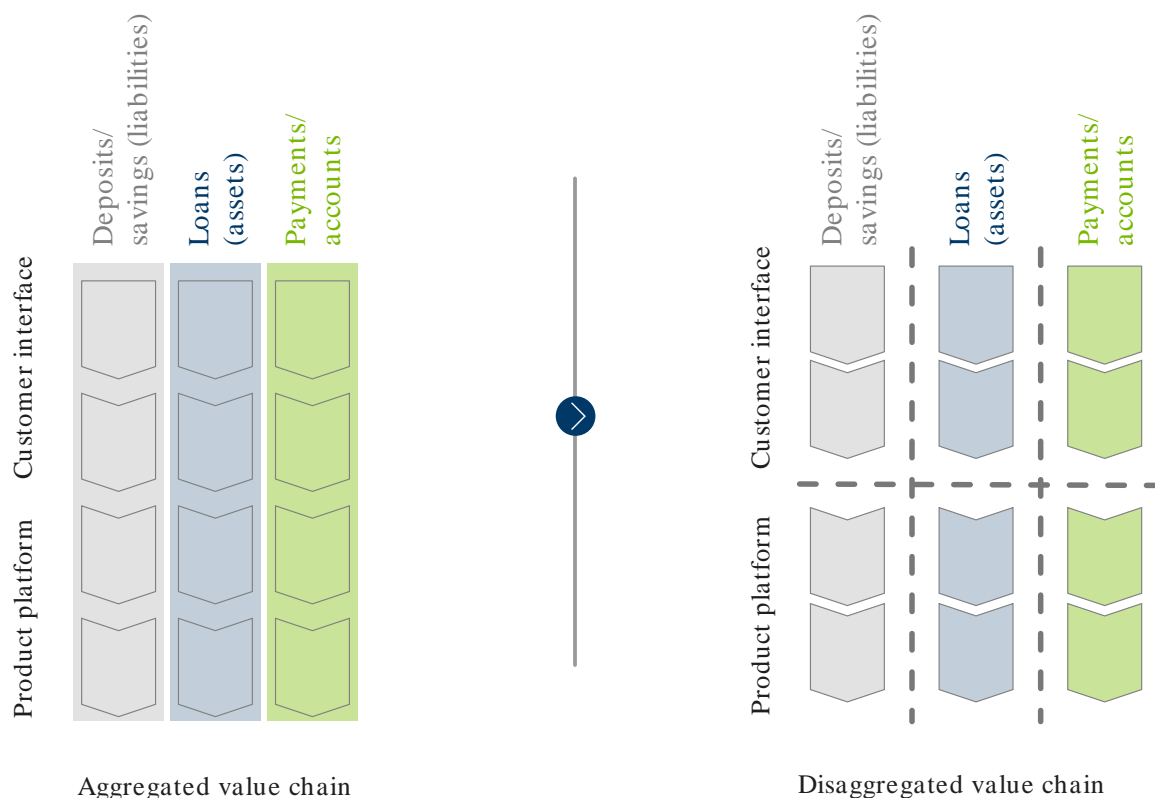
The trend toward horizontal disaggregation of the value chain is slowly producing a growing number of providers that specialise in one or only a few products. Increasing numbers of new players are emerging, particularly in payment services (cf. Chapter 5.2.3).[148]

Vertically disaggregated value chains now exhibit a stronger separation of activities involving direct contact with the customer at the customer interface and activities in middle and back office (product platform) than used to be the case. Customers regularly assume that the providers with whom they are in direct contact and who own the customer interface are actually the product owners. These providers manage the direct customer relationship and integrate other providers along the value chain as suppliers, thereby attaining a comparatively stronger competitive position.[149]

---

[148] Cf. Dorfleitner et al., 2016, The Fintech Market in Germany. Report commissioned by the Federal Ministry of Finance, p. 15. Available online: https://ssrn.com/abstract=2885931, accessed on 30 January 2018; cf. European Banking Authority, 2017, Discussion Paper on the EBA's approach to financial technology (FinTech), p. 22. Available online: https://www.eba.europa.eu/documents/10180/1919160/EBA+Discussion+Paper+on+Fintech+%28EBA-DP-2017-02%29.pdf, accessed on 30 January 2018.

[149] In this way a provider could, in principle, deliver solely the access to a service and secure all other activities involved from another provider who would cover all the regulatory and technological requirements. However, this separation would not be immediately evident to the customer.

In the following, the product platform is understood as comprising the middle and back office functions and processes that are necessary for delivering products and services at the customer interface (front office). These include concluding contracts, creating accounts, conducting detailed analyses to answer customer enquiries and making manual bookings as well as correcting errors made in payment transactions. This requires a technological platform, for example the portfolio management or transaction processing systems. Furthermore, and depending on the position in the value chain, this may require certain skills and capacities for meeting regulatory requirements[150], for example securing capital requirements, managing risk or creating notifications and reports for the supervisory authorities. And last but not least, another essential element is specialist know-how for developing financial products.[151]

The stricter separation of activities at the customer interface from those in the middle and back office could also trigger changes in the way competences are structured within the institution: While product design often used to take the form of separate agreements with individual customers in front office/sales, the introduction of dedicated product platforms and their use on the basis of clearly defined interfaces could shift the initiative for new products and their design more in the direction of the product platform. Then, building on standardised products, the product could be tailored to individual needs at the customer interface.

Disaggregation of the value chain could create a more heterogeneous landscape of players. Already a multitude of young companies like fintechs[152] and some bigtechs[153] can be found alongside the established banks. Provided they comply with the regulatory outsourcing requirements, all of these providers could concentrate on specific sections of the value chain. It is likely in such a scenario that several providers would focus on the customer interface and buy in the product platform required for value creation from a third party. The goal would be to own the comparatively attractive customer relationship while minimizing the complexity and costs associated with the provision of the product platform. However, the provision of the product platform could either constitute an active part of the business model – as part of the product platform business model (white labelling) – or a passive part, that is the corollary of the abovementioned strategy of new players taking over the customer interface. Examples of product platform strategies are mainly found among fintechs typically specialising in specific products and services. On the other hand, this could also be a strategy for incumbents who could actively decide to separate the front office from the back office and offer different parts of their value chain in a "one-stop shop", while keeping them separate both organisationally and technically.

There will also be providers who take on the customer interface but also offer the product platform.[154] These providers would provide all products and functions in the form of a full-service bank.

---

[150] Chapter VI addresses the supervisory and regulatory implications of the disaggregation of the value chain.

[151] For an example of vertical disaggregation of the value chain in the case of payments, cf. Badi et al., 2017, Global Payments 2017: Deepening the Customer Relationship. Available online: https://www.bcg.com/publications/2017/transaction-banking-financial-institutions-global-payments-2017-deepening-customer-relationship.aspx, accessed on 30 January 2018.

[152] In this way, the number of active fintechs rose from approx. 4,900 in 2010 to approx. 10,350 in 2017; BCG FinTech Control Tower.

[153] Bigtechs are primarily active in payments and in providing cloud services (cf. the explanations in the "Customer Interface" section and in chapter 4.2).

[154] This can be dedicated to one product (in the sense of the horizontally disaggregated value chain as a specialist provider) or indeed cover several banking products as a traditional full-service bank. Other substrategies are conceivable but will not be explored at this point. Similar approaches have been pursued by the Basel Committee on Banking Supervision; cf. Bank for International Settlements, 2017, Sound Practices: Implications of fintech developments for banks and bank supervisors. Basel Committee on Banking Supervision. Available online: https://www.bis.org/bcbs/publ/d415.pdf, accessed on 23 January 2018.

Incumbents, fintechs and bigtechs reveal different strengths along the relevant competitive dimensions.

For success in the market, four different competitive dimensions involving BDAI need to be considered. The relevance of each varies depending on the strategic focus of the provider:[155]

- **Consumer trust**. This can be understood as trust in the security and handling of personal data (cf. Chapter 4.1). [156]

- **Technological BDAI platform**. A flexible and modern system architecture permitting rapid response to new requirements (transformation capability), a suitable data architecture and tools for employing BDAI is meant here. It also means having the requisite employee skills and corresponding experience as well as the ability of the organisation to work agilely throughout the company (see Chapter 4.2).

- **Customer reach**. This refers to the number of customers that a provider can actually reach. Under the right circumstances, a high customer reach allows the provider to roll out operations very quickly to a large number of customers and therefore to potentially secure a first mover advantage.

- **Regulatory competence and industry know-how**. This includes processes, methods and tools to be established as well as employees with the relevant knowledge of regulations and the industry. Compared with other sectors, regulatory competence is particularly relevant because the banking industry is a strongly regulated sector. Both processes and expertise are subject to the constant pressure to adapt to changing regulations. In addition, bank licences are a prerequisite for many services.[157]

The groups of providers under study – incumbents, bigtechs and fintechs – have different positions in the market in terms of the four dimensions described. Figure 16 gives a very simplified overview of the relative competitive dimensions. At the same time, these providers have different business objectives and therefore pursue different strategies in terms of market positioning.

---

[155] Overall, the providers are also pursuing different strategies in terms of their position in the value chain. It is expected that the relative strategic positions will not be clearly allocated to one provider group, but that there will be several variations and intermediary steps. Furthermore, the positions can change over time as framework conditions and preferences/objectives change.

[156] This assessment is based on surveys on the topic of data protection and the handling of personal data; cf. the Special Eurobarometer 431 for information on the trust levels in the protection of personal information by authorities and private companies, June 2015, Question: "Different authorities (government departments, local authorities, agencies) and private companies collect and store personal information about you. To what extent do you trust the following authorities and private companies to protect your personal information?"); cf. also the survey: Die Kunden auf dem Weg zur Digitalisierung mitnehmen, March 2017 (Taking customers with you along the path toward digitalisation), commissioned by the Association of German Banks, Question: "Customer data is protected from third-party access by the following providers: Banks, bigtechs, fintechs".

[157] The licence requirements will not be examined in detail.

Figure 16: Diagram illustrating the relative characteristic strengths along the competitive dimensions for incumbents, bigtechs and fintechs
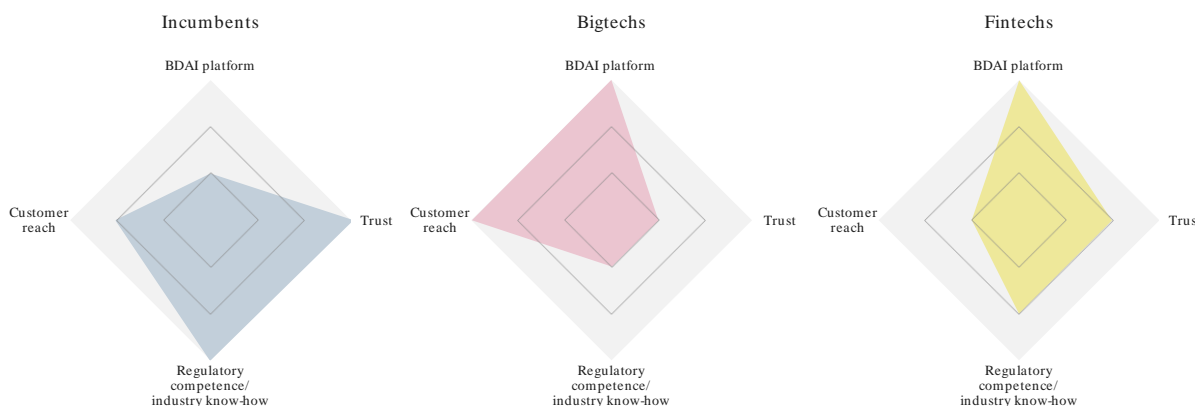
Traditional banks/incumbents enjoy a comparatively high level of consumer trust (cf. Chapter 4.1) and demonstrate greater regulatory competence due to the fact that, for years or even decades, they have had to meet a whole host of constantly changing regulatory requirements and can build on a long-standing relationship with the supervisory bodies. Furthermore, customer reach levels are high and they are therefore in a position to potentially roll out new business models to a large number of existing customers. In addition, banks enjoy the advantage of having comprehensive transaction data of their customers at their fingertips – this includes in particular an overview of the transactions of all their existing customers (customer portfolio). The introduction of PSD 2 will potentially diminish this advantage somewhat because, providing the customer's consent of course, it stipulates that transaction data must be made available to third parties. These providers could then successively develop and expand their own customer portfolio overview. At present, the technological platform and, in particular, the availability of capabilities in areas needed for BDAI still often present incumbents with a challenge. The requisite investments in a technological platform are regularly higher than for fintechs or bigtechs because the incumbents have IT architectures that have grown over the years.

Compared to incumbents, **bigtechs** typically boast the opposite strengths in these four competitive dimensions: They have a modern technological platform for their core business divisions, they use BDAI widely throughout the company and have the employees and methods they need to rapidly and efficiently develop new technological capacities. They are also in a position to exploit their particularly broad customer base and their flexible platforms to scale their operations quickly, thus attaining a strong or even dominant market position in new areas of business.[158] This ability is decisive in those markets in which bigtechs are typically active.[159] Bigtechs are also in a position to make large investments to develop their technology and open up new business areas. These investments can be made both through in-house research and development as well as through investment in fintechs.[160] Moreover, in making these investments, bigtechs are often in a position to take greater risks with new business models or products than traditional banks can

---

[158] For example, the music streaming services of Amazon (Prime Music, 2014) and Apple (Apple Music, 2015) very quickly secured top market positions behind Spotify thanks mainly to their large customer base and their platform; cf. MIDIA, 2017, Amazon Is Now The 3rd Biggest Music Subscription Service. Available online: https://musicindustryblog.wordpress.com/2017/07/14/amazon-is-now-the-3rd-biggest-music-subscription-service/, accessed on 30 January 2018.

[159] These markets can in principle be considered "winner takes all" markets, in which the big companies get bigger and bigger, cf. the comments on "centralised production" in the context of digital and information assets in Loebbecke et al., 2015, Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. In: The Journal of Strategic Information Systems. Available online: http://dx.doi.org/10.1016/j.jsis.2015.08.002, accessed on 28 February 2018.

[160] In this way, the bigtechs have been involved in over 100 funding rounds for fintechs since 2007, cf. BCG FinTech Control Tower.

or prefer to.[161] In terms of consumer trust, bigtechs currently face the challenge of consumers trusting them less than they trust other providers with handling their personal and financial data (cf. Chapter 4.1). Furthermore, bigtechs do not have the same wide range of experience that traditional banks do with regard to the regulation and development of financial products and, in the past, have had comparatively little involvement with supervisory authorities and regulators. To date, this group has not secured any full banking licence in Europe.[162]

Fintechs typically have a modern and flexible technological platform because they do not have to build on any existing IT architecture. This means that complexity and IT maintenance costs tend to be lower than for traditional banks – also often due to the general concentration on comparatively few products and/or services. On the other hand, due to their low levels of customer reach, fintechs are comparatively slower at rolling out their new business models because they generally have to build their customer base from scratch. In addition, it can be observed that consumers currently show less trust in fintechs than in traditional banks. In terms of their regulatory competence, the fintech group tends to be more heterogeneous. Many providers concentrate on the customer interface and purchase their product platform from third parties with the requisite licences, meaning that their regulatory competence is not so strong. Other fintechs concentrate on offerings from the product platform and have built up the requisite regulatory competence. BaFin has already granted a few German fintechs a licence to trade as a credit or financial services institution (full banking licence).

Regtechs, as a special type of fintech, prove excellent regulatory competence as they specifically use BDAI applications to enhance, for instance, the effectiveness and efficiency of the bank's compliance process.

---

[161] For example, a statement made by Scott Galloway (Professor at the NYU Stern School of Business): "They're fantastic at taking big risk. Putting metrics on them and just as importantly pulling the plug and performing infanticide on projects that aren't working and then moving on to the next thing. Whether it was auctions or the phone with Amazon or Facebook and some of their targeting, all of these companies have had huge missteps, but it just doesn't matter because they on average win." Available online: http://www.businessinsider.com/scott-galloway-amazon-facebook-apple-google-failure-success-tech-2017-10?IR=T, accessed on 22 January 2018.

[162] Amazon, Facebook and Google each have a payments service licence in line with the respective national implementation act for the PSD directive. Facebook has had an e-money licence in Ireland since 2016, Amazon in Luxembourg since 2011, Google in Great Britain since 2007. Amazon EU has also had a licence as a credit broker in Great Britain since 2015. However, in Germany these companies do not have any deposit or credit licences (full banking licence). PayPal, on the other hand, has been licensed to operate as a credit institution in Luxembourg since 2007.

5.2.3 The impact of BDAI on the customer interface

Those providers that offer the best customer experience in the respective customer segments will most likely prevail at the customer interface, resulting in a potential redistribution of earnings.

BDAI is of great significance for the customer interface because it can help meet new customer expectations. Many customers now also expect their banks to provide efficient processes that allow communication and the real-time conclusion of contracts fully online. Here, BDAI can support the gradual optimisation and, in an ideal scenario, complete automation of key processes (cf. the next section). For instance, credit decisions could be made for customers in real time on the basis of (transaction) data already available, thus improving the customer experience. BDAI could also facilitate the 24/7 availability of banks by means of chatbots in call centres or through the institutions' online offerings.[163] Examples might be immediate replies to questions arising during the contract conclusion phase – even outside of banking hours – or to questions arising during the contractual relationship, for instance regarding means of payment, statements of expenditure or account movements and transfers. Some customers still expect simple user interfaces and interaction options tailored to their individual preferences, such as the possibility of voice control or voice recognition instead of passwords and the like.[164] Furthermore, it is conceivable that banks use the information available to present customers with customised offerings that include exactly those products and services relevant to their present circumstances. Although many banks already offer products and services tailored to different phases of life (for example, upon graduation or marriage), this might offer further potential for individualisation. Improvements to the customer experience are important in the large corporate customer segment, too, and could be decisive on the competitive front because many of the end users at a big corporate client might expect to see the standards they know from their own private banking in their professional environment.

Fintechs and bigtechs could specifically target the customer interface because there they can leverage one of their core competences. Their strongly data-driven business models help them forge innovative paths, for instance through the use of externally available data, such as the French website "Infogreffe"[165]. For bigtechs in particular, collecting more customer data at the customer interface is vital because this data could supplement their core business operations beyond banking services. One possible approach would be to provide financial services free of charge.

For incumbents, on the other hand, the main concern is often to maintain customer relationships. Furthermore, due to the high level of investments necessary for transforming technological platforms, traditional banks are that much more dependent on the profitability of their original business operations. Many incumbents are already cooperating with fintechs to supplement their offerings at the customer interface. These types of cooperation are also attractive to fintechs because they give them access to a broad, ready-made customer base. Typical examples of fintech offerings at the customer interface include payments, investment advice, for example by means of robo-advisors, and crowdfunding. Over the long term, fintechs are primarily dependent on the profitability of their original business models. However, when they start out, they often concentrate more on growth and less on profitability. And in the medium term, the monetisation of data can also play a role.

---

[163] The increasing proliferation of such chatbots has also been addressed by the Financial Stability Board; cf. Financial Stability Board, 2017, Artificial intelligence and machine learning in financial services. Available online: http://www.fsb.org/wp-content/uploads/P011117.pdf, accessed on 23 January 2018.

[164] A further example of the personalisation of customer interactions is a fintech that relies on data-based and dynamic addressing of debtors in the event of late payment by choosing specific sentence formulations, the style of address or the communication channel based on an automatically generated customer profile. This mode of address is then adjusted to the customer's response.

[165] As an equivalent to the German Chamber of Commerce or the Federal Gazette, Infogreffe offers the possibility of finding out a lot of information about companies online.

In addition, fintechs in particular are establishing themselves as aggregators or platform providers at the customer interface. The business models of these providers exploit the relatively simple portability of some financial products. This is the case, for example, with business models that optimise call money investments. These interest rate platforms generally operate as call money aggregators and work together with a partner bank to allow their customers to invest call money with various different banks. The partner bank of the interest rate platform invests this call money with various banks without the customer having to open an extra account. The customer only has to go through the identification process once – with the partner bank of the interest rate platform. For incumbents this means the risk of losing customers.

Despite the potential that BDAI offers at the customer interface, no BDAI-triggered growth is expected short term in the market as a whole and in overall banking revenues. Rather, due to higher levels of competition at the customer interface, it could lead to a redistribution of earnings.

Data from payments and account management could prove valuable particularly for incumbents and bigtechs. For fintechs the main focus is on income from payment transactions.

Payment transactions play a key role in generating income.[166] They furthermore represent a major part of the customer relationships, as the number of interactions with the customer is relatively high compared with other financial services. Payment transactions thus contribute to increasing the amount of high-quality data held. For instance, the number of transactions made by card in particular has seen strong growth in the last few years in Europe, although there are significant regional differences.[167] Transaction-based data can also supply providers with the basis for personalised and individualised offerings beyond banking services. Currently, access to this data varies widely by provider group.

In their role as the main banks, traditional banks find themselves well-positioned to leverage transaction data at the customer interface. They already have a comprehensive overview of the transaction data of all their existing customers.[168] Unlike other providers that have no historical database, traditional banks can conduct cross-sectional as well as longitudinal analyses based on transaction data,[169] which gives them extensive possibilities for optimising the customer interface. In this way, they can personalise offers to address individual customer needs even more effectively – for example in the form of a "next best offer" based on similar customer profiles.[170] Earnings could also be increased in this way, especially if the bank knows how solvent its customers are because it could then sell them higher-quality products and services.

Furthermore, it is also possible that banks will employ BDAI to analyse transaction data in the small to medium-sized enterprise segment with the aim of identifying potential new corporate customers. This could be done, for example, by conducting a network analysis of customer relationships or a reverse search. This would allow the bank to identify potential new customers that are similarly structured to the bank's particularly profitable customers.

---

[166] In 2016, payment transactions generated global income amounting to $1.2 trillion, and this figure is forecast to grow in the next ten years by an average of six percent per year globally; cf. Badi et al., 2017, Global Payments 2017: Deepening the Customer Relationship, p. 5. Available online: https://www.bcg.com/publications/2017/transaction-banking-financial-institutions-global-payments-2017-deepening-customer-relationship.aspx, accessed on 30 January 2018.

[167] In the EU in 2016 alone, the number of card-based transactions rose by 12.2 percent to 59.6 billion. However, the percentage share of card-based transactions varied by country, ranging from below 20 percent right up to 81 percent; cf. the European Central Bank, 2017, Payments statistics for 2016. Available online: https://www.ecb.europa.eu/press/pdf/pis/pis2016.pdf, accessed on 23 January 2018.

[168] As already stated, it is expected that PSD 2 may have a detrimental effect on this competitive advantage in the near future.

[169] The prerequisite is a sufficiently large customer base.

[170] The bank determines the offer best suited to the respective customer. It is based on the customer data available and takes into consideration the interests and requirements of the customer as derived from the data.

Transaction/payment data could be of particular value to bigtechs in bolstering their core business. This data could supplement existing data on customer preferences with knowledge of the customers' ability and willingness to pay. Bigtechs could thus monetise this data within their own business models.

Payment services are also an attractive business segment for fintechs. This becomes evident when considering, for instance, investments in fintechs offering payment services: every year since 2011, 20 percent or more of all fintech investment rounds have targeted the business models of payment service providers. Fintechs could also monetise the data from this business.

BDAI can also be used to analyse the customer's willingness to pay in even greater detail. But for customers, this entails the risk that some providers may attempt to excessively extract the consumer surplus[171] based on the additional information they have on their customers' ability and willingness to pay. Indeed, this gives even providers of non-financial services with platform-based business models the possibility of extracting the consumer surplus. On the other hand, the greater transparency surrounding offers could help customers increase their consumer surplus, for instance by making use of call money aggregators or loan comparison portals, provided these portals actually are independent providers.

Automation potential offered by BDAI at the customer interface could improve the customer experience and raise efficiency potential at the same time.

BDAI allows already automated processes to be optimised and the corresponding fields of application to be expanded. For example, in addition to automated investment recommendations in the product segment of investment/savings, BDAI-optimised robo-advisors at the customer interface could also offer automated, dynamic portfolio management based on customer preferences. This could simultaneously lower operating costs. Similar effects could be realised by using chatbots: For instance, some non-European banks can already answer up to 80 percent of all customer enquiries using chatbots, processing the enquiries more cost-efficiently. At the same time, customers enjoy the advantages of a round-the-clock service.[172] Moreover, regulatory requirements can be realised more efficiently at the customer interface (e.g. MiFID[173] II) if BDAI applications are used.

In what is already a widely automated lending business for retail customers, BDAI could simplify the process particularly for more complex lending products such as construction financing, and thus contribute to faster availability of the service. For instance, there are already initial offers that permit faster financing enquiries by using BDAI and creation of cross-channel, partially completed application forms in order to speed up the process significantly. For example, there are mobile applications that compile additional information on the property. Processing the necessary documents, such as salary slips, can also be considerably simplified and/or automated.

---

[171] Cf. explanatory notes in footnote 63.

[172] Cf. for example Crosman, 2017, Chatbots to humans: Move aside, I got this. Available online: https://www.americanbanker.com/news/chatbots-to-humans-move-aside-i-got-this, accessed on 23 January 2018. Furthermore, Juniper disclosed in a report that the average customer interaction was four minutes shorter than in a conventional call centre, cf. Juniper, 2017, Chatbot Conversations to deliver $8 billion in cost savings by 2022. Available online: https://www.juniperresearch.com/analystxpress/july-2017/chatbot-conversations-to-deliver-8bn-cost-saving, accessed on 23 January 2018.

[173] Markets in Financial Instruments Directive.

With small and medium-sized corporate customers, especially in the customer segments for liberal professions and small businesses, the greatest impact would probably be in standard loan products (cf. "Credit Decision" use case) and in investment and savings products. In these segments, as in the retail customer business, the number of transactions is typically comparatively high and the processes tend to be more standardised so that greater automation could contribute to improved efficiency as well as a better customer experience. There is potential for automating administrative tasks across products and segments, for example manual transfer of information from forms and customer slips where these are still processed by front office/sales (cf. section "Core Processes" on the potential for automating administrative processes in the product platform). The options presented for more efficient front office processes courtesy of BDAI would complement another current trend of restructuring front office processes (e.g. reorganising the channel makeup of branch office, mobile office, online and telephone, including hybrid approaches).[174]

5.2.4   The impact of BDAI on the core processes of the product platform

In terms of the product platform, providers with efficient core processes or high-quality services enjoy a competitive advantage.

In providing the product platform, profitability can typically be realised through high levels of standardisation and scaling. Alternatively, it can be attained by offering services of extraordinarily high quality that can only be produced with the provider's specialised know-how. Other providers, on the other hand, concentrate on individual parts of the value chain which they can then scale. Considering their scaling skills, bigtechs could conceivably become active as product platform providers, although the typical bigtech goals mentioned above and their current market positioning currently do not speak for such an offering.

Using BDAI could generate sustainable efficiency potential in (core) processes involving large quantities and standardised procedures. Most of the potential is found in settlement and compliance processes.

The potential for efficiency gains is typically particularly high in segments involving high numbers of sufficiently similar processes. This structure is usually found in the retail customer segment and among small and medium-sized corporate customers with what tend to be standardised products. Several application areas are presented below, areas in which BDAI usage promises different levels of efficiency potential.

Significant potential could be secured particularly in the middle and back office-based processing tasks for various products and customer segments because these tasks are comparatively homogeneous for high quantities. Potential is particularly high where processes can be automated using BDAI applications. For example, robotic process automation (RPA) applications are currently used in many areas. However, these are often limited to highly rule-based scenarios where both input and output are available in a comparably structured way. This reduces the number of processes that can be addressed. By using BDAI, the scope of RPA could be increased and less structured input processed, because the BDAI applications would prepare it for the use of RPA. BDAI applications can provide, for instance, the option of machines reading digital/scanned data aided by optical character recognition (OCR) applications and then classifying it automatically (e.g. birth certificates, pay slips, credit reports). [175] This then lays the foundation for an RPA-based process. Examples of such processes would be the application process[176] and the process of setting up contracts and accounts.

---

[174] European Central Bank figures reveal a decline of 13 percent in the number of branches in the Eurozone between 2012 and 2016; European Central Bank, 2017, EU structural financial indicators. Status as of 29 December 2017. Available online: http://sdw.ecb.europa.eu/reports.do?node=1000002869, accessed on 22 January 2018.

[175]

[176] Referring here to the middle and back office part as opposed to the front office part addressed in the section "Customer Interface".

Further efficiency gains could arise when reconciling the accounts of small and medium-sized customers.[177] This could be automated even more by using BDAI to allocate deposits not directly linked to receivables, such as bank transfers with no clear reference, partial transfers or transfers relating to more than one invoice. Furthermore, account reconciliation could be automated across several accounts or even banks.

On the product side, processing retail customer loans such as consumer and car loans as well as construction financing (this in the medium term due to the higher complexity) could offer the greatest efficiency potential.[178] Similar potential could be realised in the processing of simple corporate customer loans. In many cases, efficiency gains can be secured through cooperation ventures with specialised companies. Middle and back office processes involved in payments tend to be very highly automated already so that any efficiency potential arising from BDAI usage can be expected to be rather low.

Furthermore, BDAI shows efficiency potential in compliance processes. The main areas for application are in the prevention of money laundering and in ensuring compliance with sanctions. BDAI applications could help to analyse transaction and communication data in order to identify any irregular patterns as potentially suspicious (see "Compliance" use case). For example, one large universal bank is cooperating with a fintech in order to use BDAI to help streamline and speed up its compliance processes, especially its anti-money-laundering processes.

BDAI applications could yield efficiency potential in other compliance processes, for example in those for payment transactions and in particular in the credit card business. BDAI has been used in this product area for quite some time now to identify fraudulent behaviour. For example, it is possible to combine data from credit card customers and transaction data from retailers with existing fraud and non-fraud cases to continually re-evaluate the risk of fraud for card users. This raises the number of hits and could potentially lead to more fraud cases being reported to the law enforcement authorities.

Besides these two large areas in which BDAI can help improve efficiency, there is a wide range of other options, for example in assessing loan agreements or migrating data to new data formats and structures. In addition, using BDAI for various cross-product and cross-segment analytical processes could help prepare data and information automatically rather than manually: for instance, borrower units and groups of associated customers could be developed and validated using BDAI.

BDAI applications could offer the potential to improve effectiveness particularly in the areas of compliance and risk assessment.

It is to be expected that the above-mentioned efficiency potential is greater than the impact resulting from improvements in effectiveness. This is partly due to the fact that the models currently used in the area of risk assessment are already very well advanced and partly that a large number of suspicious cases are already being identified in the compliance processes. Nevertheless, for both areas it is more important to identify the risks involved in large transactions than to process them efficiently – potential uses of BDAI are thus of relevance.

---

[177] How BDAI is used to grant corporate loans is explained more fully in the use case "Automated Loan Decisions in Corporate Customer Business". In the following, examples are given at this point of cooperation between providers at the customer interface and product platform providers.

[178] Cf. Chapter 5.4.3 for the assessment of illiquid assets.

In addition to the above-mentioned efficiency gains, improvements in the effectiveness of compliance processes are also conceivable. BDAI could probably further increase the quality of the hits: There would be fewer false hits and it would be possible to detect and track previously unidentified patterns. Furthermore, network approaches could help to optimise the processes both between individual banks and through one central authority.[179] If institutions were to share more information on possible cases of suspected fraud or any fraudulent behaviour observed, they could benefit from each other's findings. In addition, some fraud patterns might not be detected by looking at an individual institution, while a consolidated view would stand a better chance of identifying them.

The use of BDAI promises effectiveness potential with risk assessment models for lendings products: The probability of loan default for a customer could be accurately predicted on the basis of their transaction data for instance. However, over the past several decades considerable analytical and data-based optimisation measures have already been introduced precisely here so that the quality of assessments is already relatively high. These models could be further honed using BDAI. Improvements in the quality of assessments could have an impact on the capital requirements. More precise assessments could mean that individual, more risky deals that would not currently be signed could be concluded so that additional capital backing would be required. At the same time, changes could be made to the risk assessment of individual borrowers in the existing portfolio, whereby no statement can be made at the portfolio level on the trend toward an increase or a decrease in capital requirements.

5.2.5 New business models through BDAI

The "monetisation of data" is a new potential business model that cannot replace but could supplement earnings from the traditional banking business.

In addition to the impact of BDAI on banking described above, several individual new BDAI-based business models are conceivable, especially for traditional banks. This applies in particular to the monetisation of data as an additional source of income. We can distinguish between various approaches, provided they are legally permissible and customers have given their consent (see Chapter 4.1). First, providers could monetise aggregated, anonymised information on certain transaction and master data, for example information on payment transactions (number of and amount involved in the transactions).[180] Second, and assuming the customers have given their express consent and upon their request, a provider could offer, for instance, to make legitimation data available to third parties in other industries. Third, another possibility is offering data sets with anonymised customer profiles drawn from transaction and master data. These could then be supplemented with information on creditworthiness or the like. Initial examples can already be found in the banking sector for all three types of business model, all at varying stages of maturity.

Although the monetisation of data presents banks with a potential additional source of income, it is not to be expected that these earnings can replace traditional banking revenues in the short term. Providers must also weigh up the value added by monetisation against a potential undermining of consumer trust (cf. Chapter 4.1), whereby any damage to this basis of trust caused by any individual market participant could affect the entire industry. Incumbents potentially stand a chance of playing their "trust advantage" card by offering what consumers see as the controlled use of data under the auspices of a trustworthy institution (cf. the "Monetisation in Banks" use case).

---

[179] For example, some countries already have utility approaches in operation, where one central authority takes on the services relating to compliance processes; cf. for example Pulley, 2016, Know Your Customer. Banks place trust in KYC as a Service. Available online: https://blogs.thomsonreuters.com/financial-risk/know-your-customer/banks-place-trust-kyc-managed-services/, accessed on 30 January 2018.

[180] On this basis, it would be possible to issue statements on suitable locations for different companies, to determine the success of advertising campaigns or to track how many people frequent the business at different times during the day.

In addition to the monetisation of data as a new business model, market participants could offer individual sections of the value chain to third parties. As mentioned earlier, traditional banks could leverage their regulatory competence and knowledge of the industry as well as their customer reach to make white label products available to other providers at the customer interface.

5.2.6   Use cases in the banking sector

Bank-related processes are certainly relatively strongly characterised by rule-based and automated procedures, and the introduction of information technology has already produced huge amounts of data, but it is only now that banks are beginning to exploit this data on a greater scale thanks to BDAI support. As a corollary, greater efficiency potential can be expected in the loan decision process, for example. In the same way, BDAI applications can aid the compliance process as already stated, especially in making money laundering checks more efficient and effective. Furthermore, there are possibilities of monetising the data.

Chapters 5.2.6.1, 5.2.6.2 and 5.2.6.3 explore the following three use cases in more detail:

Automated loan decisions in the corporate customer business. This first illustrative use explores how BDAI applications can be used to extend the automation of credit decisions already in place in the retail customer business to cover simple corporate customers, too. In this way, BDAI could help to automate the process from credit application to credit decision in less complex cases in the corporate customer business.

Optimisation of compliance based on the example of money laundering investigations. This example highlights the potential that BDAI applications offer for core banking processes. Often, after a manual check, the automatically generated reference flagging a suspected case of money laundering proves not to be worthy of further investigation. It is conceivable that BDAI processes could help raise the quality of the hits, independently identify undetected patterns and significantly reduce the number of false alerts.

Monetisation of customer data by banks. This illustrative use case shows how transaction data can be used as a new source of income, thus opening up new business opportunities to banks in particular. Using BDAI applications, banks can evaluate the data available in-house in order to create personalised offerings, which permits banks to position themselves as an intermediary between customers and retailers. Retaining customer data at the bank and simply passing on aggregated and anonymised data to retailers could be the key to ensuring data protection while generating additional income.

5.2.6.1 Automated credit decisions in the corporate customer business

1. Introduction

The automation of credit decisions is not a new process. As early as 1997, the authors of the book "Zukunft der Universalbank" [The Future of Universal Banking] wrote about an ATM in Washington that grants small loans in only a few minutes. At that time, the authorisation of credit card sales had already been automated as well.[181] Using BDAI, this rather deterministic type of automation employed in the simple, more granular and more homogeneous consumer credit business can now be further developed to suit the more complex, yet still small-enterprise corporate lending business.[182]

Automation of the credit decision means that a machine makes the final decision to grant a loan on the basis of input parameters and decision-making models. If this is not completely feasible from a legal point of view, at least the credit (decision) process can be automated as far as possible. Nonetheless, there will always be exceptional cases that do not qualify for automation and will have to be processed manually.

BDAI is used in this process to achieve a better and simpler cash-flow-based analysis using, for instance, transactional data. This eliminates the need for balance sheet analyses, a process which tends to be more time-consuming. In turn, fewer documents and data points are required in the application process, and the process becomes shorter. At the same time, a better understanding of the cash flows reduces dependence on complex assessments of collateral.

As a result, it is no longer strictly necessary to interact with a customer adviser, and the approval (or denial) can be transmitted through digital channels after only a short time. In addition to increasing the speed of the credit process, structured use of known or additional data and AI also improves risk differentiation.

Besides speeding up the process in the interest of the customer, this development is driven in particular by the reduction of internal process costs, particularly in the small-scaled yet unit-driven SME business[183]. This process is therefore not currently applicable to asset-based lending, for instance in the high-volume commercial real estate lending business.

---

[181] Cf. Böhme, 1997, Die Zukunft der Universalbank. Strategie, Organisation und Shareholder Value im Informationszeitalter. Wiesbaden: Deutscher Universitätsverlag, p. 118.

[182] This has to be differentiated from regulatory risk models, which do concern capital requirements, yet are only indirectly related to credit decisions. These are not the object of this use case. As of today, models that need to be authorised by the supervisory authority (internal models for companies using the internal ratings based approach –IRB) have not yet paired big data and artificial intelligence. If BDAI were to be used in a supervisory and regulatory context, several questions should be considered that reach beyond these use cases.

[183] Small and medium-sized enterprises.

2. Application

BDAI can be used for less complex cases in the corporate customer business to automatically process loan applications up to the credit decision. In a streamlined credit process, the algorithm uses available customer data to dynamically and adaptively calculate the customer's credit rating as well as a maximum credit limit.

Figure 17: Use of BDAI in the credit decision
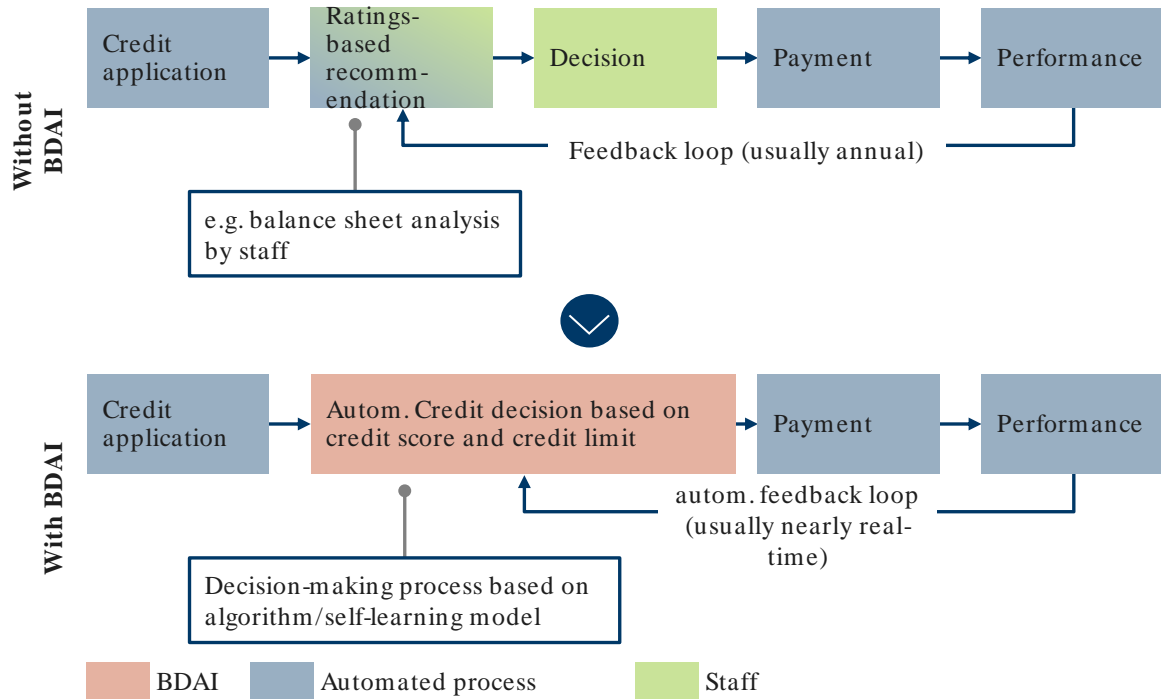
The analysis is based on a broad set of transactional customer data, including for instance the customer's credit history, current account balances, types of bookings and reasons for the respective bookings (e.g. salary and dividend deposits on the one hand and penalty fees and returned direct debits due to insufficient funds on the other).

Essential parameters for using BDAI are

■ Data basis: A broader data basis provided by the customer gives better insight into payment flows or the customer's behavioural pattern, yields a higher capacity to achieve resilient results by using BDAI, and provides a greater ability to process a loan application based purely on cash flows. Principal banks still have an advantage in this respect. However, this could change following the introduction of PSD 2.

■ Degree of standardisation of collateral structures[184]: BDAI can only be used efficiently if this area is sufficiently standardised. If the process is to run automatically, only collateral that does not at that point require manual assessment can be included. However, this does not limit the field of application solely to unsecured loans, loans secured through cash or guaranteed loans, or to physical collateral that can be evaluated in a standardised way, like motor vehicles and construction machinery. If there is physical collateral that has already been assessed, it can certainly be taken into consideration by automatically adjusting the collateral value.

■ Degree of standardisation of contract structures: The trend toward standardisation has already been going on in this area for some time now. Individual terms and conditions (e.g. lender guarantees) in the contract structures, in particular, are decreasing. This is usually easy to implement, as complex contractual clauses were frequently only included in the contracts at the instigation of the bank.

Dynamically and adaptively means that the algorithm does not depend on a defined data set being completely available and instead can work with the data digitally available, even if this data is incomplete. If there is already sufficient information available, no new data has to be collected manually. The quality of the collateral evaluation is assessed based on the type of data and how up-to-date and complete it is. For a low-quality collateral evaluation, the maximum credit limit is reduced by a safety deduction. The customer could then provide additional data upon request by the bank, such as transaction data, and thus reduce that deduction and potentially increase his credit limit. The additional data provided could serve to either increase or reduce creditworthiness.

With regard to the scope of the data requested, the bank can lower the amount of time and effort required both for the bank itself and for the customer by making an additional request for data only if the evaluation confidence calculated on the basis of existing data is not high enough for the desired loan amount.

In future, this principle of focusing on a cash-flow-based analysis can be supplemented by using additional BDAI components not yet in use today. In particular, this refers to the inclusion of additional relevant characteristics, some unstructured, such as:

■ Text mining in annual reports and annual financial statements as well as in business correspondence

■ Analysis of collateral provided using externally available data, such as traffic frequency in certain locations, and the valuation of more illiquid assets

■ Behavioural patterns during the online application procedure, such as care in reviewing documents/GTCs and test entries

These measures could serve to significantly increase the scope of information. Text mining could, for instance, also involve counting the frequency of several tens of thousands or even several hundreds of thousands of words and word combinations per document. Words like "collections" with debit amounts would have a negative impact, while "rental income" or "dividend payment" with credit amounts would have a positive impact.

---

[184] The importance of this driver of complexity is expected to decline with further development of BDAI.

This would make it possible to address not only the elements that had already been automated, like balance sheet analysis or the calculation of decision-making parameters (e.g. loan-to-value), but also more complex matters like a simple verification of the pledged collateral (land charge, vehicles, etc.) and the analysis of an economic corporate structure in terms of borrowers and groups of associated customers.

3. Technology and data

Analysis of transaction data on the level of individual bookings

The analysis is based on the evaluation of the customers' granular transaction data available at every bank down to the level of individual bookings including memo fields (if permitted by law), counterparty account numbers and the timeline (as well as the changes over time shown there). The information obtained from this can efficiently predict credit rating changes and, in terms of up-to-dateness as well as granularity and digital availability, is traditionally better than the data sources used traditionally to make the credit decision, like balance sheet and income statement. Many special topics that are checked for in traditional credit decision processes, such as dependence on single suppliers, customers and currency risks, can be monitored automatically. In many cases, a complete balance sheet analysis is only necessary for assessing credit rating at much longer intervals. The automation of the credit process is thus already widespread in the lending business with smaller corporate customers. Internal test calculations of individual banks show that it can also be used for medium-sized and large corporate customers.

Effects of PSD 2

The larger the share of all customer transaction data available, the more reliable the results and therefore the risk differentiation. It is this that makes the introduction of PSD 2 and its incorporation into national law especially relevant. This directive allows banks, with the consent of the customer, to analyse the transaction data of other banks, thus further increasing data availability. Although PSD 2 thus makes automation easier for banks, it also creates additional competition from fintech companies and non-banks because it applies well beyond the borders of the banking industry.

In addition to transaction data, the broadest possible dynamic database is created using data from other internal and external sources. And dependent on availability at a national level, public records are also becoming increasingly important. Customer consent is always a prerequisite, and compliance with the provisions of the EU GDPR with regard to transparency and intelligibility is particularly important in this regard.

The fact that external data is being used should be related to the credit applicant in a transparent and easily comprehensible manner and must depend upon the customer's approval (cf. Chapter IV and Chapter VI). If denying approval entails an increased interest rate as a result of higher risk premiums, the bank must notify the applicant.

Customer incentives for data disclosure through value-added services (e.g. automated account settlement functions, transaction analysis tools) or benefits in terms and conditions are also used to expand the data set. On the one hand, the forecast quality of the algorithms used can be increased by assembling a data set that is as complete as possible. On the other hand, ongoing plausibility checks of the available data through cross-checking improves the model quality.

Because the risk differentiation achieved in this way – due to the use of numerous data points and sources, and the multidimensionality of the analysis – is generally much more objective than the individual subjective assessment of customer advisers, their more expensive input is being used less frequently. Comparative calculations indicate significant advantages in selectivity and advance warning.

A variety of analytical methods are used to analyse the data. However, banks do not yet use self-learning algorithms in production. This is due to their more complex replicability and the compliance requirements related to this. In parallel to the analytical, linear models used in production environments, test environments already employ self-learning algorithms in challenger roles with the aim of continually monitoring the efficiency of the production environments.

4. Opportunities

For the vast majority of customers, automated credit decisions will drastically reduce the administrative burden and waiting times until a binding approval or denial of the loan is issued, improving the quality of the customer interface. In addition, better risk differentiation results in better identification of customers at risk of excessive debt, thereby protecting the bank against over-indebtedness by denying the loan.

Increasing efficiency and effectiveness in credit decisions

From the bank's perspective, automating the process from the credit application to the credit decision will make the core process of lending more efficient and effective. For less complex corporate customers, banks will also rely on digitally available data, and the manual work from the credit application through the decision will be greatly reduced.

In addition, a more detailed analysis of transaction data using more up-to-date data will serve to improve risk differentiation and early warning functions – meaning better protection against credit default. Negative developments that are overlooked or seen too late in the balance sheet or income statement due to their lack of granularity or the delay associated with preparing the balance sheet will become evident and can be included in the decision-making process.

From the market's perspective, market transparency improves. Because there is less time and effort involved for both the customer and the bank, it becomes much easier for customers to gather and compare "realistic" offers, i.e. those that have been relatively thoroughly checked.

5. Risks

For the customers, widespread use of the method mainly entails the risk of discrimination. By omitting the manual examination during further steps of the credit process, it is possible, when using BDAI, that certain customer groups have but limited access to traditional bank loans or are only offered loans at relatively higher costs. While algorithms can contribute to more objective processes and decisions, all of those instances with little representation in the underlying data set would still necessarily lead either to denial of the corresponding loan request or handling in a more manual – and therefore from the bank's perspective more expensive – process. Although there is this risk of discrimination in a manual credit application process, there is no opportunity in an automated credit application, however, to relate to a bank adviser any circumstances that could result in a loan being granted despite parameters indicating a negative outcome.

At the same time, subjective (price-driven) and objective (access to the product at all) pressure on the customer to disclose private or business data classified as highly confidential can increase in order to gain access to a loan. If the customer decides to disclose data, it should be ensured that the use of sensitive data is restricted to the lending process.

BDAI lowers barriers to entry in the lending business

For banks, competitive pressure grows when new participants enter the market because BDAI lowers the entry barriers to the lending business – algorithm-based lending processes are more easily transferable. In the traditional lending process without BDAI, experience with lending was deeply embedded in the DNA of the banks, for instance through process documentation, credit risk strategies and the experience and assessment skills of the employees. Through BDAI, lending experience becomes a value that can be generated through technology. This is further exacerbated by PSD 2 and the related access to transaction data, provided with customer consent. The existing relationship of trust between the customer and the bank should not suffer, for instance as a result of the misuse of data or even data usage that is merely perceived by the customer as misuse of data.

In addition, when a high-volume business is automated, model errors have a faster and comparably more severe impact than is the case for human mistakes. Incorrect assumptions, lack of data quality, exceeding model limits or improper use of the model can quickly result in negative consequences in the form of unintended credit decisions. This is no different in itself from the manual status quo in which an unsuitable credit risk strategy can have a negative impact on the portfolio. However, due to the higher processing speed, BDAI can on the one hand increase the demands on the reaction time. On the other hand, there is no more natural distribution of decisions due to their evaluation by different loan officers; using BDAI, all decisions are made by "one" decision-maker. This situation can be mitigated by applying strict governance that ensures, for example, that the process can be explained in a simple and robust way (cf. in particular Chapter III).

The following example uses a game theory approach to show how small differences in risk differentiation can impact competitiveness.

Example: Risk differentiation and competitiveness through credit assessment models

We are considering a hypothetical application portfolio of consumer loans. The average default rate after one year is three percent. For the sake of simplicity, the number of the individual loans is assumed to be one percent – the same as the loss given default. Two competitors are using different price models to compete for customers from this portfolio.

In the following, we look at two credit assessment models that the competitors are using to estimate the probability of default for individual customers: Model A is the base model with an AUC[185] of 0.8625. Model B includes slight improvements and has an AUC of 0.87. There is a moderate difference in the quality of the models which can be achieved, e.g., by changing the modelling from a linear to a non-linear method like gradient boosted trees or neural networks.

We now look at the effects in direct competition between two providers – comparing offers on loan comparison websites, for example. In each phase, both providers recalculate the probability of default for each customer in two steps. First, they calculate the model score to arrive at a ranking of the customers. Then, the dependence of the probability of default on the model score is recalculated based on the customers in the previous step.

---

[185] The area under curve (AUC) is a typical quality indicator that documents how well default customers can be differentiated from non-default ones. For this, the area below the receiver operating characteristic curve (ROC) is calculated. This curve represents, for classification problems, the relation between the rates of true positives and false positives.

The rules for this thought experiment are:

- In the thought experiment, the providers' prices differ only by the different premiums for the expected loss, which is proportional to the probability of default.

- Customers always choose the lowest interest rate offered.

- Only customers with a probability of default below ten percent receive offers.

In phase 1, both providers use model A based on a constant history. This results in identical pricing, and the two banks share the market. The expected rate of default (PD – probability of default) and the realised default rates (DR) match exactly.

	Market share	PD	DR
Denied	7.95%	20.76%	20.34%
Model A	46.02%	1.50%	1.50%
Model B	46.02%	1.50%	1.50%

In phase 2, provider 2 switches to the slightly improved model B. In direct competition, model A underestimates the default rate for the acquired customers by more than one percent, while model B overestimates the default rate. The market shares shift dramatically at the same time. Model B gains 38 percent more customers in relative terms. The bank using model A gets the customers at the bottom of the barrel.

	Market share	PD	DR
Denied	6.68%	22.91%	22.91%
Model A	29.53%	2.11%	3.14%
Model B	63.78%	1.24%	0.85%

In phase 3, both banks notice that their models are not calibrated correctly and they adjust the default rates – without adjusting the underlying score model. Both banks know that they cannot calibrate based on the customers acquired in phase 2 alone. They therefore choose a conservative approach in using the maximum probability of default based on the phase 1 and phase 2 calibrations. As a result, the market share of model B once again grows significantly. Model B now only slightly overestimates the default rate. Model A still underestimates the default rate by 0.5 percent.

	Market share	PD	DR
Denied	6.68%	22.14%	22.91%
Model A – Phase 2	11.93%	1.61%	2.11%
Model B – Phase 2	81.39%	1.58%	1.50%

The sample calculation shows:

■   The quality of the credit assessment model can be essential to a bank's competitiveness, particularly in highly competitive markets.

   o   The bank with the better model gains considerable market share.

   o   The bank with the poorer model underestimates the losses significantly, causing the profit margin per customer to fall or even become negative.

■   A new model introduced by a competitor can influence the results of other models in the market, as the data from the past is no longer representative.

Particularly the last point challenges traditional models based purely on the evaluation of historical data without considering market dynamics and game theory approaches.

It should be noted that, in actual fact, customers are not purely rational in their decision-making, and factors like marketing, sales and loyalty play a major role. Moreover, the bank has an inherent competitive advantage with its existing customers due to its information lead through previously collected data on payment flows and behaviour. Market participants are at least partially aware of their competitors' activities.

6. Supervisory and regulatory issues

The aspects depicted in this use case involve various supervisory and regulatory implications.

BDAI models can be used for matters relevant to regulatory authorities, raising the question of how these models are authorised and monitored.

At the same time, standardisation of the credit checking process will lower the entry barriers for new market participants in the banking industry, leading to more competition, but also to more interconnectedness and disaggregation of the value chain.

The automation of credit decisions using BDAI also has implications for consumer protection with regard to the discrimination of individuals or marginal groups. Moreover, consumers or corporate clients could feel pressured to release sensitive data in order to obtain a loan in the first place or to obtain better conditions for the loan offered.

Increased use of BDAI models also raises further questions concerning model transparency/traceability as first referenced in chapter III.

Chapter VI addresses the supervisory and regulatory implications of the points mentioned in more detail.

5.2.6.2 Optimising compliance: the example of investigating suspicions of money laundering

1. Introduction

Money laundering is a global and large-scale problem. The total annual volume of money laundering in Germany alone is estimated to range from €50 billion to €100 billion.[186] Banks process national and international payment transactions, making them central players in the fight against money laundering. When investigating suspicions of money laundering, banks are legally required to identify suspicious payment flows and report them to the authorities. Methods using BDAI can increase efficiency and effectiveness in this context.

Banks currently generate thousands of automated indications every day on the basis of detection rules defined by experts. These indications are then manually examined, and only a small share prove to be worth delving into further. One of the reasons for this is that individual detection rules can often only recognise hints of suspicious transactions that also apply to numerous other normal transactions. Because professional money launderers are continually improving their methods, and the detection rules are only adapted to these changes retrospectively, there is a danger that many actual money laundering transactions remain undetected.

Investigating suspicions of money laundering involves analysing large volumes of data from customers' transactions to detect conspicuous behaviours. BDAI methods can help to recognise previously undetected patterns, like inconsistencies in amounts and regular transactions below threshold amounts, independently through self-learning, and increase the effectiveness of existing detection rules by means of feedback loops. A few banks are currently introducing such applications in a test phase, i.e. they are mostly being used alongside the existing, partly manual process. Initial market tests indicate that the number of false reports is dropping and that, simultaneously, the quality of the remaining indications can be significantly improved.

2. Application

Automated support in combating money laundering normally involves many rules created on the basis of expert opinions that are aimed at identifying money laundering and used to monitor transaction data – together with other known customer characteristics – in a first-level process. As soon as a transaction is classified as suspicious based on these rules, a bank employee must manually handle this suspicion in a second-level process and decide whether the suspicion is confirmed and subsequently requires sending an official suspicious activity report (SAR) to the authorities for further evaluation.

---

[186] Cf. Bussmann, 2015, Dark figure study on the prevalence of money laundering in Germany and the risks of money laundering in individual economic sectors, p. 4.

This second-level process can be extremely work-intensive for the bank, particularly when the automated monitoring conducted during the first-level process generates a large number of reports that are ultimately determined to be false. Processing these in the subsequent manual second-level process involves a good deal of unnecessary work (cf. Figure 18 showing a diagram of the use of BDAI).

Figure 18: Optimising the investigation of suspicions of money laundering through the use of BDAI
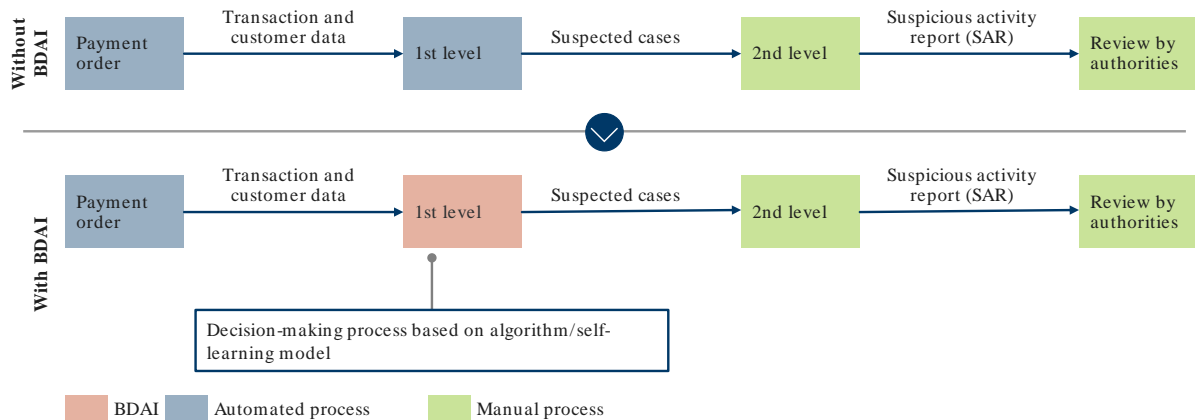
The rules to detect money laundering can usually be distinguished as either absolute or profile-related detection rules. Absolute detection rules are based on threshold amounts, payment destinations and reason for payment, for example, while profile discrepancies are generated on the basis of differences from the customer's historical profile and differences compared to a benchmark group.

Absolute detection rules are easy to bypass and result in unfounded technical indications particularly in high-volume business relationships. Profile-based detection rules are particularly susceptible to seasonal effects, a lack of comparability or representation in benchmark groups and shifting of business activity among customers' main banks.

BDAI can help improve the effectiveness of threshold amounts through sensitivity analyses, generate more selective profile analyses, covering multiple periods if necessary, and recognise previously unknown patterns (anomalies in amounts, regular transactions between payment initiators and recipients that initially appear to be independent of each other).

In addition to reducing the number of unfounded suspected cases, it can also serve to improve the quality of the remaining findings of the first-level process. Targeted filtering of suspected cases can increase the processing time available per case at the second level. This can serve to further increase the quality of the findings at the second level.

3. Technology and data

The first step in standardising the process of detecting and reporting suspected money laundering and making it more efficient is to prepare various forms of data. This data includes customer data and characteristics, transaction data, data on individual payment behaviour, on the payment behaviour of the benchmark group and changes in payment behaviour (of the benchmark group or the customer), and historical data on confirmed suspicions of money laundering.

In practice, this first involves classifying transaction data – for example by the criteria "cash deposit" or "transfer to a high-risk country". These are in turn translated into variables like "total deposits in the last 30 days" and "ratio of cash deposits in the last 30 days to the last 360 days".

Next, algorithms from both supervised and unsupervised learning can be used – the former especially in order to identify patterns from past experience in the current data sets, the latter especially to identify patterns and anomalies in the data sets that have not yet occurred (cf. Chapter III). In this process, unsupervised learning represents a particular improvement over purely rule-based models because it is able to detect indications of fraud or money laundering patterns that have not yet occurred.

The resulting suspicious cases then undergo additional manual analysis by staff specialised in this area; if the suspicion is confirmed, an SAR is reported to the authorities for further evaluation and follow-up.

To optimise the recognition of patterns in money laundering, supervised learning uses algorithms that are trained using past confirmed SARs. Using the most current (suspected) cases of money laundering as input for the algorithm so that it can learn and subsequently identify the corresponding patterns is therefore crucial to the quality of these automated investigations. In this sense, the computer learns as it were from many different staff members how to identify money laundering patterns in the data. Algorithms are capable of identifying patterns in complete account histories and the accounts linked with them through transactions.

4. Opportunities

The opportunities related to the method for using BDAI in monitoring transactions as described here are found in better identifying money laundering and thus achieving a higher rate of detection. Better methods will significantly increase the cost and risk of detection for criminals.

From the noncriminal customer's perspective, better systems for monitoring money laundering will result in a lower probability of mistakenly coming under scrutiny in a money laundering investigation. Such an event is very arduous for customers, as they are often required to provide additional information and documentation in order to allay the suspicion.

From the banks' perspective, the cost of the monitoring process is cut significantly given the higher rate of hits using the automated method and the resulting lower number of cases to be examined manually. Reducing the number of erroneous cases of suspected fraud is important for the bank's reputation because the customer relationship suffers significantly when a customer is targeted by a criminal investigation. In addition, it is essential that banks continually improve their money laundering prevention methods, as systematic errors in their methods result, under certain circumstances, in the imposition of large fines.

Regularly updating the detection rules by means of BDAI limits the opportunities for criminals to deliberately bypass familiar rules aimed at detecting money laundering.

5. Risks

One risk is the more complex traceability and more laborious verifiability of the BDAI models. The scope of this problem depends on the algorithm used. Traditional, rule-based methods are more transparent regarding the defined rules and derivations. The verifiability of results of other, more advanced methods of BDAI, particularly those involving unsupervised learning, is more laborious, specifically posing a problem for those falsely accused.[187] It is also important to ensure that what the algorithms learn is factually accurate.

In addition, there is a risk that significant advances in the detection quality, resulting from the use of BDAI, will cause criminals to target banks that are less developed in this area in future. This effect can be exacerbated by the fact that extensive digitalisation decreases the number of sales staff who may detect irregularities at the customer interface.

Furthermore, there are two potential challenges involved in maintaining consumer trust. First, it must be ensured that banks do not use the data and profiles collected to investigate money laundering for other purposes. Second, banks must not access data beyond their remit.

6. Supervisory and regulatory issues

The aspects set forth in this use case on monitoring money laundering have supervisory and regulatory implications that need to be addressed.

For instance, the use of BDAI can significantly improve the processes for detecting (punishable) money laundering violations. This would allow banks to deploy existing resources to investigate actual suspicions. These issues are discussed in more detail in chapter VI of this study in order to derive interdisciplinary implications for supervisory and regulatory authorities.

5.2.6.3 Monetising customer data

1. Introduction

Banks have access to data about their customers that is both highly sensitive and valuable, particularly information gathered through transactional banking, such as payments. The value of this data, including account transfers, is determined both by how current it is – transfers are usually booked in near-time – and by its veracity, such as when it provides precise information about how much a consumer is willing to pay for a product or service. These aspects make this kind of data much more valuable to the company compared to mere expressions of interest, like visiting a website for instance.

As a result, the customers' main banks[188] are able to precisely determine their customers' disposable income available for spending and investment, and also to forecast their willingness and ability to pay in future.

---

[187] This can also represent an opportunity insofar as it reduces external transparency for potential criminals, thus making it more difficult for them to circumvent detection.

[188] This is also true for other banks/non-banks with the introduction of PSD 2.

BDAI enables more precise customer segmentation

In this way, banks can segment private customers according to disposable income to be invested for instance in insurance, vehicles, travel, education or charitable donations. Furthermore, the type of transaction itself, i.e. debit note or transfer, credit card or point-of-sale payment, could provide profound insight into customer typologies. This segmentation is important information not only for the bank's marketing and customer classification, but also for third parties offering products and services. One application would be the identification of specific products that a customer is likely to purchase, for instance certain car brands, insurance policies, electrical, telephone and natural gas suppliers, or transport services at an aggregated and anonymised level.

For banks, this type of data theoretically has the potential for optimising the management of the bank's own sales as well as the sales activities of third-party suppliers, like retailers, as a fee-based service.

However, this potential to optimise sales and revenues for the banks is countered by the risk of sustainably damaging the high levels of consumer trust through unauthorized or unreflected disclosure of highly sensitive financial data (cf. Chapter 4.1). This is a real risk because many consumers consider this transactional data to be highly personal (cf. Chapter 4.1).

Safeguarding consumer trust

The actual technology and communication challenge arising from the monetisation of customer data by banks is therefore protecting consumer trust. Due to the fact that the customers are in a sense the originators of the data produced on and by them and that the right of informational self-determination is valued very highly in Europe (in Germany, even at the constitutional level), any intended monetisation of customer data always raises questions of whether customers actually and actively approve the monetisation of their transaction data, and if so, how they can share a reasonable amount of the value created from their data.

Currently, the additional income received by banks from fees generated by the monetisation of their customers' transaction data is not expected to achieve a scale even close to revenue from the primary banking business. This might be related to the fact that, when it comes to consumer trust, banks are now subject to different expectations than are big tech companies. Monetising data does not represent an effective way to significantly improve a bank's solvency, although it could make a small contribution or constitute a differentiating feature in competition with other banks, for example where fee income is used to cross-subsidise small-margin services.

The purchase price or revenue from transactions with data observed in the past in other industries has usually resulted from different types of business models, e.g.

- Sale of complete customer data sets such as those seen in the sale of data-driven platforms[189] and

- Monetisation of customer data sets unrelated to a banking relationship, yet very wide disclosure of non-bank-related data[190]

---

[189] To compare this, approximately $30 was paid per user when WhatsApp and Instagram were sold to Facebook, cf. Glikman et al., 2015, What's The Value Of Your Data? Available online: https://techcrunch.com/2015/10/13/whats-the-value-of-your-data/, accessed on 30 January 2018.

[190] To compare this, Facebook has earned per user in the USA and Canada in the third quarter 2017 approximately 21 USD, in Europe approximately 7 USD, cf. Facebook, 2017, Facebook Q3 2017 Results, p. 8. Available online: https://s21.q4cdn.com/399680738/files/doc_financials/2017/Q3/Q3-'17-Earnings-Presentation.pdf, accessed on 30 January 2018.

Transactions like these are therefore not suited as a benchmark for banks.

Initial business models – outside the banking sector – from the United States, based on the comprehensive disclosure of data by the customer (e.g. data from financial and other transactions, communication data, digital profiles, health-related data) offer a low single-digit dollar amount per month as compensation for the consumer. The share attributable to the bank-related data alone is likely to be only a fraction of that amount.

Several banks in Europe and the Asia-Pacific region have already monetised customer data in a way that tries to reconcile the conflicting goals of protecting consumer trust and generating additional income or reinforcing customer loyalty to the bank as well as boosting value creation for third parties.

It should be noted that the general rule in international banking practice has been to disclose only anonymised data to third providers. Anonymisation options either use customer groups with similar profiles (cf. Chapter III) or forgo the disclosure of individual customer's personal data, since no customer-related data leaves the bank's systems. An example is shown in the following section.

2. Application

A bank offers retailers the option of contacting a certain group of the bank's customers whose profiles match parameters defined by the retailer, for example in order to offer them a certain range of discounted products. Such profiles can be based, for instance, on regional affiliation, payment transactions to certain addresses in the past or on a certain amount of free liquidity.

Figure 19: Dual process structure of the use case
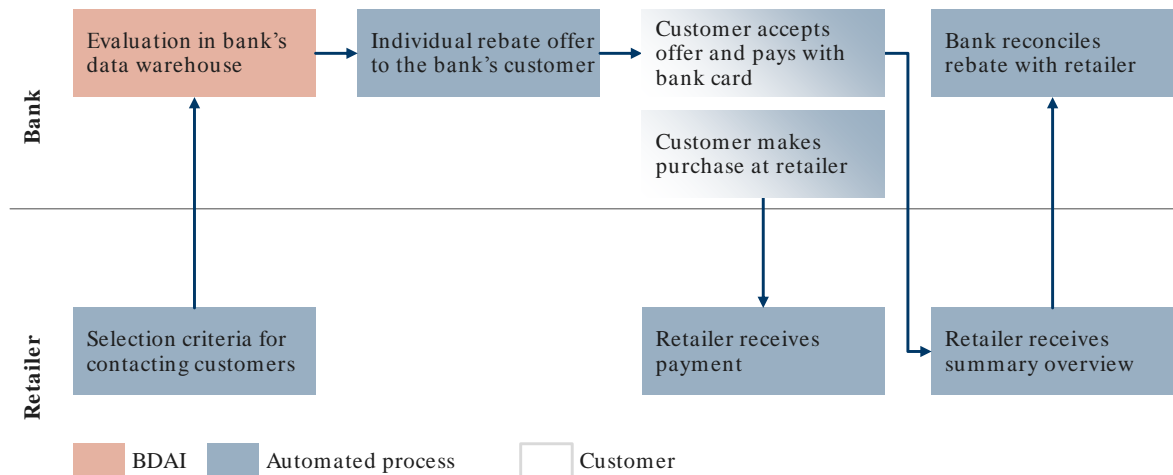
Example: How the monetisation of customer data could operate

After the retailer has prepared the selection criteria and passed them on to the bank, the bank evaluates its data sets on the basis of the selection criteria provided using the algorithms for the retailer. The bank customers who are identified receive specific rebates from the retailer through the bank's platform, e.g. its app.

The bank customer can accept this offer by carrying out a specific transaction with the retailer – without showing the app or the specific offer – using the bank's card or through his account with the bank. In this way, the retailer cannot identify the individual customer who has conducted the business transaction as a result of these marketing measures.

The bank deducts the rebate when settling the transaction by crediting the amount to the customer.

After a period of time, the retailer receives a report from the bank summarising the customers identified using the selection criteria who actually made a transaction, albeit without identifying the individual customers (cf. Privacy-Preserving Data Mining in Chapter III). At the same time, the retailer compensates the bank by reimbursing the rebates granted plus a fee.

The necessity of customer consent

Customers need to actively consent to the use of their data for this additional feature. They can also limit the service to those retailers with whom they have previously conducted business.

This model can also be extended and combined with bank products such as consumer loans for larger purchases or car loans, in order to additionally increase the bank's interest and fee income.

This case illustrates the value of the transactional customer data of banks. However, with a view to PSD 2, it is worth noting that this competitive advantage of banks over non-banks can be reduced, as the guideline also gives third parties access to transactional data –provided that the customers have given their consent for this.

3. Technology and data

This use case is based on harnessing the banks' large data warehouses. This can consist, for instance, of the structured payment history across contractual partners, payment channels
and payment locations as well as payment amounts over several past periods or the bank's limit assessment, customer master data and other banking-related attributes such as the classification of risk for investment products. This data warehouse can, of course, be either expanded by adding banking attributes, or reduced.

The algorithm can be based on a traditional, deterministic process. An even more productive approach would be to use self-learning algorithms that take into account the individual customer's past behaviour in response to each of the retailer's sales measures.

In addition to including rebates in sales, the feedback loop can also draw upon future behavioural changes of the customer such as increased visits to the respective retailer.

4. Opportunities

As the bank's customer, the consumer receives personalised offers which may contain rebates. The retailer, also the bank's customer, can individualise its customer retention and marketing efforts without creating its own database and analytics.

Better customer retention

The bank can use its position as the contractual partner to both the consumer and the retailer to improve its own revenues. Internally, it can do so by better marketing its own products, and also externally or indirectly by improving the offers of retailers. To redeem the discount anonymously, the customer must also use a bank card or bank account. This improves customer retention both objectively, by creating additional transactions, and subjectively as well, by creating a feature setting the bank apart from its competitors, for instance with regard to customers who are very digitally savvy.

From a market perspective, this means an advantage for banks over bigtechs. In addition to electronically documented and reviewed customer and transaction data, banks usually also have access to extensive historical datasets that allow digital learning to take place. Individualising their banking services can help the banks to compete with data-driven providers like big tech companies.

5. Risks

In participating in such loyalty and rebate programmes consumers risk a lack freedom of choice and a lack of information regarding and control over how their personal banking data is used. This risk is even more acute if the provision of ordinary banking services were linked to participation in programmes like these, putting the customer in a take-it-or-leave-it situation. By withholding their consent to participate in the loyalty and rebate programme, the customer could as a result be excluded unfairly from the bank's deposit and payment services, for instance. A lack of information and control in programmes like these could mean that private customer information is disclosed or used for purposes other than the usual banking services or indeed even passed on to third parties without the customer requesting this.

Potential loss of trust related to a non-transparent use of data by banks

It is important for the bank to avoid any potential loss of trust among its existing and potential new customers by making participation in such programmes as transparent, traceable and controllable as possible for the customer. Although customers do have the freedom to choose, as in the model described above, there is still a risk that banks will suffer damage to their image if data is misused or is simply perceived by the customer to be misused. Customers could also transfer their dissatisfaction with the advertised service providers, dealers and their products to the bank. In this context, it would be helpful if banks ensured full transparency by proactively communicating ahead of time to the customers participating in these programmes exactly what information about them is transmitted to third parties.

If bank customers are flooded with too many consumer offerings, they could also be encouraged to make purchase decisions beyond their actual ability to pay – a situation which would not have arisen in the absence of such an offering. From the bank's perspective, this could damage the debtor's credit rating. From the consumer's perspective, their own liquidity situation would be endangered, or their debt could grow as a result of unusually high consumption.

6. Supervisory and regulatory issues

The aspects depicted in this use case have various implications for supervision and regulation.

The monetisation of data using BDAI could have implications for consumer protection in particular, with regard to both the principles of data minimisation and the appropriation of data, and in terms of control over the use of the data. Individual customers could also experience "negative discrimination" – for both bank-related and non-bank-related products.

Chapter VI addresses the supervisory and regulatory implications of the points mentioned in more detail.

5.3 Insurance companies

5.3.1 Introduction and status quo

The use of BDAI in the insurance sector is basically still in its early stages.

The insurance sector traditionally uses large volumes of structured data for actuarial tasks.[191] There is, however, currently no evidence that BDAI technologies as defined in this study – for instance, supervised learning methods – are widely used in the global insurance market today[192]. And yet BDAI technologies are being used or tested by several providers around the world along the traditional value chain. Large insurance companies are often more advanced in this practice than smaller competitors. In Germany a few insurance companies have already launched and implemented their first BDAI initiatives mainly in motor vehicle, household and health insurance.

Two challenges have delayed market penetration of BDAI.

Two challenges that have delayed more rapid market penetration of BDAI technologies in the insurance sector are described below. The study then also highlights why BDAI have grown in significance nonetheless – and will continue to grow – in the insurance market.

Relatively little experience of dealing with new sources of data and (self-learning) algorithms

By their own account, compared to other industries, some insurance companies still lack both the expertise and experience in dealing with the technologies described in chapter III as well as an adequate number of experts on their staff. The recruitment and training of personnel represent key challenges for many insurance companies planning to implement BDAI applications.[193]

---

[191] Cf. Feilmeier, 2016, Digitalisierung, Big Data und andere digitale Techniken in der Versicherungswirtschaft. In: Fakten & Meinungen zur DAV/DGVFM-Jahrestagung 2016.

[192] This study excludes reinsurance companies, pension funds and insurance companies specialising in multinational insurance and pension solutions. Furthermore, the focus is on the end consumer (private and corporate customers). The customers of reinsurance companies are primary insurance companies and therefore, strictly speaking, neither private nor corporate customers. Moreover, only private health insurance but no compulsory health insurance is included in the scope of this study, as the use of BDAI – for price differentiation, for instance – is currently not foreseeable.

[193] Cf. Silverberg et al., 2016, Innovation in Insurance: How technology is changing the industry. In: Institute of International Finance, p. 18f.

Current data collection often is not focused on behaviour-based data, and in parts may be unsuited to use in a BDAI context

Current data collection generally focusses on traditional master and claims data, whereas the individual behavioural data of policyholders, which may be valuable in the use of BDAI applications, is often not collected yet.[194] Furthermore, insurance companies in some sectors[195] in many instances still find themselves facing data silos, e.g. data collected in customer relationship management as well as in policy management, claims management and accounting systems. As a result, the data household is often inadequate for the use of BDAI[196] (cf. Chapter 4.2).

BDAI potential can only be fully realised within an appropriate IT architecture.[197] This requires that the core systems are harmonised. The data migrations usually necessary to achieve this, however, are associated with project risks and significant capital expenditure. In the past, for many insurance companies these risks were the main reason that such migrations were not carried out.[198] More recently, however, various insurance companies throughout the market have been showing stronger interest in implementing modern IT architectures. Publicly available information has shown that particularly the German insurance sector is investing heavily in digitalisation – primarily to harmonise core systems. On average, German insurers referenced individual budgets between €100 million and €500 million for the total period from 2017 to 2021.

There are two aspects relevant to this development. Firstly, cost pressure has grown, particularly for health and life insurance providers, as a result of sustained low interest rates. Also, there are initial signs that new competitors – often with great technological expertise – with data-driven business models and corresponding BDAI technologies could enter the insurance market.[199]

---

[194] In some insurance companies, interaction with policyholders – particularly if there are no claims – is limited to an "annual report", which only collects a limited amount of data.

[195] In Germany, there is the principle of specialisation (cf. Section 8(4) of the Act on the Supervision of Insurance Undertakings (Versicherungsaufsichtsgesetz, or VAG)). Consequently, an insurance company can, for instance, only secure a licence for either life insurance, health insurance or other insurance sectors. Also a reinsurance company is only admitted to practice reinsurance business. Often an insurance holding, however, appears to practice business under the same branding. Furthermore, customers are often asked to accept that certain parts of their data can also be used by other insurance companies of the holding (e.g. for cross-selling campaigns).

[196] Cf. Jung et al., 2017, Die Last der Altsysteme. In: Zeitschrift für Versicherungswesen 03/2017, pp. 86–89.

[197] Cf. Chapter 4.2 for further success factors for BDAI application.

[198] Findings from the evaluation of conversations with experts.

[199] Fromme, 2018, Versichert von Amazon. In: Süddeutsche Zeitung 14 November 2017.

The Wall Street Journal, 2018, Amazon, Berkshire Hathaway, JPMorgan Join Forces to Pare Health-Care Costs. Available online: https://www.wsj.com/articles/amazon-berkshire-hathaway-jpmorgan-to-partner-on-health-care-1517315659, accessed on 30 January 2018.

5.3.2 The impact of BDAI on the insurance sector

The various possibilities for using BDAI in the insurance sector, such as targeted customer acquisition using next-best products identified based on the evaluation of a policyholder's data, are generally not only of interest for conventional insurance companies, but could also motivate competitors to either strengthen their position in the insurance market or enter the market as a new participant. The following chapter differentiates market scenarios – as in the previous chapter – for three types of companies: incumbents (traditional insurance companies), bigtechs[200], previously with only isolated activities in the insurance market, and insurtechs. In this context, their respective strengths, weaknesses, opportunities and challenges are described.

The current market positions of the three types of companies mentioned above can be described using the competitive dimensions derived in chapter 5.2: Consumer trust, technological BDAI platform[201], customer reach and regulatory competence/industry know-how[202]. This is shown in Figure 20.

Figure 20: Schematic depiction of the relative characteristics of competitive dimensions for incumbents, bigtechs and insurtechs
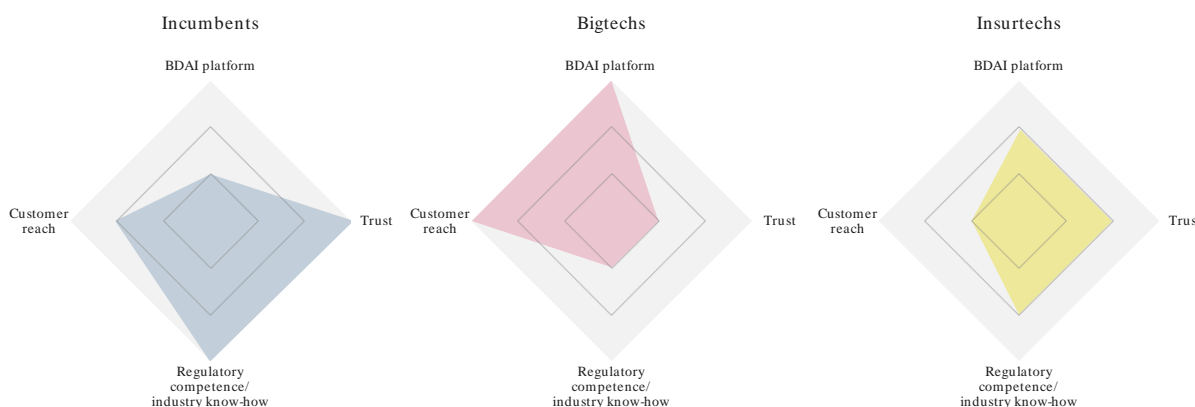
By investing in BDAI technologies, traditional insurance companies can collect new, valuable data that could complement both the standard data collected from policyholders and their claims data.

Compared to insurtechs and bigtechs, traditional insurance companies tend to have significantly more expertise in dealing with regulatory requirements because of their longer presence in the market. They also have a deeper understanding of the industry. At the same time, they benefit from potentially stronger consumer trust when it comes to handling sensitive information (cf. Chapter 4.1). As long as they keep these competitive advantages and at the same time build up their BDAI technologies and competences, traditional insurance companies could be in a good position to secure their current market position, particularly when it comes to the customer interface. At the same time, BDAI applications could also be crucial for insurance companies in their core processes because they tend to go hand in hand with efficiency gains in operations (cf. Chapter 5.3.3).[203] For this, companies would have to invest in BDAI technology and implement appropriate processes (cf. investment budgets for digitalisation in Chapter 5.3.1).

---

[200] Like Google (Alphabet), Amazon, Facebook, Apple.

[201] In this context, the technical innovation capability, the speed at which BDAI applications are implemented and an agile company culture are of particular relevance.

[202] In this context, the authority over proprietary insurance data and the expertise in developing insurance products are of particular relevance.

[203] BDAI-based standardisation of previously heterogeneous work steps will make it possible, in future, to automate (practically) all of the still (semi-) manual business processes within an insurance company.

Another conceivable scenario would be that in the medium to long term, traditional insurance companies would choose to withdraw from the customer interface or be pushed out of these parts of the value chain. This crowding out might happen if new competitors offered policyholders a sustainably better customer experience.[204] In this scenario, a traditional insurance company could focus on its core insurance functions such as the bearing of risks. This would imply greater cooperation with insurtechs and/or bigtechs.

Bigtechs could try to dominate the interface to the policyholders in order to gain access to (policyholder) data and to develop their own products.

Bigtechs currently have greater expertise when it comes to dealing with BDAI technologies on a daily basis, particularly compared to traditional insurance companies. As an example, a major social media company has been using facial recognition software to identify people on photos uploaded by other users since 2010.[205] In addition, bigtechs generally tend to have a broader customer base as well as a firm customer stock – familiar with the generally easy-to-use services offered by bigtechs – and can, therefore, more rapidly scale new business models (cf. banks in Chapter 5.2). Because of the regulatory requirements of the insurance sector and the lack of specific industry know-how, they would need to acquire the appropriate expertise externally or build it up internally, if they intended to enter the insurance market (cf Figure 20, middle).

Thus, one possible strategy for bigtechs could be to enter the insurance market with a focus exclusively on the customer interface instead of developing their own insurance products. Consequently, bigtechs could use their own platforms to sell the insurance products of insurance companies with official licences for insurance business. This would further foster a disaggregation of the value chain in the insurance market similar to the banking sector (cf. Chapter 5.2) – in other words, it would increasingly separate companies selling insurance products from those bearing the risks associated with insurance policies.

Subsequently, it would be conceivable that bigtechs themselves would develop and sell BDAI-based insurance products. If they were not themselves authorised to operate an insurance business, however, they would depend on cooperation with or being contracted by licensed insurance companies. In this scenario, bigtechs could provide consulting services for product development to conventional insurance companies. However, the actual balance of power between bigtechs and conventional insurance companies could shift, creating a scenario in which the incumbent insurers are in a sense forced to work with bigtechs, as these companies:

■   dominate the direct contact with policyholders to such an extent that the insurer would hardly be able to successfully contact and acquire interested customers through other channels/their own branding, or

■   manage to execute processes at such a low cost that the incumbent insurer would no longer be competitive in carrying out these processes himself.

As such, the authorisation to operate any insurance business would remain with the traditional insurance company, if the bigtech did not apply for and receive its own authorisation. Conceivably, in the future traditional insurance companies might act only as carriers of the risks inherent to insurance policies, while all processes associated with operating the insurance business (including product development, underwriting, claim settlement, capital investment, etc.) would be outsourced to one or several tech companies that specialise in conducting distinct processes.

---

[204] Cf. Chapter 5.2.1 for the definition of customer experience.

[205] Cf. Fast Company, 2017, How Apple, Facebook, Amazon And Google Use AI To Best Each Other. Available online: https://www.fastcompany.com/40474585/how-apple-facebook-amazon-and-google-use-ai-to-best-each-other, accessed on 30 January 2018.

It is highly likely that any market entry of a bigtech into a specific insurance segment would depend on the proximity of an insurance product to the actual core business of the bigtech. One major online warehouse is currently in the process of developing product-related insurance policies for higher-value devices such as notebooks and cameras.[206]

At this time, the market is not expected to undergo a fundamental change due to the entry of new or existing insurtechs as they can easily be integrated into the structures of traditional insurance companies.[207]

A large number of different insurtechs can currently be observed in the insurance industry, especially at the interface to the customers but also with a focus on supporting the core processes. For the most part, these companies specialise in individual insurance functions. For example, certain insurtechs provide support services in sales (e.g. comparison websites), contract management or claims/benefits management.[208] When an insurtech sells insurance products but does not have its own licence to trade insurances (e.g. insurance agent), it has to cooperate with one or several established insurance companies that have the requisite licence for insurance business and that bear the risk inherent in the insurance contract.[209]

Regardless of the presence of insurtechs in the market, there has been no significant change in the market to date. Such a change by means of insurtechs is also not foreseeable[210] as the majority lack a broad customer base. The costs of acquiring new customers can also be quite high for insurtechs. For many, one plausible strategy could therefore still be to use their expertise in dealing with BDAI technologies and their digital access to policyholders to cooperate with conventional insurance companies – in other words, to integrate their BDAI competence into the value chains of incumbents in order to offset their lower customer reach levels. In 2017, for instance, one major German financial services provider[211] took over a digital financial assistant able to show bank accounts, investments and insurance policies in a cross-platform mobile app aimed at creating transparency.

The changes described in the previous section for the different types of companies could have a variety of effects on the insurance market. The following paragraphs provide more details on a few of these possible effects.

---

[206] Cf. Fromme, 2018, Versichert von Amazon. In: Süddeutsche Zeitung 14 November 2017.

[207] Beyond the definition in chapter 5.1, this study considers insurtechs to be those companies authorised to operate an insurance business in accordance with Section 8(1) VAG as well as those which do not have the appropriate authorisation and yet – as relatively young, technology-orientated businesses – try to create a niche for themselves along the value chain of insurance companies.

[208] Offtermatt, 2017, Fintechs und Insurtechs. Die Disruption kommt (?). In: Max. 99, Deutsche Aktuar-Akademie GmbH, Versicherungsforen Leipzig.

[209] Ibid.

[210] "Insurtechs were most active in distribution. At the same time, intense competition between online brokers had already resulted in market adjustments among them." The 2017 GVNW Symposium of the Gesamtverband der versicherungsnehmenden Wirtschaft e. V., 2017, Marktbereinigung bei Insurtechs hat begonnen. In: VersicherungsJournal.de 12 September 2017.

[211] The business segment "insurance" is an essential pillar of the entire company.

BDAI-induced efficiency gains could lead to a short-term rise in the profitability of insurance companies.

However, improved profitability is unlikely to be a sustainable effect in the long term as the market mechanism – as long as markets are efficient – would adjust the insurance companies' market prices to the new cost situation. As soon as several providers could cut costs by realising efficiency gains, the strategy of individual insurance companies could be to lower their prices to undercut their competitors and in this way gain additional market share and adjust the insurance market. The consequence of this market mechanism could be that BDAI-conditioned efficiency gains are passed on to policyholders in the form of, on average, less expensive insurance products.[212]

Extracting the consumer surplus is possible.[213]

Regardless of the fact that policyholders – as described in the preceding paragraph – could benefit from gains in efficiency achieved by the insurance companies, in the long term it may also be possible to extract consumer surpluses. Insurance companies could use targeted, BDAI-supported sales management to extract consumer surplus by offering policyholders products at a price that corresponds to their respective willingness to pay and not based on their respective risk (cf. Chapter 6.2.3). A greater individualisation of insurance products made possible by BDAI can also reduce transparency, leading to customers being less able to compare services and their corresponding prices to identify the best offer. Here comparison websites could only help up to a certain point, in particular when the respective offerings reduce the intention of a consumer to "shop around".[214]

The use of BDAI could result in more destructive competition.

Through the use of new BDAI technologies, traditional insurance companies with large investment budgets could gain first-mover advantages in the insurance market. They could use this to differentiate themselves from competitors with smaller budgets by offering lower rates, and to invest in additional technological innovations. In this way, they could try to improve their position in the market.

Bigtechs and some insurtechs could benefit from similar first-mover advantages because they already have extensive experience in using BDAI technology. They could even undercut conventional insurance companies with large investment budgets by offering lower rates because a high degree of digitalisation of processes would be expected for these companies. Being free from any legacy IT systems, bigtechs and insurtechs would then be able to make the most of cost advantages in developing and offering insurance products. This would depend, however, on whether they are authorised to operate an insurance business.

Those traditional insurance companies attempting to adapt the advantages achievable through BDAI as second movers could potentially be squeezed out of the market – unless they are able to also gain a competitive advantage, e.g. by switching to niche products. As a result of the intense competition, in particular, smaller insurance companies (and insurtechs) with relatively small investment budgets could face pressure to consolidate, especially to gain access to necessary BDAI know-how.[215]

---

[212] Findings from the evaluation of interviews with experts.

[213] Cf. explanatory notes in footnote 63.

[214] FCA, 2016, Price discrimination and cross-subsidy in financial services, Occasional Paper No. 22, pp. 7 ff. Available online: https://www.fca.org.uk/publication/occasional-papers/op16-22.pdf, accessed on 9 March 2018.

[215] It is not foreseeable at present that this will result in any significant reduction of providers in the market.

It is currently not possible to make a quantitative statement concerning the BDAI-induced market impact.

It is just as difficult to foresee whether the insurance market as a whole will grow or shrink, because the use of BDAI could have a netting effect, namely occupation of market niches versus reduction of overall claims. In addition, any BDAI-induced market effects could overlap with other major trends such as the development of self-driving vehicles and greater interest in cyber insurance.

The section below describes how the size of the insurance market could be affected by both of the BDAI-induced effects mentioned above and the trends rather not resulting from BDAI.

- **Market niches.** New risks could become insurable as a result of the more efficient collection and evaluation of more data such as behavioural data. This is primarily due to a price effect, as – without the help of BDAI technology – certain risks would have been associated with very high fixed costs so that insuring them would not have been profitable from a purely economic perspective, or products would have needed to be priced so high that demand would have been too weak. With the help of BDAI-based product development, fixed costs could be significantly lowered and (new) insurance products be offered at affordable prices. For example, through automated evaluation of complex, high-volume data sets on the progression of diseases, it could become possible to offer HIV-positive people life insurance in less time and at a lower rate, or even at all.[216] Something similar applies to homeowner insurance for buildings in flood-prone areas. Using satellite pictures and geo data, insurance companies can identify such areas more precisely. Consequently, the insurance market could grow as market niches are covered.

- **Reducing claims and premiums.** One effect on market size that could offset the coverage of market segments – that had (previously) not been served – could be the reduction of overall claims through better fraud detection and loss prevention (e.g. more prudent driving through the use of telematics tariffs, leading to fewer car accidents/claims payments overall). BDAI could be used for targeted avoidance of value destruction, which might make it possible to reduce insurance premiums.[217] Reducing premiums paid would result in a smaller insurance market – when measured with respect to gross premiums paid.

- **Changes in the motor vehicles market.** The development of self-driving vehicles could permanently reduce the importance of motor vehicle insurance for the insurance sector as the number of accidents should decline.[218] In 2016, premiums paid for motor vehicle insurance in Germany comprised 13 percent of the total amount of premiums paid in the overall German insurance market.[219] Estimates show that gross premiums paid for motor vehicle insurance could drop by up to 80 percent by 2040 compared to 2016.[220] Thus, the overall insurance market could become significantly smaller – independent of potential BDAI-induced market effects.

---

[216] One US insurance company has been offering life insurance policies to HIV-positive customers since early 2016. Product development is based, among other things, on the evaluation of data sets containing several tens of thousands of HIV-positive people and was performed in cooperation with a German reinsurer.

[217] From an economic standpoint, this positive effect would not necessarily be accompanied by a reduction in the profits realised by insurance companies. It is expected that the cost advantages potentially passed on to the policyholders would only represent a proportionate share of the reduction in claim cost. At the same time, individual policyholders, e.g. those not so tech-savvy, might see higher premiums (cf. Collective consumer protection in Chapter 6.2.3).

[218] The trend depends, among other things, on how well consumers and society accept self-driving vehicles.

[219] BaFin Annual Report 2016. Available online: https://www.bafin.de/DE/PublikationenDaten/Jahresbericht/jahresbericht_node.html, accessed on 20 April 2018

[220] Morgan Stanley and The Boston Consulting Group, 2016, Motor Insurance 2.0, p. 3. Available online: https://www.bcg.com/publications/2016/automotive-motor-insurance-2-0.aspx, accessed on 15 January 2018. There are differences in the degree of decline depending on the respective country. Trends other than self-driving vehicles were also taken into consideration, such as car sharing.

- **Information security risks.** Along with growing interest in securing against data loss, damaged hardware and software, business interruptions due to IT disturbances or violations of intellectual property rights among both private and corporate customers, insuring information security risks could become a new mass product in the insurance market. Establishing this market segment could then drive revenue growth resulting from higher premium income, and thus, growth of the market. Due to growing demand for this product type, national and international initiatives and work groups have been established in recent years by the OECD, the IAIS and the EIOPA. The objective is to gain a better understanding of these products and the related risks in order to be able to develop recommended courses of action for supervisory authorities.[221]

5.3.3 The impact of BDAI on the customer interface

For customers, the use of BDAI could make interaction with insurance companies easier and faster as well as more individual.

When insurance companies use BDAI technology, it affects the customer interface in two ways. On the one hand, it serves to bind policyholders to an insurance company through ongoing interaction with them and providing them with continuous information about new personalised products (targeted sales). On the other, it is the experience of individual interaction with an insurance company (the customer experience) that is important.

Sales could change in such a way that, by using BDAI, individualised marketing[222] (lead generation, cross-selling) could be developed for each individual policyholder, making it possible to offer personalised products that take into account their respective life situation. This would also include offering short-term, situational insurance products such as protection for a test drive, bus travel insurance or a temporary skiing protection plan. Offering such highly individualised products tailored to the respective situation requires that an insurance company has knowledge about which product or service a policyholder needs at any given moment. Unlike the current situation, the policyholder could not only be made an offer upon request but rather proactively – for instance, when visiting a car dealership, embarking on a bus trip and when staying at a ski resort. For instance, first providers of such situational insurance products have become established in Germany. Yet, such offerings are often offered at prices above year-round insurance products covering the same amount of risk.[223]

The potential for targeted sales could be mainly in the private customer area of composite insurance and in the supplementary insurance policies for private health and life insurance. Insurance segments with a high degree of market penetration (measured here by the number of policies/policyholders) could be particularly lucrative. In the composite insurance segment, therefore, individually personalised products might be introduced initially for motor vehicle insurance, followed by household/homeowners insurance, liability insurance and accident insurance.[224]

---

[221] Currently, it remains unclear whether and how BDAI could be used in the collection and analysis of relevant data on the claim probability and claim size of information security risks in order to be able to better assess these risks.

[222] This refers specifically to lead generation. In this context, leads are people interested in certain insurance products and who have given their personal information to an insurance company in order to receive additional information about corresponding products (e.g. by registering for a newsletter).

[223] FAZ, 2015, Spontan per Handy versichert. Available online: http://www.faz.net/aktuell/finanzen/meine-finanzen/versichern-und-schuetzen/mini-versicherungen-spontan-per-handy-versichert-13309963.html, accessed on 9 March 2018.

[224] BaFin Statistik, Schaden- und Unfallversicherungen 2016. Available online: https://www.bafin.de/DE/PublikationenDaten/Statistiken/Erstversicherung/erstversicherung_artikel.html?nn=8228488, accessed on 20 April 2018.

In many insurance companies, the digitalisation process has already made many process steps that were laborious from the policyholder's point of view – e.g. contracting a policy (application, documentation, signatures), the administration of existing accounts (changes to address and name) and claims settlement (claim reporting including damage documentation) – less manual, and thus, faster and more convenient (cf. examples of BDAI-based claims settlement in motor vehicle insurance in Chapter 5.3.7.1). By using BDAI, some processes could be further simplified from the policyholder's perspective and even more customised to suit the needs of the policyholder in the future. In a few years, BDAI applications such as bot systems and digital assistants could become marketable for insurance companies, and then be utilised to further automate existing processes at the customer interface.[225]

Generally, all insurance segments have the potential for further improvement of the customer experience through the use of BDAI technology. In this insurance segment, the potential could be greater the more frequently policyholders interact with their respective insurance providers. Consequently, the life insurance segment should have the least potential since insurance companies usually have less interaction with their policyholders compared with companies in other segments.[226]

5.3.4   The impact of BDAI on the core processes

BDAI could result in a better understanding of the individual insured risk and make core processes more efficient across all insurance sectors.

The use of BDAI by insurance companies could result in a change to the core processes (e.g. product development, claims settlement) and, as expected, increase both their effectiveness as well as their efficiency, particularly by allowing greater differentiation in risk assessment, the reduction of risk and administrative costs and better claim prevention and fraud detection.

With regard to the effectiveness, the primary gain from BDAI could be the opportunity to better differentiate when pricing insurance products.[227] This would be because – due to the availability of new data – the majority of relevant predictors could be redefined to achieve a more granular assessment of risks (and thus, the determination of the premium). In principle, this makes it possible to price high risks higher and low risks lower than before the use of BDAI. Consequently, BDAI could reduce adverse selection in the insurance sector.[228]

---

[225] First providers of intelligent customer service assistants have already become established, in Spain for instance; overall they have yet to build a broad customer base in the insurance market.

[226] Life insurance products have longer maturity compared to other insurance policies. Generally, regular interaction with policyholders is unnecessary. Conversely, there is a significantly higher number of interactions between policyholders and, for instance, health insurance companies during the year, simply due to regular appointments with doctors (including health checks).

[227] However, only under the condition that segments that are homogeneous by today's present knowledge (this study considers a segment to be one insurance rate) are actually still non-homogeneous and can be divided into several sub-segments using external data and new assessment technologies.

[228] By differentiating policyholders, a risk collective (this study considers a risk collective to be the totality of all insured in one insurance sector for a single insurance company) becomes more granular. In other words, all of those insured are divided into more (and thus more refined) segments down to a segment of one. In a segment of one, a policyholder would pay a premium based on his or her individually estimated claims.

The possibility of creating more homogeneous segments[229] could have a positive impact on the variation of expected claims expense, and thus, on the risk costs of each segment[230]. In addition, incentivising risk-conscious behaviour (cf. product development use case in Chapter 5.3.7.2) could further reduce risk costs. At the same time, such an incentive could also offer insurance companies two additional benefits. On one hand, the policyholders could self-select: the fact that a person agrees to use a wearable or chooses a telematics rate could be a potential initial indicator of a lower risk. On the other hand, this would make additional data collectable which could allow insurance products to be further differentiated in some cases. The risks of such a development from the perspective of the policyholder are addressed in the product development use case in chapter 5.3.7.2 and in chapter VI in connection with potential supervisory and regulatory implications and key questions.

The potential for such differentiated insurance products could exist across all segments, both for private and corporate customers. However, it must first be clarified whether such data have sufficiently high explanatory power to evaluate risks for instance.[231] From an economic perspective, BDAI will probably first be used for price differentiation in high-revenue segments. Accordingly, in the composite insurance segment, for instance, greater differentiation is assumed first in motor vehicle insurance, then in homeowners/household/industrial/commercial insurance, and then in liability insurance.[232]

Efficiency could be improved, for instance, through a BDAI-based reduction of administrative expenses. For example, efficiencies could be boosted through BDAI-based automation of the process of contracting insurance policies (e.g. through automated underwriting in life insurance[233]), in the administration of existing accounts including prevention of cancellation and in claims settlement. In these examples, individual insurance companies are already using or testing BDAI applications.

---

[229] Simultaneous formation of a larger number of segments/rates within a risk collective.

[230] Less risk capital would be required as a result of lower variation related to the expected claims, thus lowering risk costs.

[231] For instance, there is also criticism with respect to the relevance of wearables: "Das sind nur Gimmicks, durch Sport und Bewegung kann man maximal eineinhalb Prozent der Gesundheitskosten beeinflussen.", O. Bäte, Chairman Allianz SE. In: "Wir brauchen Waffengleichheit", Der Spiegel, 16 January 2016.

[232] BaFin Annual Report 2016. Available online: https://www.bafin.de/DE/PublikationenDaten/Jahresbericht/jahresbericht_node.html, accessed on 20 April 2018

[233] The software tools of major reinsurers are already available for this.

In terms of reducing expenses, the potential of BDAI can be approximated by using two indicators: the number of transactions[234] and the potential for standardisation of process steps[235]. It is to be noted that BDAI potential in a certain insurance segment could be higher the more transactions occur and/or the greater the potential for standardisation – and thus, for automation – of the affected process steps is. By this logic, based on the number of transactions, BDAI potential is significantly higher in the private customer business. Analysis of the product dimension within the private customer business shows that the composite sector could hold significantly higher BDAI potential than the health and life insurance sectors due to its relatively high number of transactions and the greater number of process steps with the potential for standardisation. Comparing health and life insurance, following the logic described above, it is assumed that the health insurance sector could have slightly higher BDAI potential. Within the composite sector, motor vehicle insurance could provide the greatest BDAI potential followed by household and homeowner insurance, liability insurance, accident insurance, legal expense insurance and other indemnity/accident insurance.

In the long term, insurance companies will probably try to implement BDAI applications selectively in individual process steps and utilise new BDAI technologies to holistically transform their business, and thus, increasingly achieve end-to-end automation of processes. It would be conceivable to fully automate, for example, the process of new customer subscription in health insurance, from the point of first contact with an interested customer through to issuing the certificate of insurance, so that no employee needs to support the process. As a result, the operational risk for insurance companies will change. Risk will be shifted to model risks. However, precise estimates of the extent to which the risk will either increase or decrease are not possible. Operational risk could potentially increase as shorter process times would leave less time available for detecting automation errors and introducing corrective measures. Similarly, the operational risk could decrease overall as the usage of algorithms could at the same allow to improve the control mechanisms.[236] In addition, operational risk changes as information security risks tend to rise through the use of BDAI applications and through the correspondingly stronger interconnectedness of business processes and higher dependence on digital applications (cf. Chapter 4.2).

Furthermore, BDAI applications could support traditional actuarial tasks or in future even perform those tasks, which would also reduce administrative expenses. In particular, the use of AI could make it possible to automate the migration of IT systems within a shorter period of time and at a much lower cost. One example of this is the BDAI-supported migration of life insurance policies.[237]

Furthermore, BDAI-supported claim prevention and fraud detection could improve the efficiency of insurance companies. Both of these, together, could potentially reduce claim costs (cf. examples for loss prevention and fraud detection in the use case claims management in Chapter 5.3.7.1). Also, BDAI could be used to minimise the claim size of losses incurred; this is discussed, in particular, in the context of industrial insurance.

---

[234] Indicator measured by the number of new policies per year and sector divided by the average policy maturity for each sector multiplied by the average number of transactions per year and sector. Source for new policies: Hoppenstedt insurance database 2017. Average policy maturity 50 years (health insurance), 30 years (life insurance), 1 year (composite insurance). Avg. number of transactions p.a. 2.5 (health insurance), 1 (life), 0.2 (composite). Empirical data.

[235] Indicator measured by the personnel cost rate per sector: i.e. expenses for wages and salaries divided by the net premiums earned (= earned premiums for own account). Logic applied: the higher this ratio, the more manual the process, therefore the higher the BDAI potential. Source for wages and salaries as well as net premiums: Hoppenstedt insurance database 2017.

[236] One prerequisite for this is the error-free programming of the algorithms used. In addition, it can be expected that algorithms generally make more objective decisions than is sometimes possible for humans.

[237] This is a BDAI method aimed at efficiently structuring the data household of a life insurance. The secondary objective is to reduce the number of different products in the company while simultaneously ensuring that no policyholder is disadvantaged through the migration.

5.3.5 New business models through BDAI

Insurance companies could develop new business models based on BDAI.

This study considers a new BDAI-induced business model in the insurance sector to be a new source of revenue, separate from the traditional insurance business.[238] One potentially new business model for insurers is the monetisation of data, in other words, using the data to offer products beyond traditional insurance products, or selling sufficiently anonymised data sets to other insurers or other companies.[239] On the subject of the risks associated with the potential monetisation of data, see the use case on monetisation of customer data in chapter 5.2.6.3, on consumer trust in chapter 4.1 and on potential supervisory and regulatory implications and key questions in chapter VI.

To date, no significant instances of the monetisation of data by insurance companies have been observed in the market. Generally some market participants estimate that traditional risk data would be difficult to monetise. Monetisation of data is conceivable in two areas, as described below:

---

[238] There are country-specific differences in the extent to which business unrelated to insurance would have to be outsourced to a subsidiary. In Germany, for instance, insurance companies are prohibited from conducting business unrelated to insurance, cf. Section 15(1) of VAG.

[239] These companies could come from sectors other than financial services.

Traditional, proprietary and new external data (e.g. collected from telematics or wearables applications) could be used together for loss prevention, and thus, to reduce risk (cf. Chapter 5.3.7.1). This implies a development away from a pure claims manager towards becoming a risk manager, which could create a source of revenue in addition to the core insurance business.[240] Furthermore, in the context of the Internet of Things (IoT) this way of monetising data could open up the opportunity for insurers to offer enhanced, regular services, for example, sending early warnings of hailstorms in motor vehicle insurance, or providing information in household and homeowner insurance concerning which household devices or commercial machines need to be maintained and when, in order to prevent or to mitigate short circuits or fire hazards. More frequent and beneficial contact for policyholders could serve to improve customer retention. For example, one US insurance company offers its policyholders the option of receiving instructions about how to reduce hurricane damage through "Alexa"[241].[242]

Similarly, newly collected data could be sufficiently anonymised in data packages and sold directly to interested companies, such as motor vehicle dealers (telematics data) or fitness centre operators (wearables data). For this, however, the respective insurance company needs to possess the data which at least in Germany is not (yet) always the case.

It is currently unforeseeable how much income insurance companies could generate through the business models described. Also, the extent to which this income would be sustainable is unknown. In particular, the extent to which such transactions would be compatible with the requirements of data and consumer protection laws as well as with the requirements of individual countries governing business unrelated to insurance would have to be considered. The ways in which insurance companies use personal information within the scope of such new business models could also have implications for consumer trust.

---

[240] Cf. Bauer, 2017, Cognitive IoT – Trends & Auswirkungen auf Prozesse der Versicherungswirtschaft. In: DAV/DGVFM-Jahrestagung 2017 – Actuarial Data Science.

[241] An internet-based, personal assistant with loudspeaker function offered by US company Amazon. It remains to be seen how such smart assistants in the home will be used and what impact they will have on privacy protection.

[242] Cf. Mutchler, 2017, Insurance Companies Increase Alexa Skill Offerings. Available online: https://www.voicebot.ai/2017/03/06/insurance-companies-increase-alexa-skill-offerings/, accessed on 30 January 2018.

5.3.6 Relevance of external data

Market entry of non-insurance companies[243] would be facilitated, if these had external data relevant for risk assessment and could purposefully analyse it. The actual explanatory power of external data has not yet been adequately proven.

The systematic use of BDAI at the customer interface could gain greater significance in the insurance sector because the external data collectable here (i.e. not the primary data from an insurance company's claims history) could not only serve to digitally expand the customer interface, but could additionally be used to achieve greater differentiation in assessing risks based on new predictors.

The technical access to relevant external data (cf. Chapter 5.3.7.2) for a description of new relevant data types and sources) would, however, not only be reserved for traditional insurance companies. In future, third-party companies (non-insurance companies) could be able to collect and monetise this data as well.[244] One option for monetisation would be to use it for the insurance business. Currently, the explanatory power of external data alone is inadequate for characterising the insured risk. Only when combined with traditional, proprietary insurance data (including claims history and general information about the policyholder) greater granularity in risk assessment could be achieved. Traditional data would primarily be used for back testing.

To date, the de facto gain in explanatory power of individual new predictors (e.g. the acceleration behaviour of drivers) for assessing risks has not yet been conclusively proven.[245] In the past, empirical evidence has shown that determining new predictors can be of interest for insurance companies. For instance, one study showed that in motor vehicle insurance, there is a statistical correlation between credit ratings[246] and expected claims.[247] The question is therefore whether in future the explanatory power of the data mentioned could be sufficient for non-insurance companies to effectively conduct insurance business and whether the potentially more precise characterisation of the insured risk would be material. If both of these prerequisites were met, this could facilitate the entry of new competitors into the insurance market, given that they would meet the regulatory requirements applicable to insurance companies.

---

[243] This study defines non-insurance companies as those companies that in the past were not engaged in insurance business yet in the future could gain access to relevant data to determine actuarial risk.

[244] In compliance with legal requirements such the EU-GDPR.

[245] For an assessment of the current relevance of external telematics data in motor vehicle insurance see Morawetz, 2016, Der telematische Irrweg der Kfz-Versicherung. In: Zeitschrift für Versicherungswirtschaft, 04/2016, p. 5f.

[246] Referring here to credit scorings.

[247] Cf. Monaghan, 2000, The Impact of Personal Credit History on Loss Performance in Personal Lines. In: Casualty Actuarial Society.

5.3.7 Use cases in the insurance sector

There is currently no evidence to support comprehensive use of BDAI in the global insurance market – due to the challenges described above. However, insurers around the world have already begun using or testing initial BDAI technologies. Among other things, they are trying to design premiums on the basis of various, heterogeneous data sources. Moreover, insurance companies are trying to increase the number of policies per policyholder by offering next-best products. Thus, cross-selling is optimised as well.[248] In addition, BDAI is used in claims settlement and fraud detection (both in acquisition as well as in existing insurance policies): through automated photo and video analysis.[249]

Two of these applications are described in detail in chapters 5.3.7.1 and 5.3.7.2:

- **Data analysis in claims management.** The first use case examines how insurance companies currently use BDAI to reduce administrative costs in claims settlement, to detect fraud and to prevent losses and how they could use it in future. Furthermore, it shows the extent to which the claims settlement process in particular could be simplified for the policyholders.

- **Adding external data in product development.** The second use case details the potential for greater differentiation in pricing actuarial risk through the use of BDAI. Furthermore, it is described that the selection pressure in the market could increase, combined with the risk that certain customer segments could be forced to pay significantly higher premiums, while other customer segments could benefit from significantly lower premiums.

A third use case (cf. Chapter 5.3.7.3) addresses a potentially disruptive insurance system, one separate from the current legal framework, which could become established in the next decade:

- **Homogeneous risk groups based on BDAI.** A better understanding of actuarial risk could make it possible to form large, international, homogeneous groups. In a certain organisational form chosen by the policyholders themselves, these groups could exploit cost advantages and make use of supervisory arbitrage.[250] However, it should be noted that the possibility of local insurance offers being undermined by international insurance platforms is not currently foreseeable.

These use cases were selected on the basis of an indicative assessment of the economic BDAI potential[251] – whereby the greatest BDAI potential is currently seen in claims management and in product development/pricing – and on the basis of potentially disruptive developments in the insurance market.

---

[248] Gerbert et al., 2017, Putting Artificial Intelligence to Work. In: The Boston Consulting Group, p. 10.

[249] Findings from the evaluation of conversations with experts.

[250] The respective business model must always be individually assessed to determine whether the organisation is subject to mandatory approval and how to deal with risk collectives formed on the basis of BDAI for realising P2P insurance, in which members could be organised according to the Teambrella concept (available online: https://teambrella.com/). Generally applicable statements cannot be made.

[251] Findings from the evaluation of interviews with experts.

5.3.7.1 Data analysis in claims management

1. Introduction

Claims management consumes a significant portion of the resources in the value chain of all insurance companies. Though many claims can be settled quickly, issues surrounding claims are at times complex and require a lengthy process of manually assessing claims and insurance payouts.[252] Based on several insurance segments in composite insurance[253], this use case presents how BDAI can be used to support automated claims management, in this way both improving and speeding up the process.

BDAI will extend beyond the automation of the preliminary recording of reported claims in the insurance companies' application and benefits systems. BDAI could serve as an administrative aid by automating certain verifications or plausibility checks of circumstances or of the extent of damages on the basis of additional data sources. Furthermore, BDAI could even be used in future to completely automate claims management for a subset of claims.

Today, some companies already use the analysis of external data sources to aid their claims management for selected applications, as described below.

2. Application

BDAI – more precisely, the computer-based evaluation of publicly available (e.g. databases with prices for replacement parts) and private data sources (e.g. photographs and videos provided by customers) – can aid claims management in three different ways:

- Improved claims settlement processes through (partially) automated pricing and approvals or denials[254] of benefits in a claim; and

- More efficient and effective detection of fraud through more precise damage detection; and

- Loss prevention.

Improved claims settlement

In composite insurance, claims settlement processes can be optimised using BDAI applications to the extent that claims can be assessed automatically using photos and videos taken at the site of the accident and other data sources. This changes the entire claims settlement process, as shown below in the simplified representation in Figure 21.

---

[252] Due to its complex nature, in addition to the documents provided by the policyholder, claims settlement may also require the collection of estimates from specialists such as lawyers and doctors.

[253] This use case deals with motor vehicle collision and motor vehicle liability insurance as well as household insurance and insurance on buildings.

[254] Under the code of conduct of the German Insurance Association (Gesamtverband der Deutschen Versicherungswirtschaft e. V., or GDV), automated decisions can only be made in the case of a decision to the benefit of the policyholder, i.e. only the acceptance of an application or the approval of an obligation to provide indemnification can be presented in this manner.

Figure 21: Comparison of a simplified claims settlement process without vs. with the use of BDAI
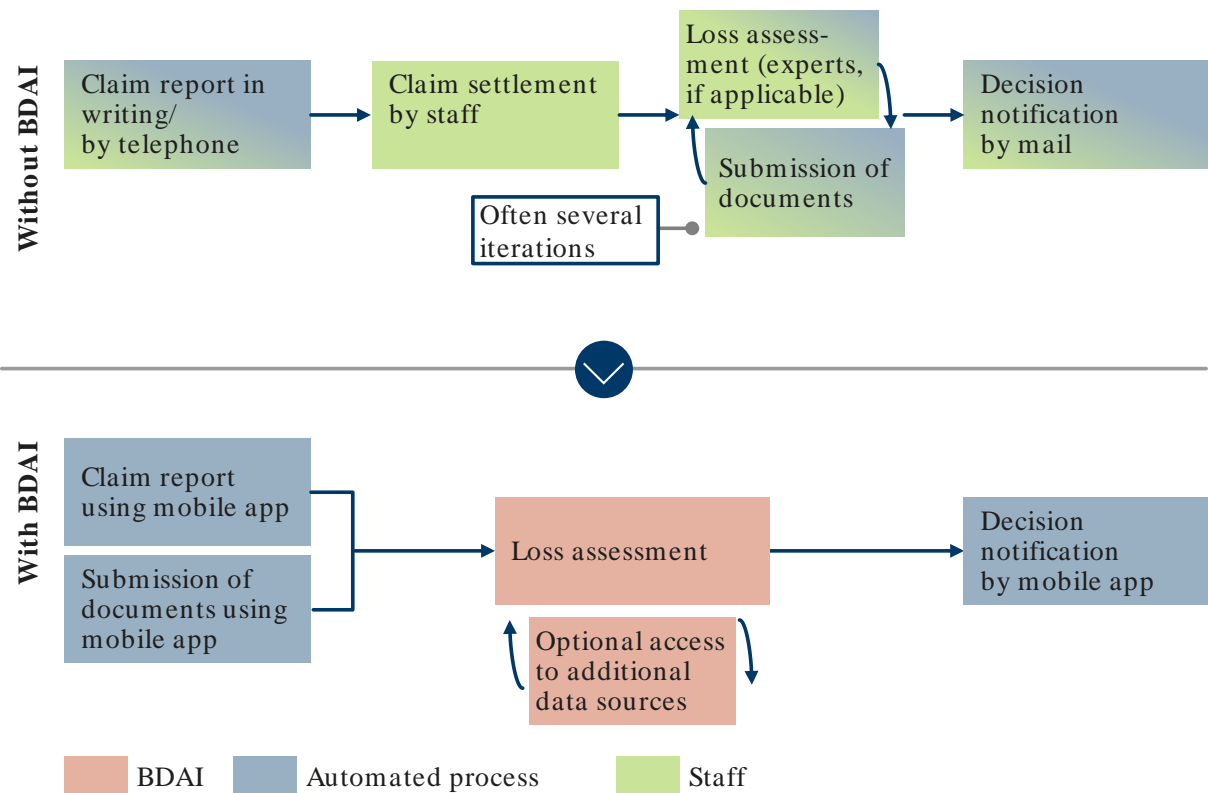
For example, motor vehicle insurers in Germany and the US already offer their policyholders a claims app that they can use in the event of damage to their vehicle to upload photos of the damage to the mobile app and transmit them to the insurer. Thereby, the location[255] is determined automatically along with the date and time and added to the data set.

Meanwhile, there are now service providers that specialise in taking high-resolution images of damage (e.g. to motor vehicles or buildings) from drones and using BDAI to draw conclusions for insurance companies. For instance, one US start-up platform provider offers such services to insurance companies mainly in the USA, but now also in Europe and Australia.

Example: Largely automated claims settlement in household insurance

The following example shows that by means of BDAI it may be sufficient, under certain conditions, for the policyholder to simply report damages without providing evidence. In 2016, a Dutch insurance group began extensively automating its claims settlement process for its household insurance division. Firstly, an algorithm was developed using the claims history of policyholders with household insurance policies (with a term of over six months). When a claim is reported, this algorithm can decide – based on the insured risk (house/household effects), the type of damage (e.g. fire) and the damage amount – whether the claim must or must not be settled in a completely automated process in the majority of cases (supervised learning, cf. Chapter 3.2.1). The algorithm also calculates the risk of this being a fraudulent claim. In disputed cases, a controller manually reviews the circumstances.

---

[255] For instance, using data from the global positioning system (GPS).

Example: Fraud detection at a motor vehicle insurer by means of predictive analytics

A US motor vehicle insurer uses predictive analytics in a two-step process to detect insurance fraud, thereby combining both automated and manual (work) steps. Firstly, all claims reported are automatically screened for fraud on the basis of a specific pattern defined by various data sources. In a second step, a specific investigating unit (manually) rules on suspected cases in order to avoid unfounded accusations due to fraudulent claims misidentified by the automated process.

Usage of weather data and data from connected devices for loss prevention

In order to avoid – or at least mitigate – claims in motor vehicle insurance, policyholders can be provided with weather forecasts and reports of hail, thunderstorms, stormy winds or icy conditions at their respective location[256], to give them time, for instance, to protect their car from hail damage. The contact can be through a mobile app (push notification), SMS or email. Individual motor vehicle insurers already offer such services in several countries (e.g. Germany and the US). This development raises the question of what consequences could arise for the policyholder if they failed to heed these reports. Nevertheless, it also has to be assumed that the respective policyholder would be assigned to a higher risk collective, which could cause his premium to increase.

Loss prevention is also possible in household insurance and insurance on buildings. On the one hand, by interpreting weather data to provide forecasts and/or advice in time, e.g. use sandbags to protect the basement from flooding, and on the other, by analysing the continuous data streams between connected devices (IoT). With the consent of interested policyholders, an insurance company could connect to their devices in order to permit the early detection and communication of risks – for example, of a toaster malfunctioning or short-circuiting, which could have caused a fire[257]. Furthermore, the insurance company could support interested policyholders by offering other enhanced services, such as information on how to safely dispose of defective devices. A US start-up for smart home sensors teamed up with major US insurance companies to offer smart sensors[258] to policyholders for their apartments and/or houses. This allows damages in the household to be automatically communicated to the insurer. At the same time, these sensors allow insurers to be immediately informed of damages (e.g. vandalism) even when policyholders are not at home and to initiate appropriate measures (e.g. inform the police) as well as to begin the claims settlement process.

---

[256] Using the global positioning system (GPS).

[257] Or detection of potential water damages from defective washing machines.

[258] Sensors to monitor e.g. temperature, humidity, movement of people and noises.

3. Technology and data

In future, recording and evaluating insurance claims using a (fully) automated process will be made possible through the availability of new data and the use of BDAI technologies. The new data include digital images and videos of damaged property and accident sites, publicly accessible weather data and market data on local prices for replacement parts or repair services. BDAI technologies include image, video and voice recognition for automated readouts of, for instance, digitised accident descriptions and damaged properties as well as built-in sensors in, for instance, smartphones and motor vehicles, which continually collect data. Furthermore, AI-based document analysis makes it possible to systematically compare insurance terms in contractual agreements against written requests and statements submitted by the policyholders and their lawyers. The information collected using these technologies serves as an aid in deciding whether a claim needs to be settled, and is also integrated into static or dynamic models used to set the amount of individual loss reserves.[259]

Insurers can pull the data required for greater automation in claims settlement from a variety of sources. They can buy weather and price data from specialist providers. Photos and video material can be acquired directly from the policyholder (or an authorised person) – by photographing and transmitting the photos (of the accident) using a mobile app for instance. Alternatively, drones can be used to collect the photographic evidence (e.g. of hurricane damage) necessary for settling the claim and to transmit it to the respective insurer for analysis.

Three-step analysis for BDAI-based fraud detection

For BDAI-based fraud detection, three automated analyses can be combined: the analysis of the damage/claim, the policyholder analysis and a network analysis. The objective of claim analysis is to automatically detect suspicious claims patterns. Using the internal claim history, data on previous claims at other insurers and general accident statistics, algorithms can address questions such as "Does the extent of damages match the accident?" or "Does the medical diagnosis match the accident?" The objective of the policyholder analysis is to automatically detect fraudulent behaviour. Based on the internal claim history and on data on previous claims at other insurers and credit agencies[260], algorithms can assess questions such as "Is the policyholder's financial situation tense?"[261] The objective of the network analysis is to identify connections to fraudulent networks. Using the same data used for the policyholder analysis, algorithms can analyse questions such as "Are there suspicious relationships with other claims?" or "Does the policyholder have suspicious relationships with criminal networks?"

4. Opportunities

The approaches for aiding claims management in composite insurance are associated with a series of opportunities for policyholders and insurance companies. In summary, policyholders could benefit from a better customer experience, lower premiums, additional services and indeed fewer damages overall. By the same token, insurance companies could improve customer retention through a better customer experience, cut administrative costs, identify fraudulent claims earlier and more easily as well as maintain more frequent and value-adding interactions with their policyholders.

---

[259] From an accounting perspective, the expenses for the claim are recognised in individual loss reserves, which are reversed upon payment of the claim. In this context, "dynamic" refers to algorithms in valuation models that continually modify themselves and improve their forecast quality using new data.

[260] Access for insurance companies is handled differently in each country.

[261] From a technical perspective, social media can also be utilised for deepening the policyholder analysis.

Policyholder perspective

In the years to come, policyholder behaviour and the expectations they have of business processes will continue to change. The percentage of digital-savvy policyholders in particular will grow, along with the expectation of digital interaction with the insurer, meaning through mobile apps. The use of BDAI in claims management could be one way to address this.

The customer experience during claims settlement could be improved by two effects. Firstly, claim processing time could be reduced. In particular when it comes to complex claims, especially under private health insurance involving prepayments by the policyholder for services, this would be beneficial to policyholders because they would learn of the decision in favour of damage cover earlier. On top of this, the claim would be paid faster. In simple claims, the ruling on claim coverage could even occur within a few minutes of the claim being reported.[262]

Secondly, claims settlement could be simplified from policyholders' perspective – by reducing or doing away with time-consuming written claims documentation and, ideally, by eliminating the need to save physical receipts[263] as well as, in the case of motor vehicle claims, by dispensing with requests for cost estimates at repair shops.

A lower number of claims and in turn lower claim costs both due to improved fraud detection as well as loss prevention could expand the range of options for lower premiums. Positive effects could result from improved fraud detection, in particular the detection of systematic fraud systems that hurt honest policyholders by causing premiums to be raised.

Insurers could also offer innovative services in line with loss prevention. For example, and assuming they consent to the process, policyholders could supply photos of their houses and then receive suggestions on how they could better protect their property.

Furthermore, as a matter of principle, insurance companies could take loss prevention measures to help their policyholders prevent claims and the related effort.

Insurance company perspective

One key objective of insurance companies in using BDAI for claims management is securing added value for all policyholders. Claims settlement is the part of the value chain on the basis of which policyholders usually decide whether they are satisfied with the services of their insurer. A faster and more convenient claims settlement process is therefore a key element of improving customer satisfaction over the long term –and improving customer satisfaction could effectively reinforce customer retention.

By effectively utilising the technologies mentioned above, insurers could realise cost savings from more automated claims settlement processes, more efficient detection of fraud cases and improved loss prevention through better interaction with policyholders.

---

[262] Findings from the evaluation of interviews with experts. Naturally, policyholders differ in their assessment of the degree to which the customer experience improves as a result of the response via mobile app after a few minutes as opposed to by mail in (as the case may be) a few days.

[263] In future, all receipts could potentially be saved digitally in mobile applications provided by insurers.

On the one hand, BDAI can be used to reduce the internal claims settlement costs (i.e. mainly the costs of the claims department). Two effects contribute to this: a larger number of claims could be classified as "easy to settle" and be processed with the aid of automation or even fully automated.[264] In addition, the average claim settlement period could be shortened overall. In particular, more complex claims that still need to be reviewed manually could be aided by BDAI and, as a result, be processed more efficiently. On the other hand, the external claims settlement costs (e.g. costs for experts) could be reduced, as photo and video analyses could replace the services of experts in a higher number of claims.[265] The cost savings potential from BDAI in claims management would certainly be specific to the respective company and would depend on automation initiatives already introduced and investments already made.

All claim reports could be screened more quickly for abnormalities using BDAI-based automation of fraud case detection. As a result, claim payments for (previously) unidentifiable fraud cases might be avoided. Based on GDV estimates, the damage from insurance fraud currently amounts to approximately €4 billion each year.[266] Even smaller claims that were previously not worthy of investigation due to the high cost of manual processing could be quickly and efficiently checked for fraud.

Ensuring more consistent loss prevention would result in fewer claims occurring, which would lower claim costs on the one hand. On the other hand, it would allow insurance companies to interact more frequently with their policyholders and achieve better customer retention.

In the past – when there were no damages – interaction with policyholders was frequently limited to, for example, sending the invoice for premiums. By means of BDAI, however, policyholders could be inexpensively supported anytime and anywhere. Applying new data and ways of analysing them, companies can utilise new service opportunities aimed at bringing policyholders and insurers closer together. Preventive services in particular make the added value of insurance transparent for policyholders – an insurer could convincingly give its policyholders the feeling that they are actually being taken care of constantly, even in the absence of an acute claim.

5. Risks

In addition to the opportunities outlined here, the approaches presented for supporting claims management are associated with risks.

Policyholder perspective

From the policyholder's point of view, there is a certain risk that the models used to assess individual claims could lead to systematically unjustified denials of claims. BDAI-based network analyses in particular could mean that policyholders are unaware of being erroneously allocated to a certain category – e.g. "relationship with criminal network" – after having reported a claim. As a result, policyholders would face avoidable expenditure (in terms of time) to pursue their claims in court. There might also be the risk of the insurer not covering the claim where preventive advice was not observed and in individual cases the claims not – or at least not fully – being covered.

---

[264] Even without BDAI, many claims can be settled with little manual effort.

[265] The extent to which BDAI-based photo and video analyses would be allowable in court is not discussed in this study.

[266] Insurance Europe, 2013, The impact of insurance fraud, p. 11.

It is assumed that damages of approximately €2.2 billion from insurance fraud remain undetected each year in the insurance market in Great Britain, cf. Insurance Europe, 2013, The impact of insurance fraud, p. 9.

The digitalisation of claims settlement processes raises the question of how the changes that inevitably accompany this could in the long term limit certain policyholders in the way they interact with insurers. Should the newly developed customer interfaces and processes – in terms of claim reporting via mobile apps – mean that individual policyholders[267] were disadvantaged, such as those who are less tech-savvy, it would be important from a supervisory perspective to ensure that all policyholder interests are still sufficiently preserved. Moreover, it must be ensured that the less tech-savvy policyholders can still insure themselves at justifiable premiums.

Furthermore, using BDAI in claims management could make it even more complicated for policyholders to be able to relate to what insurers concretely know about them by analysing their data. Also, policyholders might not be able to understand how those findings affect their individual claims settlement processes and their premiums.

Depending on the way individual preventive measures are structured, insurers will have access to quite precise patterns of movement for their policyholders. If this were legally permissible in the future, insurers would be obliged to ensure that such data could not be used for a purpose other than that for which it was collected. Also, considering the smart sensors already mentioned, IT security has to be taken into account. Here one question is central: Are the insurers' means of access to connected devices and such sensors adequately secured to prevent data attacks and manipulation?

For a higher degree of automation in the assessment of claims, policyholders would potentially have to transmit sensitive evidence when the situation arises. It is assumed here that the degree of sensitivity of evidence submitted in composite insurance cases would tend to be lower than, for example, for health data and medical reports. Nevertheless, even by providing their data under composite insurance policies, policyholders place themselves at risk of misuse of their data, in the sense that it could be used wrongfully. In future, this risk could be mitigated by setting appropriate market standards to ensure transparent data processing. A future retrospective review should determine whether the market standards laid down by the EU-GDPR, effective as of May 2018, are already adequate.[268]

Insurance company perspective

In order to utilise the potential of BDAI in the insurance business, most companies – depending on their individual starting situation – will need to invest in corresponding IT infrastructure and qualified staff. As insurance companies have had only limited experience in dealing with BDAI, such investments carry significant project risk. During the trial phase of using BDAI in claims management, claims could be settled incorrectly for systematic reasons, for instance, because the algorithms erroneously accept claims. Such a scenario could result in excessive claims payments and negatively affect profitability. Under certain circumstances, major insurers could mitigate these risks more effectively than smaller insurers, which would be forced into implementing BDAI as a result of greater competition and would perhaps not possess sufficient data to train algorithms themselves and instead would need to acquire algorithms externally. Major insurers could initially limit the extent of algorithms in their business and then gradually increase their use.

---

[267] For example, policyholders used to sending handwritten documents by post, or those who deliberately choose not to conduct their communication digitally. A policyholder must always have a true choice in the way he prefers to communicate.

[268] Among other things, the EU-GDPR gives consumers the right to receive information at any time about what personal data is used and for what purpose. Furthermore, the regulation also gives consumers the right to correct, delete and limit the processing of their data.

System flaws might also result in claims not being settled, potentially causing policyholders to lose trust in their insurance providers. In addition, negative press coverage could drive potential policyholders to competitors. The increased number of consumer lawsuits expected in connection with unjustified denials could also result in higher process risk for insurers.

It is feasible that – although the algorithms used correctly categorise claims as "to be settled" – the system incorrectly calculates the respective amounts of the individual loss reserves to be recognised, setting them either too low or too high. The results of BDAI-based models for assessment of individual claims serve as input for actuarial processes to value IBNR[269]/belated claims and for additional downstream processes, e.g. in calculating capital requirements. There is thus a risk that BDAI-induced improper measurement of reserves affects the adequacy of the total loss reserve (actuarial reserves) and could result in under-reserving. In particular, model adjustments and recalibration of these models could cause temporary instability of the individual loss reserves, and thus potentially lead to structural breaks in the input data used for actuarial assessments and risk capital calculations. This presents a particular challenge for actuaries, who also in the age of BDAI would bear the responsibility of ensuring that actuarial reserves meet the legal requirements[270]. [271] Hence, BDAI presents a certain risk of under-reserving and thus a risk that the actual solvency situation lacks transparency as long as the insurance employees do not attain the skills necessary to understand and maintain self-learning algorithms. Systematically inaccurate assessments of individual claims would be particularly serious for insurance policies with long horizons, as potential improper measurements may in some cases only be discovered after several years when payments are made.

Assuming corresponding pressure from competitors, insurance companies would likely be encouraged to procure algorithms provided by third parties to save in-house development costs, underscoring the sensitivity of this operational risk. Ensuring the explainability of how these external algorithms function would be especially challenging for the insurance companies. The topic of explainability of models is covered in detail in chapter 3.5.2. At this point the question should be raised about the extent to which insurance companies can in future guarantee the quality of data sets and thus the decision-making basis of BDAI-based models, particularly when purchasing data from external service providers.

Potentially erroneous information being distributed for alleged prevention of loss also constitutes a comparable operational risk. If policyholders rely on timely information from insurers, the latter have a special responsibility towards their policyholders. Erroneous information as a result of the misinterpretation of data could expose policyholders to avoidable dangers in traffic or in the household. As a result, the insurance company would no longer be considered a reliable partner.

6. Supervisory and regulatory issues

This use case highlights various issues that need to be addressed by the supervisory authorities. These are referenced in the use case conclusion in the next section and will be discussed again in chapter VI.

BDAI-based data analysis for claims management most notably raises the supervisory issue of dynamic models for the assessment of individual claims and the question of future standards in terms of their quality and verifiability. This is joined by the issue of dependence on external data service providers and the question of appropriate quality standards for data input. Another issue would be the appropriate use of data by insurance companies in terms of consumer protection – to retain customer trust and to curb the potential danger of misuse of sensitive data by criminal networks or the insurance companies themselves.

---

[269] Incurred But Not Reported.

[270] In Germany, the legal requirements include, among other things, the adequacy of loss reserves, cf. e.g. Sections 75ff. of the Act on the Supervision of Insurance Undertakings (VAG) in conjunction with Articles 19ff. of the Implementing Provisions (DVO).

[271] This could represent a problem particularly for those with long horizons, especially as there the cost data influenced by BDAI are usually the dominant data set. In contrast, insurance policies with short horizons, which mainly use payment data, should prove relatively robust with respect to the use of BDAI in individual claim valuation models.

5.3.7.2 Adding external data in product development

1. Introduction

Product development[272] is another central process in all insurance companies. Across all sectors, the use of BDAI could speed up this process and increase its current level of automation.

The process of developing insurance products can be divided into three steps. The first step involves establishing the insurance terms and conditions. These define both the insured risk and the exclusions. The latter include certain behaviours, certain risks or special circumstances that are excluded from insurance cover. Additional options can be included in the insurance terms and conditions depending on the product.[273] The second step in product development includes determining additional contract clauses, e.g. stipulating payment terms, sales channels and other contractual terms.[274] Pricing is the third step in the process of product development. Stipulating all relevant insurance terms and conditions (risks, risk classes and clauses) makes it possible to determine a unique price for the insurance product. Herewith each risk is related to a specific property (e.g. a motor vehicle of a certain brand) and provides relevant information to derive the risk classes.

For health, life and motor vehicle liability insurance, this use case shows how insurance companies could use BDAI to further improve, individualise and automate the calculation of premiums commensurate with the risk – as part of product development – by using additional, mostly external data.

2. Application

More differentiated pricing of insurance by adding external data

In health and life insurance, external data could be collected or purchased and analysed using algorithms to enable more differentiated pricing for insurance coverage. In many cases, risk models[275], which are expected to be more precise, could thus replace the cost-intensive, largely manual health exams for a priori assessment of the risks of potential policyholders. These data-based models could possibly predict potential health risks not only more cost efficiently and faster than blood and urine testing but also just as effectively.[276]

---

[272] For the purposes of this study, an insurance product is defined as a conditional contract offering coverage for distinct, clearly defined risks for a specified price – in other words, the insurer agrees to pay predefined compensation when specific events take place. This report uses the terms products and tariffs synonymously.

[273] One example is the option to dynamically adapt the disability amount in accident insurance policies to changing life circumstances.

[274] Cancellation conditions, for instance.

[275] Based on policyholder data (e.g. creditworthiness scores, television usage, online shopping behaviour, recreational activities, queries of medical databases and receipt records).

[276] Cf. Scism et al., 2010, Insurers Test Data Profiles to Identify Risky Clients. In: The Wall Street Journal.

Example: BDAI-based pricing of life insurance

Since 2015, a US insurer has offered life insurance policies with differentiated pricing that policyholders can purchase online in only a few minutes and initially without a medical exam.[277] Risk assessment is based on a BDAI-aided evaluation of online questionnaires and public (health) databases almost in real time. According to the same principle, a German start-up insurance company uses a paid IT-platform focused on corporate customers (B2B) for selling BDAI-based life and occupational disability insurance policies online.

Example: Pricing based on image recognition or behavioural data

Another US life insurance company offers potential policyholders a price estimate for a life insurance policy based on photos uploaded digitally to its website. In cooperation with the technology company, Lapetus, the insurer uses AI to extract information on age, gender[278] and body mass index.[279]

The example of a South African bonus programme – aimed at supplementing health and/or life insurance policies, and meanwhile also offered by major insurance companies in Germany and France – demonstrates that BDAI makes it possible to establish new and innovative business fields. The continuous collection of data on policyholder behaviours[280] allows insurers to create detailed profiles. Premium refunds or payment of bonuses (vouchers etc.) are used to reward risk-conscious behaviour[281] and simultaneously motivate policyholders to continue living in a way that prevents risk. Incentive-based insurance products could expand the insurance business of health and life insurance companies by the aspect of preventive health care.

Rating motor vehicle liability insurance by means of telematics data

In motor vehicle liability insurance, telematics, in particular, are being used to automatically record external personal and vehicle-related information and subsequently to allow differentiated pricing for insurance coverage. Herewith risks are assessed not only based on secondary data (residence etc.) but also using primary data from driving behaviour and driving conditions (weather conditions, traffic, construction sites etc.). The data collected can form the basis for assessing risks and calculating insurance premiums more precisely, potentially even in real time. Figure 22 illustrates how policyholders within a risk collective can be assigned newly-developed motor vehicle tariffs using additional risk predictors.

---

[277] The insurance company does not completely rule out medical examinations, they are merely the exception to the rule.

[278] Use of information on gender within pricing is not permitted in the EU.

[279] BusinessWire, 2017, The Power of a Selfie: Legal & General America Launches SelfieQuote.com. Available online: http://www.businesswire.com/news/home/20170731005616/en/Power-Selfie-Legal-General-America-Launches-SelfieQuote.com, accessed on 8 November 2017.

[280] By means of fitness trackers or health monitors.

[281] In Germany, tariff classification based on a medical examination at the policy inception cannot be changed should the policyholder's circumstances change following conclusion of the contract (cf. federal government response to the enquiry by the BÜNDNIS 90/DIE GRÜNEN faction dated 9 November 2016, printed paper 18/10259; also the statement by the German Association of Private Insurance Providers (PKV-Verband) on big data and health dated 31 March 2017).

The PKV-Verband is of the view that incentive systems rewarding health-conscious behaviour by means of premium refunds are not prohibited, if it can be shown that insurance benefits can be cut (cf. PKV-Verband, Big Data, no. 6). The PKV-Verband also considers it possible from an actuarial perspective to offer from the outset an insurance tariff that is linked to the policyholder's willingness to have health-related data collected regularly.

Figure 22: Illustrative assignment of policyholders to motor vehicle tariffs without vs. with BDAI

Furthermore, telematics can be used for effective loss prevention. Policyholders may thereby be motivated to adapt their driving behaviour in order to receive premium refunds.

The idea of taking driving behaviour and the driving situation into account in evaluating risks already existed in the 1930s. At the time, however, recording such situational data was not technically feasible.[282] In the meantime, utilising telematics tariffs is becoming more established in both North America and in Europe. In 2016, an estimated ten million telematics policies were active in both North American and Europe. By 2019, these amounts could reach 33 million policies in North America and 27 million in Europe, which corresponds to annual growth rates of 40-50 percent.[283]

Very few insurers currently offer telematics tariffs in Germany. To date, major German insurers have only offered telematics tariffs to drivers under the age of 30. This customer segment traditionally includes many high risks so that telematics is particularly well suited as an incentive for safe and compliant driving behaviour. A smaller German motor vehicle insurance company piloted telematics in 2013; however, it was discontinued in 2015. Despite high overall customer satisfaction, the price point was (still) set too high. Thus, the company was unable to attract the critical quantity of policyholders. Moreover, the insurer faced high operating costs for such a comprehensive product (e.g. feedback on driving behaviour, emergency calls for accidents).

---

[282] Dorweiler, 1929, Notes on Exposure and Premium Bases. In: Proceedings of the Casualty Actuarial Society XVI, p. 319; reprint: PCAS LVIII, 1972, p. 59.

[283] Berg Insight, 2016, Insurance Telematics in Europe and North America. In: M2M Research Series.

3. Technology and data

The section below describes how BDAI-based product development and pricing works, first for health and life insurance, and then for motor vehicle liability insurance.

Product development and pricing in health and life insurance

Data collection for the purposes of BDAI-aided risk evaluation for pricing a health or life insurance policy for a new policyholder could be done by applying a purely online method or in a sales meeting with an insurance broker. To do so, the potential policyholders would complete interactive health questionnaires (either online or in the broker's office) and would then receive the result of their risk assessment in the form of a differentiated price. Running in the background, an algorithm would collect, digitise and structure the data submitted and transfer them to a risk model. This would be a supervised learning model as described in chapter 3.2.1. The risk model would require varying degrees of detail in the data sets, depending on individual circumstances. Based on cost-benefit analyses, additional data sources would be automatically accessed in order to achieve a more detailed differentiation of risks. More inexpensive data sources are used first. In other words, the data requested by the questionnaire is first augmented with internally available, already-digitised data sources (e.g. existing and/or former policies, disease history, photos, lab reports). Further, data could be generated with facial scans or an algorithm-based analysis of social media profiles. If all this data were still insufficient for precise risk assessment, unstructured data would be selected using OCR[284] solutions. If still no clear decision could be made regarding the application, the next step would be to access external (health) databases, which provide previously digitised and structured data for a fee. If it were still not possible to reach a final assessment of risk despite all of the information taken into account, the final step might have to be manual assessment.

For incentive- or behaviour-based insurance tariffs, data would be continually gathered from sensors installed in wearables or similar devices and, potentially, sent to the insurer or a third party. Algorithms could then evaluate the data to determine the extent to which financial compensation is justified.[285]

Figure 23 shows data types and sources that could be used for assessments. The comparison of today's traditional pricing to potential future BDAI-based pricing shows how much more precisely the risks of health and/or life insurance policyholders could be determined: based on initial expert estimates, BCG reckons that the average probability of assessing risks accurately could be improved from around 65 percent to approximately 80 percent. Yet it also becomes obvious that predicting risk with 100 percent accuracy is not possible even with the help of BDAI. It also remains unclear what effect the variance of the remaining approximately 20 percent would have on the price of the risk.

---

[284] Optical Character Recognition.

[285] Depending on the respective legal situation.

Figure 23: Comparison of data types and sources usable for pricing without vs. with BDAI. BCG assessment according to initial expert estimates[286]
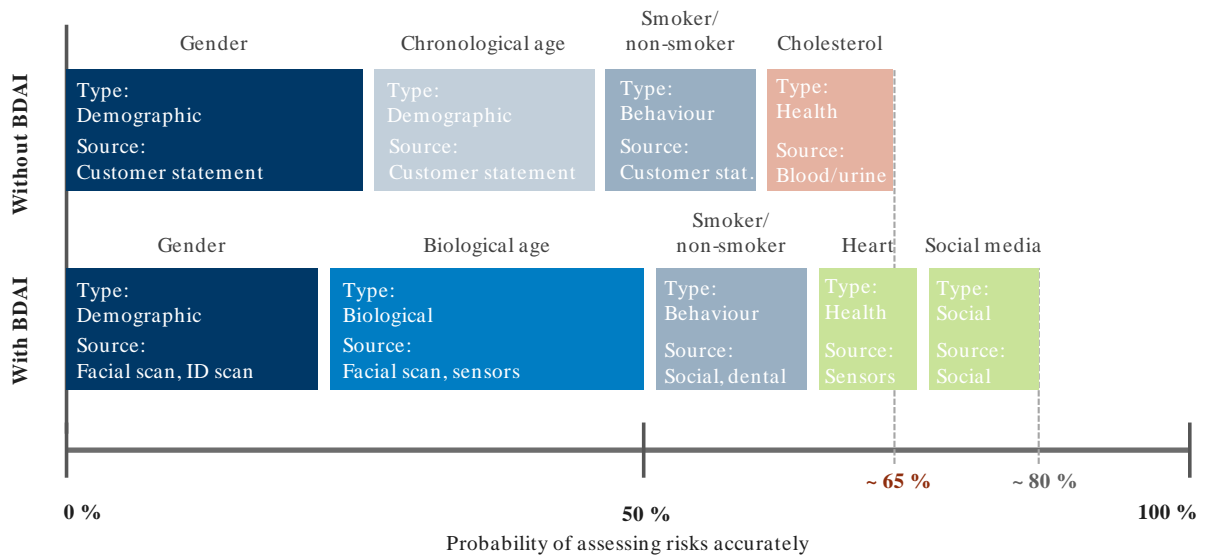
Product development and pricing in motor vehicle liability insurance

Telematics comprises the collection of all kinds of information (e.g. speed) from remote, moving objects (e.g. motor vehicles) using telecommunication networks.[287] In this way, telematics makes it possible to collect and save numerous data sets on the go. To accurately determine situational risk, data can be collected on driving behaviour (driver) and on the driving situation (context). This data specifically encompasses sociodemographic and vehicle-specific contractual data, telematics data such as speed, acceleration, braking and cornering behaviour, maintenance of proper distance, trip mileage, weather and traffic data differentiated by traffic volume, route section, date, time, weekday and season.[288] With the help of BDAI, various interrelated analyses can be conducted based on the data. For example, data could be automatically analysed to determine whether the driver is adhering to speed limits and has an anticipatory and uniform driving style. Abrupt acceleration and braking manoeuvres could indicate higher-risk driving behaviour and corresponding higher risks of accidents. The time and duration of a trip could also be associated with various risks. Rush-hour traffic in the evening and night driving at the weekend, for instance, are riskier than driving in the morning.[289]

Data collection usually occurs through a permanently installed black box or through a telematics app for mobile phones. The black box is usually connected to the car's OBD2 port (vehicle diagnostic system). In this way, the black box can access data through control devices during trips and transmit it over radio frequencies to insurers or other service providers, who analyse the data and then sell it back to the insurers. The second option mentioned uses a smartphone's sensors to record the relevant data. These include GPS sensors for positioning, acceleration sensors for measuring speed and gyroscope sensors for evaluating cornering behaviour. The vehicle data is continually transmitted and analysed using a telematics app.

---

[286] As mentioned above, the use of information with regard to gender for individual price setting is not permitted in the EU.

[287] Berg Insight, 2016, Insurance Telematics in Europe and North America. In: M2M Research Series.

[288] Transchel et al., 2016, Telematik: Was KFZ-Versicherer bereits heute nutzen können. In: Zeitschrift für Versicherungswesen, 15–16/2016, pp. 491–496.

[289] Transchel et al., 2016, Classification of scale-sensitive telematic observables for risk-individual pricing. In: European Actuarial Journal 6(1), pp. 3–24.

Which data is used to calculate risk according to what criteria depends on the individual provider. Likewise, the frequency of recalculating premiums for telematics tariffs varies from insurer to insurer. For example, a German insurance company selected a method under which the maximum possible discount on the monthly insurance premium was calculated after only 25 hours of data collection. The resulting new price for motor vehicle insurance is then applicable for all the remaining months of that year. After that, the insurance premium is readjusted at the beginning of each year on the basis of all data gathered up until that point. Data older than 12 months is no longer included in the calculation as it is considered no longer current.

The reliable calibration of telematics models, in particular, is usually very data intensive. This has led to the emergence of telematics providers that specialise in offering data collection, data processing and data modelling services to insurance companies. A global telematics provider from the UK combines driving behaviour data with contextual environmental data to generate a risk score. This score is used as a telematics factor (generally, a percentage discount or surcharge) in calculating the premium. The company uses a risk model that is continually adjusted on the basis of currently measured claims data. For every new driver or motor vehicle, the initial score is first determined on the basis of conventional differentiating factors such as age, vehicle and number of accident-free years (ex ante risk assessment). Driving data specific to the policyholder is subsequently fed into the calculation and periodically analysed to develop a new score (ex post risk assessment). The individual risk score is sent to the insurance company which, in turn, can reprice a policyholder's insurance premium.

4. Opportunities

The approaches presented for aiding product development or pricing in health, life and motor vehicle insurance are associated with a series of opportunities for insurance companies and policyholders.[290] Insurance companies could benefit from even more differentiated policyholder segments and a reduction of acquisition costs, while some policyholders might be able to enjoy lower premiums and may be motivated to make positive behavioural changes.

Insurance company perspective

Access to new data sources such as individual behavioural data from telematics may allow insurance companies to achieve more differentiated policyholder segments (cf. Chapter 5.3). As a result, policyholder risks previously priced the same could now, increasingly, be differentiated using BDAI. By using BDAI in product development, an insurer could thus secure an informational advantage and differentiate itself from its competitors as well as expand its customer base. This advantage, however, would presumably exist only until other competitors could achieve comparably effective segmentation through the use of BDAI (cf. analogy regarding short-term efficiency gains in chapter 5.3.2). With incentive-based tariffs, data recorded on an ongoing basis could be used to carry out (automated) continuous adjustment of individual policyholder segments.

One consequence of the larger price spread for insurance products is that individual segments would technically decrease in size. However, this would not conflict with the compensation of claims wihtin the collective as the collective would continue to exist.

---

[290] No consistent differentiation is made between sectors in the description of opportunities and risks as they are generally applicable across sectors. The following includes only selective differentiation, where necessary.

A higher concentration of lower risks across all tariffs in the insurance company would go hand in hand with a reduction of the overall claims payments expected. The lower expected claims payments could be used both for lowering premiums as well as for increasing margins. Specifically, by dividing the financial benefits of BDAI-induced positive selection, the insurance companies could improve their profitability while simultaneously offering policyholders less expensive products promising lower risk.

Furthermore, using BDAI can speed up the process of product development/pricing through (partial) automation. The resulting efficiency gains could result in lower acquisition costs. The greatest potential for realising improvements in efficiency is found in non-advisory online sales. Using BDAI, this would also be feasible for products with extensive application processes associated with more complex risk assessments (such as life insurance). To date, online sales have tended to focus on the less complex sectors and insurance products. Here, too, total savings potential from BDAI is certainly specific to the company and depends on automation initiatives already introduced and investments already made.

Moreover, individual segments would become more homogeneous due to greater granularity in policyholder segmentation. As a result, claims could fluctuate less, ultimately meaning that less risk capital should be required. This could also lead to a reduction in the cost of capital.

Lower acquisition costs could, in turn, make it feasible for insurers to sell products with margins that are still not considered lucrative and tap new market segments. Technological advances could allow for smaller, situational insurance policies – such as trip-delay insurance for air and rail travel, previously not offered on a large scale due to their high per-policy cost – to become established on the market (cf. Chapter 5.3.2). Also, for those risks previously unable to be priced commensurately due to the speed and extent of changes in the risk environment, the faster availability of new data and better processing technologies[291] could make it possible to properly assess that risk and to quickly and flexibly adjust the valuation, if needed. For instance, in future pay-as-you-drive motor vehicle liability and collision insurance could be offered at prices reflecting the short-term changes in risk related to changing road conditions (such as fog and ice).

Policyholder perspective

Policyholders could benefit from BDAI – those who previously had not purchased certain insurance policies as they evaluated the price as too high in comparison to their perceived risk could be offered more risk-based, i.e. more individual, products at commensurate prices. Therefore, the spread of premiums mentioned above could benefit certain customer segments in the form of lower insurance costs.

Through incentive-based insurance products, insurers gather valuable feedback regarding the behaviour of their policyholders. This is valuable to the extent that insurers can use the data collected to derive targeted information about policyholder expectations and needs in their respective life situation. This, in turn, could benefit (potential) policyholders as soon as insurers approach them with bespoke new products and services based on this information.

Furthermore and as outlined above, incentive-based insurance products could motivate policyholders to maintain a healthier lifestyle, for example to drive in an anticipatory and disciplined way[292] in order to benefit from lower premiums.

---

[291] Cf. Chapters II and III of this study.

[292] According to the British Road Safety Foundation, beginner drivers with telematics tariffs have an approximately 40 percent lower accident risk than those with traditional tariffs.

5. Risks

In addition to the opportunities outlined here, the approaches for BDAI-aided product development and pricing for health, life and motor vehicle insurance are also associated with risks. Worth mentioning here, in particular, is the risk of losses of reserves for the insurer as a result of missing the product development goals, and from the policyholder's perspective the risk of rising prices in certain customer segments.

Insurance company perspective

For insurance companies, using BDAI-based models in product development could result in the systematic over- or under-assessment of actuarial risk. Subsequently, insurance coverage could be approved inaccurately and premiums calculated incorrectly (negative selection). This would be particularly relevant for long-term insurance policies such as life and accident insurances. In the event of under-assessed risks, companies would receive inadequate premiums for many years which could, in turn, result in losses of reserves when payout is due.[293] Thus, for BDAI-based models, the risk remains that premiums may be calculated inadequately. Insurance companies[294] could face significant challenges when attempting to guarantee more risk-adequate premiums on the basis of self-learning models.[295] They would potentially have to set higher security surcharges in order to prevent underrating. Insurers would then be forced to trade higher security surcharges off against the price attractiveness of their products.

The data basis used is crucial in avoiding incorrect premium calculation. The quality of data, collected by the company and externally procured, must be ensured (mainly completeness, consistency, validity, accuracy and timeliness). In addition to data quality, it is necessary to ensure that data being used to calculate premiums demonstrates a statistical correlation with the risks of the policyholders.[296] The aim is to avoid spurious correlations and mitigate systematic improper assessment of risks. Evidence of what new data actually results in sound, risk-based differentiation is to be provided across all insurance sectors. In the past, incorrectly assessed risk drivers and dependencies were among the reasons why telematics tariffs were discontinued in some isolated cases.[297]

BDAI in product development is still a relatively new thing and experience with it is limited at present. For this reason, it is important to handle new data and algorithms with care in order to avoid underrating. This also involves launching an in-house learning process and building up the necessary BDAI expertise.

Against the backdrop of consumer protection, it may be necessary to further specify in legal terms which collected data and which procured data may be used under which conditions for analysis at all, in order to avoid discriminatory data mining[298]. Moreover, from a legal perspective, the question arises as to how to inform the respective policyholder of the use of individual, personal data. If an insurer were careless with the data or neglected to secure the consent of the policyholder, it could damage the company's reputation. Also, there is the possibility of a general loss of trust in the financial sector if, in future, cases of data being used for unauthorised purposes were to accumulate.

---

[293] For incentive-based tariffs, special features such as the payment of bonuses has to be taken into account and priced in when the policy is underwritten in order to ensure that the premiums are adequate.

[294] Cf. explanation of evaluating self-learning models in Chapter 3.3.1.

[295] Cf. explanation of risks from BDAI-based models in use case in Chapter 5.3.7.1

[296] Cf. for e.g. life insurance and substitutive health insurance Section 138(2) and Section 146 of the Act on the Supervision of Insurance Undertakings (Versicherungsaufsichtsgesetz, or VAG) and appropriation in accordance with the German Federal Data Protection Act (Bundesdatenschutzgesetz, or BDSG).

[297] Morawetz, 2016, Der telematische Irrweg der Kfz-Versicherung. In: Zeitschrift für Versicherungswirtschaft, 04/2016, pp. 68 ff.

[298] Cf. explanation of non-discriminatory data analysis in Chapter 3.5.5.

When introducing a new BDAI-aided product, all data for managing and monitoring insurance products should be recorded in a way that ensures effective product controlling. The objective of product controlling is to retrospectively verify that the actuarial assumptions made upon development of a product can be confirmed after introduction as well as the extent to which undesired effects can occur, such as addressing the wrong customer segments. This also involves verifying the profitability of new products. One of the main challenges of BDAI-based products is monitoring how new parameters – such as acceleration behaviour in the case of telematics tariffs, or the shape of "red capillaries" around the nose in the case of health insurance – are actually suited to calculating risk-adequate premiums. Further, one has to regularly re-evaluate whether or not the physical characteristics of policyholders should be allowed to be used to calculate premiums.

Even though product controlling is generally a core function of insurance companies, its significance differs between the different insurance sectors. This has to do with the specific maturities. In BDAI-based health insurance products, there could be a gap of several years between the launch of a tariff and the collection of resilient data points (e.g. claims payments) depending on the sector. As a result, there would be fewer in-house data points available to validate the above-mentioned assumptions using "real" data.

Also in product controlling, particularly during a new product's trial phase, there might be more data collected from policyholders and analysed than ex post is actually necessary for assessing their risks. This would conflict with the principle of data parsimony or minimisation as reflected by the Federal Data Protection Act or the EU GDPR.

When introducing new rates, primary insurers often use reinsurance companies to better distribute risk. This will most likely also apply to the introduction of BDAI-based tariffs. In this way, reinsurers could be seen as enabling BDAI-based tariffs to some extent. Yet, assuming that BDAI-based tariffs were unprofitable upon introduction, reinsurers would also participate in that lack of profitability. It goes without saying that reinsurers will not underwrite unprofitable BDAI-based tariffs in the long term.[299] A longer period of time could, however, pass between underwriting[300] and the possible realisation that a BDAI-based tariff is indeed unprofitable. As a result, a concentration of unprofitable BDAI tariffs could build up on the balance sheets of reinsurers.

To ensure resilient risk assessments and premium calculations, insurance companies must – as previously mentioned – ensure the validity, accuracy and timeliness of the data basis used. This also includes mitigating the risk that policyholders may manipulate the data. In health/life insurance, potential policyholders could attempt to deliberately manipulate their risk characteristics to achieve more inexpensive policies when entering into an insurance contract, for instance by setting up fake social media accounts.

There is a comparable risk in the field of motor vehicle liability insurance. Policyholders could knowingly drive on their "best behaviour" during the phase when potential premium reductions are determined, later transitioning to significantly riskier driving behaviour. This danger of data manipulation could be avoided if motor vehicle insurance premiums were determined on a situational basis only, i.e. from trip to trip, recalculated each time based on driving behaviour.

---

[299] Unless they receive corresponding compensation in another area.

[300] Individual tariffs may have been underwritten in the past.

Policyholder perspective

With the incentive-based tariffs mentioned above, policyholders leading ostensibly less healthy lifestyles could be faced with higher insurance costs than with non-incentive-based tariffs. A very large proportion of policyholders with low risks switching to incentive-based tariffs could cause the price of non-incentive-based tariffs to climb, as a disproportionate number of policyholders with high risks would remain.[301] This trend could be viewed as price discrimination for the purposes of consumer protection. Assuming such products became established as the standard throughout the market, this could also have negative consequences for policyholders as they would then no longer be free to choose the incentive-based rates. One particularly relevant issue in this connection is the fact that what is perceived as an "unhealthy lifestyle" from the outside is often not the conscious choice of the individual. In other words, policyholders with impaired health through no fault of their own could potentially face discrimination due to incentive-based health insurance. The argument that the policyholder alone is responsible (the notion of personal responsibility) would not hold here.

Due to the potential price spread of insurance premiums mentioned in chapter 5.3, it is possible that a new policyholder's premium (e.g. for a new health or motor vehicle insurance policy) could be set significantly higher than if that same person had been assessed on the basis of traditional risk models. As a result of this price spread, individual policyholders could in extreme cases be technically uninsurable because of the high premiums. It can be assumed that the use of BDAI applications in product development could result in more cases of uninsurability. In the health and life insurance sectors, denial of insurance policies on the basis of new data could prove particularly disadvantageous for people with chronic illnesses or disorders, whether hereditary or accidental. For the purposes of equality, such a development could be viewed as unwanted in both a social and political sense.

To collect the data required for differentiated risk assessment, policyholders would most likely have to consent to a wide array of different uses of their data. In future, it may even be necessary to set up a social media account to be able to purchase certain insurance products at the lowest possible price. Interested policyholders without such an account or those who would not want to set one up would have to either choose a non-BDAI-based and potentially more expensive tariff or do without the respective insurance coverage. This would limit the consumers' freedom of choice.

Despite the required formal consent for data usage, there is a risk that policyholders are not sufficiently aware of the potential uses of the data they submitted. There is also a general danger that policyholders' data could be misused. In consenting to the use of their data, the policyholders trust that their data is absolutely safe and that the company will handle it responsibly. As a consequence, insurance companies must guarantee comprehensive protection of data and trust by providing transparent information and give their customers a real choice in consenting to the use of their data. In the long term, this transparency will strengthen both the consumers' trust in the insurance company and the customer relationship. In addition, this also means that insurers must guarantee that algorithms used for assessing risks only use and analyse the data that is actually relevant, i.e. the data that is directly related to the risk to be assessed (data economy and minimisation with regard to personal data). It is also vital to ensure that the recorded data is not misused to monitor policyholders[302] or used other than was intended.

---

[301] Whether such a development would take place also depends on other aspects such as competitive pressure, market saturation and/or policyholders' willingness to change providers.

[302] Particularly in the context of behaviour-based tariffs such types of surveillance could take place via GPS data.

6. Supervisory and regulatory issues

The aspects set forth in this use case have implications that must be addressed by the supervisory authorities. These are referenced in the use case conclusion in the next section and will be discussed in more detail in chapter VI.

Alongside the discussion on claims management, BDAI-based product development raises the question of the necessity of the acceptance/certification of BDAI-based models, including the development of the requisite BDAI skills within insurance companies. Self-learning models/algorithms could represent a particular challenge. Another key point is the discussion of how to avoid the disproportionate or illegal discrimination of certain policyholders. When it is practically impossible for certain groups of people to obtain insurance coverage as a result of the greater individualisation of insurance products and, in particular, premiums – leading to uninsurability in extreme cases – then there are social and political consequences. Likewise, this use case points out that the importance of protecting consumers' data and their trust that it will not be misused grows as new technologies and business models are introduced.

5.3.7.3 Homogeneous risk groups based on BDAI

1. Introduction

It is likely that the assessment of actuarial risk on the basis of BDAI applications will change in the years to come through the addition of new predictors and/or replacement of traditional ones – provided more differentiated risk assessment becomes feasible.[303] In future, the insurance companies themselves will no longer necessarily be in possession of the data required to assess risk, rather it could potentially be controlled by large data providers.[304] These providers could use their data and analysis tools or pass them on to third-party platforms to compile large, international groups of policyholders that would be homogeneous in terms of their actuarial risk (homogeneous risk groups).

It would make sense for these groups to organise independently as a P2P network[305]. Such insurance networks would mainly approach private customers for purchasing composite insurance and supplemental health and life insurance policies.

The size, homogeneity and organisational form could offer significant cost advantages compared to traditional insurance companies so that this could, in future, represent a potentially disruptive business model. The underlying business model is, however, somewhat similar to a mutual insurance association (cf. example box on page 133).

2. Application

Under the influence of BDAI applications, the process of evaluating actuarial risk will change. New data able to provide insight into actuarial risk will be generated, collected and analysed. Telematics and wearables are the best-known examples that could influence the assessment of motor vehicle liability risk or health risk. Practically all sectors could benefit from using BDAI applications by supplementing and possibly even replacing existing data from insurance claims.

---

[303] Cf. use case product development in Chapter 5.3.7.2.

[304] For example Google (Alphabet), Amazon, Facebook, Apple.

[305] P2P network corresponds to P2P insurance. For the purpose of this study, P2P insurance is distinguished from traditional insurance companies.

Identification of fitting candidates to set up homogeneous groups of policyholders

From the current perspective, there is still a steep learning curve associated with using alternative data sources to evaluate risk. For this thought experiment, however, it is assumed that this step has already taken place: Actuarial risk is assigned on the basis of various data sources – including non-traditional ones – which can differ depending on the sector and the type of risk. In future, it will be possible that the insurance companies themselves would no longer own the information relevant for assigning actuarial risk but rather bigtechs would control the customer interfaces and collect the corresponding relevant data. With this data and the expertise required to assign actuarial risk on the basis of unsupervised learning[306], for example, bigtechs would be in a position to identify from among their customers (or users) those individuals who are comparable in regard to one or several actuarial risks, and thus to form large homogeneous risk groups.

Assuming that the insurance companies would be in possession of this relevant data, they could operate in the same way, only limited to their (usually local) business. For their part bigtechs, however, have significantly higher numbers of users and could thus assemble larger homogeneous groups.

*Data monetisation: A potential business model for bigtechs*

Bigtechs would presumably try to monetise the information they obtain. One option would be to offer insurance coverage to the intended user groups themselves. The likelihood of that happening is rather low, however, because doing so would mean dealing with both the licensing processes as well as the ongoing work necessary to fulfil regulatory requirements. Another means of monetisation could be to sell the information to traditional insurance companies or digital insurance platforms.[307] Alternatively, bigtechs could provide their customers with the added value of the information from their respective risk group free of charge as an added benefit or for a fee (cf Figure 24). By doing so, they could potentially improve customer satisfaction, and thus increase customer loyalty to their actual core business.

---

[306] Cf. explanation of unsupervised learning in Chapter 3.2.2.

[307] In turn, these would be able to offer insurance products within homogeneous groups.

Figure 24: Simplified illustration of the transfer of relevant information to customers and the creation of homogeneous risk groups
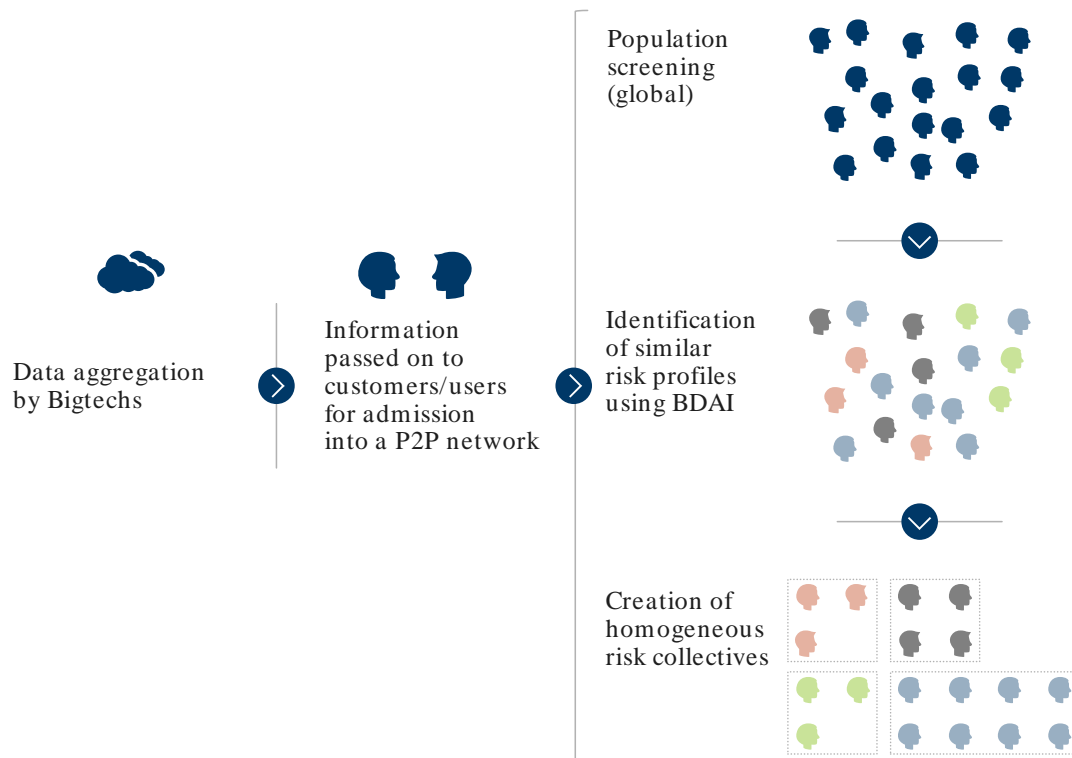
Put simply, all customers (or users) who have shown themselves to be particularly risk averse (or risk seeking) or those who have revealed a certain risk appetite through their online shopping habits could be considered part of a very large and homogeneous risk group. This thought experiment is based on the assumption that in future the information relevant for optimally characterising actuarial risk is not the proprietary data of the insurance companies but will be collected by bigtechs at their customer interfaces. It is also assumed that bigtechs will use their information to inform their customers about their actuarial risk.

Specifically, what could the customers of bigtechs do with such information? How could this situation give rise to insurance coverage and how could the groups organise themselves? Below, this report outlines possible answers to these questions and illustrates a potential organisational form for the homogeneous risk groups described using a contemporary practical example.

3. Technology - based on the example of Teambrella

Russian software architects and entrepreneurs developed a P2P insurance system (Teambrella) that allows the members of self-regulating groups to insure one another.[308] Members of a group assume control of all of an insurance company's core processes including acquisition of new customers, risk assessment and claims settlement. The provider only provides the technical infrastructure but does not make any payouts, does not administer the funds and offers no insurance coverage. The business model is based on a complete uncoupling from traditional insurance companies. The policyholders become the insurers. Pilot projects for the system have been underway in the Netherlands and South America since 2017.

---

[308] Cf. Teambrella. Available online: https://teambrella.com/, accessed on 11 December 2017.

A Teambrella group is made up of at least two members. Those interested can apply to join one or more groups. When joining a group, the potential future member chooses a certain type of insurance coverage – equivalent to the application process in traditional insurance. Teambrella is currently based on the idea that new members will be acquired through social media invitations and recommendations of friends. In our thought experiment, however, groups are assumed to be created based on the information of bigtechs. In the beginning, each Teambrella group defines its own rules, including determining what claims are covered and the maximum insurance amounts that will be paid out. In this case, the rules defined by the group are the counterpart of a traditional insurance contract. Through these rules, every member can decide, for instance, on expanding the insurance cover, on taking on new members and on the amount of claims. The group can make decisions by voting or by delegating to experts. These can be members of the group with broad expertise in insurance or professional insurance consultants. The developers of Teambrella recommend that group members use insurance consultants to promote sound and qualified decisions. What these consultants are paid (compensation is a percentage of the payout for claims) is intended to guarantee that the insurance system is highly professional and efficient. This could, however, also constitute a conflict as there could theoretically be an incentive for consultants to accept into the group interested persons with high risks or to agree to pay out high claims.

Members submit claims directly to the team and provide the information agreed at the outset (e.g. photos or police reports) as evidence. After that, every team member decides on the total amount to be paid out for the claim. When a claim is accepted by the group, every member pays the member being reimbursed. The share of the claim to be paid is calculated individually for each member based on their personal risk coefficient. Every member's risk coefficient is, in turn, calculated from the product of the relative loss probability and the relative expected loss amount. Group members with the same risk coefficients pay the same amount for each other in the event of a claim, while group members with differing coefficients pay prorated amounts for each other, ensuring comparability of the claim payouts within a group. This means that members with higher risk coefficients have to pay higher amounts to members with lower risk coefficients. Conversely, the lower the risk assessment, the lower the payout for claims.

Every member of a group pays a certain amount (equivalent to an insurance premium) into a personal e-wallet. The amounts remain in the personal wallet until claims are submitted and payments have to be made. The developers of Teambrella chose to use a cryptocurrency to keep transaction costs as low as possible.

The Teambrella concept is primarily aimed at property claims. Unlimited personal injury cannot by covered through Teambrella as no reserve funds in the true sense exist, and this would require working with a traditional insurance company. The system is also currently not an option for insurance policies required by law such as motor vehicle liability insurance.

It is not possible to say for certain whether the Teambrella concept described here will succeed, as the project is still in the pilot phase. Yet it is still suited to our thought experiment in terms of demonstrating how groups of policyholders could organise themselves in the future to hedge risk and use the information on group classification.

Since policyholders and insurers are interchangeable under BDAI-based P2P insurance, the opportunities and risks for both are described below. The perspective of the technical platforms (the infrastructure providers) is not considered.

4. Opportunities

From the perspective of a member of a BDAI-based P2P insurance group, there are two aspects that could be viewed as opportunities. The insurance coverage is expected to be less expensive and the underlying processes – particularly those for processing claims – are, potentially, more transparent and autonomous as compared to traditional insurance companies because they are induced by the members themselves.

One major advantage is the cost structure of a P2P insurance – the members potentially receive the full benefits thereof. Acquisition costs are low by design and the administrative costs including internal claims settlement costs amount to a small percentage of the insurance premium.[309] Unused paid-in insurance premiums are distributed back to the members. This means that it is generally possible – if no claims are submitted – for group members to have no expenses within a certain period of time except for the small fees paid to the infrastructure provider.

In addition, membership offers the benefit of designing insurance processes dynamically within the insurance group and the ability to adapt them to the specific needs of the group. This could serve to improve a member's satisfaction with the insurance cover offered. At the same time, claims assessment becomes a more transparent process for users of P2P insurance because they can evaluate and authorise the payouts themselves. Compared to traditional insurance, this could serve to further increase trust in insurance processes.

5. Risks

If using customer data from bigtechs to create an insurance profile, it may well be that personal data is not adequately protected. This raises questions about data protection for members of a P2P insurance. In addition, the desired transparency in calculating risks and claim payouts requires that all members of a group are able to see all the claims submitted and the corresponding details about these claims. With regard to sensitive data (e.g. health data) this could mean that members are negatively affected by somewhat weaker data protection and accompanying loss of privacy.

With P2P insurance, insurance benefits are not backed by a traditional insurance company. Insurance coverage could tend to decrease regardless of the adequacy of the calculated member premiums. Insurance default could occur, especially in years with extremely high total damages (whether as a result of a few unexpected, very high individual claims or a rare cumulation of smaller claims). This raises the question of how this type of risk could be absorbed within the group. One option would be to define a mandatory additional contribution, as (partially) required for mutual insurance associations. When the insurance cover exceeds the premiums received in a given period, the group would have to make an additional payment in order to pay the outstanding obligations to policyholders. It remains to be seen how this requirement could be realised in the example of P2P insurance. Also a reinsurance solution would be conceivable.[310]

A different risk could arise as experience has shown that when it comes to larger groups, in particular, only a smaller number of members will be intensively involved in handling the individual claims. In the decision-making processes described above, these persons would then be assigned a significant amount of power. As a result, it is possible that a type of favouritism could occur with regard to claims settlement: this would generally question the transparency of a P2P insurance. In such a case, claims settlement would be based on the relationship of the person submitting the claim to the person in power rather than being based on the facts of the case alone.

---

[309] Average administrative cost ratio in the German (traditional) composite insurance market is around 14 percent compared to approximately 5 percent at Teambrella (as of December 2017). Hoppenstedt Corporate Database, 2017, gross administrative expenses as a percentage of gross premiums earned; Teambrella. Available online: https://teambrella.com/, accessed on 11 December 2017.

[310] Alternatively, the requirement to fund buffers similar to the equalisation reserves pursuant to the German Commercial Code (Handelsgesetzbuch, or HGB) could be implemented throughout Europe.

Another risk is that members of a group could be confronted with significantly higher administrative expenses and decision-making efforts. When joining a group, it would be difficult to estimate how many individual claims (particularly in large groups) would occur within certain periods of time and then would need to be processed with the involvement of the members. As a result, it might be impossible to assess the true opportunity cost for the group members (more time needed to process claims versus potentially lower payments/costs compared to an insurance contract with a traditional insurance company).

In non-institutional business models with inadequate contractual security, processing payments between members could also be accompanied by further risks. Using bitcoin as a cryptocurrency, for instance, bears a certain risk due to the volatility[311] of the exchange rate[312] and the danger of cyberattacks (which could cause a member to irrecoverably lose access to his e-wallet).[313]

6. Supervisory and regulatory issues

The aspects set forth in this use case have implications that must be addressed by the supervisory authorities. These are referenced in the use case conclusion in the next section.

Groups of policyholders formed on the basis of BDAI for realising P2P insurance, in which members are organised according to the Teambrella concept, could be subject to approval. The individual countries will have to check this. But even if approval were required, the business model might not be limited to jurisdictional boundaries. Should that be the case, it would be possible to circumvent local supervisory authorities and possibly practise regulatory and supervisory arbitrage. In that instance, this would also apply to data and collective consumer protection.

---

[311] Volatility is a measurement of how the price of an asset changes over time. The value of the bitcoin depends on supply and demand as well as from its acceptance in the economic cycle. Price fluctuations and bubbles can occur when speculators enter the bitcoin market. Cf. BaFin-Fachartikel, 2013, Bitcoins: Aufsichtliche Bewertung und Risiken für Nutzer. Available online: https://www.bafin.de/SharedDocs/Veroeffentlichungen/DE/Fachartikel/2014/fa_bj_1401_bitcoins.html), accessed on 8 December 2017.

[312] Cf. Bitcoin-Volatilitätsindex. Available online: https://www.buybitcoinworldwide.com/de/volatilitatsindex/, accessed on 7 December 2017.

[313] By stealing the personal (digital) access key. Cf. BaFin-Fachartikel, 2013, Bitcoins: Aufsichtliche Bewertung und Risiken für Nutzer. Available online: https://www.bafin.de/SharedDocs/Veroeffentlichungen/DE/Fachartikel/2014/fa_bj_1401_bitcoins.html), accessed on 8 December 2017.

Example box: P2P insurance

P2P insurance addresses policyholders' general lack of trust vis-à-vis traditional insurance companies with regard to uncomplicated settlement of claims and the calculation of insurance premiums, which is at times extremely difficult to understand. Interested policyholders can join together into groups as "peers" in order to provide each other with mutual insurance cover based on personalised insurance terms and conditions. In this sense, the principle of P2P insurance is comparable to that of a mutual insurance association (Versicherungsverein auf Gegenseitigkeit, or VVaG). The legal form of a VVaG is also based on the principle of reciprocity: "One for all, all for one."[1] Risks are shared and claims are covered from a shared fund.

In the years to come, P2P insurance could grow in popularity based on new technologies. It is only through the analysis of social media data that people with similar risk profiles can join together, forming collectives (called "groups" in the context of P2P insurance) independent of location and profession.

The first P2P insurance policies observable in the market (insurtechs from Germany, the UK and the US, though without utilising BDAI) are found in the segments of motor vehicle insurance and household insurance. To date, there have been three growth waves for P2P insurance. In the first wave, pure distributors entered the market and provided for a pool of customers with policies from various insurance companies. In the second wave, insurers entered the market selling their own, personalised insurance products to policyholders through online channels. In the current third wave, providers are rushing into market but providing only the technical infrastructure and using new data sources to identify suitable policyholders. These providers (or new ones) could use BDAI to improve their technical infrastructures in a fourth wave.

[1] Farny VW 1975, 90 (92); Weigel. In: Prölss VAG, Section 15, Marginal Note 10.

5.4 Capital markets

5.4.1 Introduction and status quo

The capital markets are an interaction structure in which various players negotiate with each other. In a simplified view, these can be divided into two groups: the sell side (e.g. banks and brokers) and the buy side (investors such as asset managers and insurance companies). Beyond these groups, the capital markets utilise a market infrastructure provided by various providers, including trading venues, clearing houses or custodians. There are many other providers, such as for data, analytics and IT services. On the whole, capital markets are thus already significantly fragmented or specialised. The sell side and buy side together account for a majority of revenues in the capital markets (estimated at around 82 percent in 2016).[314]

The capital market can also be differentiated according to asset classes: The large markets are the fixed income and currency markets[315], the stock market and the commodities market[316], which also all include the associated derivatives markets. In both the fixed income and stock markets, there is a primary market[317] for issuing new securities and a secondary market[318] for trading in existing securities and/or OTC derivatives. The advisory area, which looks – inter alia – after mergers and acquisitions (M&A), can also be considered part of the primary market.

Data and analysis methods are already a significant part of the capital markets. The amount of data generated and made available is growing steadily.

The capital markets are inherently characterised by high levels of technology and strong use of data and analysis methods. Classically, the capital markets use structured data (e.g. securities prices). Market participants have long made use of specialised data and analytics providers. The spectrum ranges from classical providers such as Bloomberg and Reuters to special providers of satellite imagery and the evaluation thereof[319]. The use of data of provable value is subject to many restrictions, e.g. regulations regarding insider trading and the associated subject of front running.

Data generation and availability are expected to continue to grow. The volume of available data increases[320] through such factors as shifting of business to trading venues, which can bring transparency to transactions even for players that are not directly involved in the trade in question. Regulatory specifications can also contribute to growth of data inventories and/or the availability thereof. Some examples of this include the pre- and post-trade transparency rules in MiFID II or the ESEF[321] of the ESMA. One additional driver of the growing data availability and usage in the capital markets is electronic trading in certain asset classes and

---

[314] Cf. Morel et al., 2017, Global Capital Markets 2017. Mastering the Value Migration, p. 7. Available online: http://image-src.bcg.com/Images/BCG-Mastering-the-Value-Migration-May-2017_tcm108-153443.pdf, accessed on 30 January 2018.

[315] Fixed income, currency, and commodities markets are often referred to together as FICC (Fixed Income, Currency and Commodities).

[316] A distinction can be made here between a physical and a derivatives market (e.g. futures in oil or gas).

[317] They are often referred to as DCM (Debt Capital Markets) for debt instruments or ECM (Equity Capital Markets) for equity instruments.

[318] Over-the-counter derivatives are those that are traded outside of stock exchanges.

[319] Cf. Allison, 2017, How Hedge Funds Can Trade on Data Constantly Collected from the Sky. Available online: http://www.newsweek.com/how-hedge-funds-can-trade-data-sky-satellite-fintech-562762, retrieved on 30 January 2018.

[320] In 2014, for example, the NASDAQ OMX Group archived over 0.5 petabytes of data per year; Microsoft, 2014, NASDAQ OMX Group Reduces 2 PB of Data to 500 TB with Microsoft In-Memory Technology. Available online: https://customers.microsoft.com/en-us/story/nasdaq-omx-group-reduces-2-pb-of-data-to-500-tb-with-m, last accessed on 17 February 2018.

[321] European Single Electronic Format.

types enabled by a high degree of process standardisation (electronic trading is, for instance, particularly prevalent in the stock and futures markets, but currently less common in high-yield bonds).[322]

Today's use of BDAI in the capital markets is heterogeneous. It varies between market participants and between the individual areas of each market participant.

Each capital market player deploys BDAI in very different ways. For example, compared to other capital market players, certain hedge funds rely extensively on BDAI to make their investment decisions. Indeed even the purpose for which BDAI is deployed varies. In the case of active asset managers, the purpose of use is, for example, in forecasting price developments or for portfolio analysis and management[323]. The benefit of more data or of new methods of analysis is viewed critically by some market participants, however.[324] Within sales for an investment bank, BDAI may be used in such areas as the representation and assessment of the current market situation or in sales lead forecasting.

Individual players deploy BDAI in very different ways. For instance, BDAI usage in the front office in algorithmic trading is already far advanced (see also the use case "Algorithmic Trading for Order Execution by Institutional Investors"). In other areas of the front office, e.g. for issuing promissory note loans, it is much lower, however. The heterogeneity of BDAI usage continues in the areas of middle office, back office and in the control functions, and is also reflected in the associated data infrastructures.[325]

It must also be noted that many of the existing models (e.g. hedging) in capital markets are already of high quality. Increased use of BDAI may thus only offer small increases in model quality. However, in certain areas of the capital markets, the necessity for continuous improvement and the resulting gains may lead to greater use of BDAI methods. For example, the use of new or BDAI-expanded trading, hedging, risk, or alpha[326] models as well as portfolio and capital optimisation models could lead to increased revenues or improved risk and capital efficiency.

---

[322] In the largest individual market, the currency market with approx. US$5 trillion in daily volume (net-net basis) in April 2016, around 50 percent of transactions are executed electronically; cf. Bank for International Settlements, 2016, Triennial Central Bank Survey. Global foreign exchange market turnover in 2016. As early as 2010, the proportion of automated spot trading transactions was estimated at approx. 25 percent; Bank for International Settlements, 2011, High-frequency trading in the foreign exchange market. Markets Committee, p. 11.

In the US Treasury market, which, with a daily volume of approx. US$500 billion at the end of 2017 is another large single market, a high proportion of benchmark bonds are traded automatically. The electronification is comparable to that of futures; cf. Bank for International Settlements, 2016, Electronic trading in fixed income markets. Markets Committee, p. 10; cf. Sifma, 2018, US Treasury Trading Volume. Available online: https://www.sifma.org/resources/research/us-treasury-trading-volume/, accessed on 19 January 2018. In the case of stocks and futures, 80 percent and 90 percent respectively were being traded electronically as early as 2015, while it was less than 30 percent in the case of high-yield bonds; Bank for International Settlements, 2016, Electronic trading in fixed income markets. Markets Committee, p. 9.

On the stock exchanges of the Federation of European Security Exchanges (plus London Stock Exchange Group and BATS Europe) , approx. 77 percent of the stock trading volume was executed via electronic order books in October 2017; Federation of European Security Exchanges, 2017, European Trading Statistics – European Electronic Order Book Equity Trading.

[323] Cf. also the use case "Integrated Analysis and Portfolio Management Platforms".

[324] Cf. Wigglesworth, 2016, Fintech: Search for the super-algo. In: Financial Times. Available online: https://www.ft.com/content/5eb91614-bee5-11e5-846f-79b0e3d20eaf, accessed on 21 December 2016.

[325] The focus on revenue generation may have represented an obstacle to bolstering BDAI usage in these areas in the past.

[326] Alpha generally refers to the excess returns over a comparison market.

The market is characterised by opportunistic and somewhat "antagonistic" business models of the capital market participants.

The opportunistic and somewhat antagonistic business models[327] lead to a comparatively high pressure to adjust, the fast spread of trends[328] and thus to a high level of readiness to experiment and innovate. The associated mobility is supported by the margins, some of which are high, that allow for the necessary investments in technology and capabilities[329]. Examples of this include the (roughly) sevenfold increase in the CDS volume from 2004 to 2007[330] and – more BDAI-specific – the nearly doubling of quant hedge funds' assets under management to almost US$1 trillion in the past seven years.[331] On the buy side, some large BDAI using hedge funds have attained double-digit billions in assets under management. Currently, bitcoin futures, funds or analyses are a demonstration of the dynamic adaptability of business operations to new market opportunities.

As the market participants currently pursue several strategic thrusts (or may be experimenting with them), there will be more changes to the capital markets as a result of BDAI. It is assumed that these will lead to improvements in efficiency and effectiveness, some of which are more likely than others. Furthermore, more significant changes cannot be ruled out as they have in the past been intrinsic to the sector. For example, at the beginning of the 2000s, approximately 600 cash equity traders were apparently working at a large investment bank in New York, while it is said that there are now – largely as a result of automation – only two (and 200 IT experts).[332]

Overall, increased use of BDAI could lead to a shift in revenues to the benefit of data and analytics providers, niche provider and market infrastructure providers.[333] Earnings from increases in efficiency would presumably only remain with the players temporarily and could be lost in additional margin compressions[334], for example.

---

[328] From 2019, for example, the exams to become a Chartered Financial Analyst are expected to contain questions on AI; cf. Hunnicutt, 2017, "CFA exam to add artificial intelligence, 'big data' questions". In: Reuters. Available online: https://www.reuters.com/article/us-investment-tech-cfa/cfa-exam-to-add-artificial-intelligence-big-data-questions-idUSKBN18J2R4, accessed on 14 January 2018.

[329] Large investment banks are among those included in the top 10 most popular employers among undergraduates in business and engineering worldwide; see Ziegler et al., 2017, World's Top Employers for New Grads. In: CNN Money. Available online: http://money.cnn.com/gallery/news/companies/2017/06/30/worlds-top-employers-new-grads/10.html, accessed on 12 January 2018.

[330] Bank for International Settlements, 2018, OTC, credit default swaps, by type of position. As of: H1 2007. Available online: http://stats.bis.org/statx/srs/table/d10.1?p=20072&c=, accessed on 12 January 2018.

[331] Cf. Wigglesworth, 2018, Quant hedge funds set to surpass US$1 trillion management mark. In: Financial Times. Available online: https://www.ft.com/content/ff7528bc-ec16-11e7-8713-513b1d7ca85a, accessed on 30 January 2017.

[332] Cf. Byrnes, 2017, As Goldman Embraces Automation, Even the Masters of the Universe Are Threatened, In: MIT Technology Review. Available online: https://www.technologyreview.com/s/603431/as-goldman-embraces-automation-even-the-masters-of-the-universe-are-threatened/, accessed on 12 January 2018.

[333] See Morel et al., 2017, Global Capital Markets 2017. Mastering the Value Migration. Available online: http://image-src.bcg.com/Images/BCG-Mastering-the-Value-Migration-May-2017_tcm108-153443.pdf, accessed on 30 January 2018.

[334] Potential margin increases due to cost reductions could be consumed by competitive dynamics.

5.4.2 Likely developments: "more of the same, only faster and better"

Given the already-widespread use of data and analysis methods in capital markets, it is likely that the increased use of BDAI could result in a "more of the same, only faster and better". BDAI could lead to optimisation in the front, middle and back office areas and to additional automation and "algorithmisation" of trading and products as well as to fintechs acting as an "extended workbench" or as an innovation driver in partnership with, or as provider for, incumbents.

BDAI can contribute to optimisation (increases in efficiency and effectiveness) in the front, middle and back office as well as in neighbouring areas.

In principle, BDAI applications in the front office are aimed at effectiveness (e.g. more alpha, locating sources of revenues) as well as efficiency (e.g. use of robotic process automation[335]). Unlike in the middle and back office, however, the effectiveness aims might dominate. For the middle and back office areas, there are also presumably potential increases in efficiency that could partially be tapped by using BDAI methods. These include generic applications such as processing and evaluating unstructured data (e.g. contractual texts) but also the use of predictive analytics in trade processing. It is also probable that additional automation could be driven forward by means of BDAI, for example by using RPA along all touch points (front/middle/back office) including onboarding/Know Your Customer.[336] One example of this would be closing automation gaps in fixed income settlement. All of these applications, however, have a relatively small influence on the structure and shape of the capital markets.

Specific areas of application in the middle and back office can, however, be seen, e.g. in collateral management/optimisation[337], model validation[338] and in the use of liquidity risk models, specifically for asset managers[339]. Significant changes to collateral management could also affect capital markets due to changes in demand for high quality assets. Finally, there has also been progress in the automation of contract creation for derivatives.

Regulatory reporting duties also require effective processes that can be bolstered and made more efficient by the use of BDAI. These include real-time reporting, internal compliance and monitoring. BDAI is already used for compliance and monitoring, e.g. trade monitoring and evaluating electronic communication such as email traffic. The individual capital market players (particularly banks) can also use this type of communication internally in the area of compliance, for example in the evaluation of unstructured data (e.g. communication data).[340] Activities are already being observed by regtechs in this field in particular.

---

[335] Disregarding specific models, the front office is showing evidence of increased use of robotic process automation, e.g. for the automatic recording of specific email content and the execution thereof; cf. Arnold et al., 2017, Robots enter investment banks' trading floors. In: Financial Times online. Available online: https://www.ft.com/content/da7e3ec2-6246-11e7-8814-0ac7eb84e5f1, accessed on 30 January 2018.

[336] Research in investment banks is not considered separately here.

[337] This is the optimisation of collateral provisions, which also have a pricing component in derivatives, cf. e.g. Kondratyev et al., 2017, MVA Optimization with Machine Learning Algorithms. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2921822, accessed on 31 October 2017.

[338] Model validation means showing that models fulfil their task as expected and in accordance with their design and the business requirements (e.g. use of machine learning; cf. Woodall, Model risk managers eye benefits of machine learning. In: risk.net. Available online: https://www.risk.net/risk-management/4646956/model-risk-managers-eye-benefits-of-machine-learning, accessed on 12 January 2018.).

[339] Cf. Kilburn, 2017, Firms race to apply machine learning to liquidity risk models. In: risk.net. Available online: https://www.risk.net/asset-management/5328636/firms-race-to-apply-machine-learning-to-liquidity-risk-models, accessed on 12 January 2018.

[340] Cf. van Liebergen, 2017, Machine Learning: A Revolution in Risk Management and Compliance? Institute of International Finance Research Note. Available online: https://www.iif.com/publication/research-note/machine-learning-revolution-risk-management-and-compliance, accessed on 30 January 2018.

In conclusion, it can be noted that BDAI applications are becoming more common in more functions supporting the front, middle and back office areas and in compliance processes.

BDAI allows continued automation and "algorithmisation" and thus increased machine/machine interaction in trading, sales and products. This could lead to an associated acceleration of processes and could also affect advisory areas.

Today, many trade interactions between capital markets participants are already heavily standardised, digitalised, automated and "algorithmised" (see also the use case "Algorithmic Trading for Order Execution by Institutional Investors"). For this reason, BDAI is not expected to lead to any significant changes in the short term here. All the same, an increase or an expansion in algorithmic trading – not necessarily of the high frequency type – can be expected across all asset classes. This could allow existing automation and digitalisation gaps to be closed. One example of a market with still-greater automation potential and/or that still offers significant niches is fixed income, there for example in the electronic trading of corporate or high-yield bonds. The driver behind this could be algorithmic trading (as can be seen, for example, in corporate bonds trading[341]), which requires shared standards and efficient processes. This would allow the proportion of machine/machine interactions in capital markets to continue to increase.

In addition to the use of BDAI applications in trading, increased automation and "algorithmisation" in sales for asset management is also apparent. On the one hand, the use of BDAI can be seen in pricing strategies[342], while on the other an increase in the number of robo-advisors[343] can also be observed.[344] However, many tend to involve comparatively low levels of BDAI complexity at the moment. Furthermore, such offers from incumbents show some considerably greater volumes than those from new fintech providers[345]and, in some cases, certain incumbents maintain shares in these fintechs[346]. Furthermore, BDAI is seen to be used increasingly to support cross-selling in the retail area of asset management (e.g. of asset management products to customers of bank products). In wealth management, machine learning can also be seen to generate proposals for customers as a tool for employees (known as "next-best action").[347]

---

[341] Cf. The Economist, 2017, Digitalisation shakes up corporate-bond markets. Available online: https://www.economist.com/news/finance-and-economics/21721208-greater-automation-promises-more-liquidity-investors-digitisation-shakes-up, accessed on 19 November 2017.

[342] This includes the use of BDAI to determine optimal levels of management fees; a similar situation applies to security services (e.g. custodians, clearing houses).

[343] (Retail) asset management or intermediation products with a fully automated customer interface that use algorithms to determine and, where applicable, manage portfolios based on customer information. Hybrid approaches (human/machine, cyborg advisor) and advisor platforms can already be observed in the (fintech) market.

[344] Cf. Chapter 5.2, Banks.

[345] Cf. World Economic Forum, 2017, Beyond Fintech: A Pragmatic Assessment Of Disruptive Potential In Financial Services, p. 136.

[346] As an example, Jessop et al., 2017, BlackRock takes Scalable Capital stake in Europe "robo-advisor" push. In: Reuters. Available online: https://uk.reuters.com/article/us-blackrock-scalablecapital/blackrock-takes-scalable-capital-stake-in-europe-robo-advisor-push-idUKKBN19A322, accessed on 12 January 2018.

[347] Cf. Son, 2017, Morgan Stanley's 16,000 Human Brokers Get Algorithmic Makeover. In: Bloomberg. Available online: https://www.bloomberg.com/news/articles/2017-05-31/morgan-stanley-s-16-000-human-brokers-get-algorithmic-makeover, accessed on 15 January 2018.

Processes and decisions that are traditionally based more on human interaction and appraisal, such as in advisory, could be supported in the future by more BDAI applications. For example, BDAI is already being used to evaluate when assessing what or if a time might be suitable for an ECM[348] or DCM[349] engagement. The use of BDAI is also conceivable in personnel-intensive applications such as due diligence or other legal processes. Legaltechs could make these processes scalable. Furthermore, the use of BDAI is also conceivable for affecting opinions regarding certain aims in capital markets (e.g. in shareholder activism) – something not entirely unproblematic from the consumer viewpoint.[350]

Finally, overarching BDAI usage could also be found in algorithmic product development.[351] Simply put, it is conceivable that BDAI could allow market developments and trends to be incorporated automatically into the design of financial products. Bigtechs might enjoy an advantage here, as already stated, because the constant data-driven optimisation of products is already part of their business and operational model.[352]

Overall, the additional linking of machines and applications across market participants could generate higher speeds, resulting in situationally more fluctuating liquidity and episodes of higher volatility.[353] Self-reinforcing interactions could thus result in increasing market instabilities.

Fintechs have, so far, evolved well particularly in partnership with incumbents in capital markets. They often act as an extended workbench, ideas generator or a specialist provider for process changes and efficiency increases.

In the case of fintechs, only a limited direct influence[354] has been observed on the capital markets as a whole.[355] So far, they have been seen particularly in the areas of data and analytics and in execution (and/or in execution technologies and associated infrastructures). Broadly speaking, a distinction can be made between three approaches:

1. Fintechs cooperate with or are backed by and/or are integrated systematically into incumbents' processes. That way fintechs would not act as independent "displacers", but more as an "extended workbench" and as a provider of ideas for quick implementation of BDAI applications.

2. Fintechs function as a specialist provider for financial service providers and other incumbents in capital markets. One such example includes suppliers of program libraries[356] for fast numerical differentiation, something that is already widely used for assessing exposures and risks. This could allow efficiency increases in the relevant areas of banks or asset managers. Another example would be the optimisation of portfolios of (exchange traded) derivatives with regard to various market forecasts. This would entail

---

[348] Equity Capital Markets.

[349] Debt Capital Markets.

[350] The recent discussion regarding the influencing of voter opinions suggests that methods of this type are indeed a reality.

[351] This is used in the textile sector by some start-ups, for example; cf. e.g. https://www.fastcodesign.com/90128248/how-stitch-fix-is-using-algorithmic-design-to-become-the-netflix-of-fashion, accessed on 30 January 2018.

[352] One example of this is the use of A/B tests to identify promising product changes.

[353] See the discussion in the use case "Algorithmic Trading for Order Execution by Institutional Investors".

[354] For example, according to a survey in 2017 only about 36 percent of AI-based capital markets fintechs were disruptive in nature. BCG FinTech Control Tower.

[355] Cf. the aforementioned example from a study by the World Economic Forum; World Economic Forum, 2017, Beyond Fintech: A Pragmatic Assessment Of Disruptive Potential In Financial Services. Overall the number of fintechs in the direct capital market is small compared to retail or corporate banking: out of approx. 8,000 fintechs tracked by BCG in 2016, only approx. 600 were located in the capital market; see Morel et al., 2016, Fintechs in Capital Markets: Land of Opportunity. Available online: https://www.bcg.com/de-de/publications/2016/financial-institutions-technology-digital-fintech-capital-markets.aspx, accessed on 30 January 2018. For principal trading firms and hedge funds using crowdsourcing for investment algorithms see next section.

[356] A collection of functions and routines to be used by computer programs.

consideration of billions of combinations. One possible option that would help reduce legal uncertainty and/or enable decisions to be taken faster and more effectively includes the legal interpretation and forecast of court decisions – think "legaltech".[357]

3. Fintechs also act as independent providers of financial services (e.g. trading platforms). This approach could be more difficult to implement. For core areas such as SEFs[358], for example, the market share of independent fintechs found there has historically been low.[359] One reason for that could be the ability of incumbents themselves to adapt as well as their potentially self-disruptive actions[360], which could make independent entry of fintechs into (core) areas of capital markets more difficult.

However, the following must be noted in terms of the actual development of fintechs: Incumbents such as investment banks in particular invest in regulatory projects or have already invested significant resources in them (e.g. in the implementation of MiFID II). As a result, as relevant players they have already contributed to the development of fintechs in capital markets as part of the first two approaches. If incumbents' capacities were to free up again, however, projects previously outsourced to fintechs might again be carried out by the incumbents themselves.

5.4.3   Further developments: higher connectivity and complexity

In addition to "more of the same, only faster and better", the increased use of BDAI could also have further-reaching and deeper consequences – potentially even reorganising the market landscape. On the other hand, BDAI applications could lead to increased transparency in capital markets both for interaction with business partners as well as for creating products. Each new model- and data-driven interaction or product so produced strengthens the links between players and markets. This could lead, overall, to higher connectivity and a higher level of complexity in the capital markets.

---

[357] Discussed e.g. in Alarie et al., 2016, Regulation by Machine. In: 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

[358] Swap Execution Facility, a trading platform.

[359] Cf. World Economic Forum, 2017, Beyond Fintech: A Pragmatic Assessment Of Disruptive Potential In Financial Services, p. 178.

[360] According to a study by the World Economic Forum, for example, market infrastructure providers can act self-disruptively; cf. World Economic Forum, 2017, Beyond Fintech: A Pragmatic Assessment Of Disruptive Potential In Financial Services, p. 170.

5.4.3.1 Reorganisation within the market landscape

BDAI makes specialisation and the emergence of new market participants easier, thus potentially leading to further fragmentation of the value chains in capital markets.

As the capital market is heavily data-driven, BDAI in principle makes it easier to turn value chain components into (several) independent models, i.e. fragmentation due to specialisation[361]. Already specialist providers such as principal trading firms[362] and providers of post-trade analysis tools to verify best execution[363] have established themselves in the market. This development can, at the same time, be reinforced by the "freeing up" of certain roles in capital markets, such as incumbents focusing on trading in an agent capacity[364].

Furthermore, a combination of already-completed specialisation and fragmentation of the value creation can be observed along with the pursuit of greater effectiveness when using BDAI. Examples of this include insurance companies and pension providers. These players already procure certain BDAI capital market services externally (e.g. in the area of alternative investments). The indirect use of BDAI can now be seen in the selection of the suppliers here, e.g. hedge funds.

Bigtechs are already profiting from BDAI usage in capital markets because they are important infrastructure providers supplying cloud computing. However, it is also conceivable that bigtechs extend their core business and thus secure a stronger bigtech presence in the capital markets. To this end, for example, the use of the existing customer reach, existing process chains and client analytics capabilities could form part of the large-scale (pure) distribution of retail financial products. New business models in certain areas of capital markets are also imaginable. For instance, bigtechs could potentially act as asset managers.[365] It is possible that the focus would then not only be on the collection of fees, but also on expanding the (data) ecosystem in question. An affiliate of Alibaba already operates the largest money market fund in the world (Yu'E Bao). Direct monetisation of relevant data by means of special offerings for financial service providers is also conceivable.

BDAI could establish new players in the capital markets, which as already mentioned could mean an increase in the degree of interdependence and complexity in the capital markets.

---

[361] E.g. due to the ability to automate a wide range of tasks using BDAI and easy access to powerful BDAI tools such as corresponding program libraries.

[362] These are (high frequency) trading firms that take on risk and use algorithms to, for example, execute customer orders or carry out market making and to trade for their own account. They may act as liquidity providers.

[363] Banks and financial services institutions must take the necessary steps when executing orders to make sure that the best possible result is generally attained; cf. Federal Financial Supervisory Authority, 2015, Wertpapiergeschäfte – Was Sie als Anleger beachten sollten. [Securities Trading – What to Look Out for as an Investor] Available online: https://www.bafin.de/DE/Verbraucher/GeldanlageWertpapiere/Wertpapiergeschaefte/wertpapiergeschaefte_artikel.html, accessed on 14 January 2018. The brochure is currently being revised.

[364] Trading without taking on any own risk.

[365] See Mnyanda, 2017, Tech Firms Are The Next Threat to Asset Managers, Moody's Says. In: Bloomberg Markets. Available online: https://www.bloomberg.com/news/articles/2017-11-14/tech-firms-are-next-big-threat-to-asset-managers-moody-s-says, accessed on 12 January 2018; see Takeo et al., 2017, World's Biggest Pension Fund Says AI Will Replace Asset Managers. In: Bloomberg. Available online: https://www.bloomberg.com/news/articles/2017-12-14/world-s-biggest-pension-fund-sees-ai-replacing-asset-managers, accessed on 18 December 2017.

Using BDAI could cause the number of interfaces to grow and standardisation to become more widespread in the front, middle and back office. Indeed, make-or-buy decisions may have to be re-evaluated.

BDAI applications benefit from easy and standardised data access. As a result, the implementation of BDAI in front, middle and back office can contribute to higher standardisation there. This, in turn, could further the spread and utilisation of interfaces in the capital markets. One example would be a universal middle and back office communication protocol, allowing post-trade information to be exchanged between them in a similarly universal manner as generally pre-trade and trade information. Increased use of regulatory interfaces would also be possible. Another example of this would be data exchange by model-driven machine executable regulatory reporting, which is direct machine/machine communication as part of reporting between the supervisory authority and the entity being regulated.[366]

The creation of new interfaces and further standardisation of processes could lead to re-evaluation of make-or-buy decisions. Some investment banks, for example, are already providing clients with access to their risk management systems.[367] Smaller players which operate their capital market business subcritically and as an adjunct to their main business,[368] could potentially make greater use of white labelling.[369] Standardisation and available interfaces could mean that these players might be motivated to counterbalance their lack of scale in their subcritical adjunct business in this way. At the same time, the market risks of these players would be reduced.

In extreme cases, the consistent use of interfaces could lead to even greater fragmentation of value creation up to and including complete digitalisation of all of a player's tasks. The use case "Crowdsourcing of Investment-Algorithms" already shows the first moves in this direction.

The role of data, analytics and process providers in capital markets becomes more relevant with the use of BDAI. As a result, there could be further concentration within these provider groups.

As already mentioned, the ongoing "algorithmisation" and automation is based on the availability of data and algorithms. This increases the relevance of providers in that field. Existing providers sometimes expand their offering by including additional new providers and by expanding their own data and analytics offerings. These providers of data and information platforms open them up to third-party providers who can then provide their data and applications to a larger clientele.[370] One prime example of the expansion of one's own offerings is the use of machine learning in liquidity models for asset managers.[371]

As value chains break up, more capital market players could start to focus on core competencies. In turn, a few select players could close the gaps that opened in the value chain with appropriate offerings. This approach can already be seen in integrated analysis and portfolio management platforms (see use case "Integrated Analysis and Portfolio Management Platforms").

---

[366] See Financial Conduct Authority, 2017, Model-driven machine executable regulatory reporting. Available online: https://www.fca.org.uk/firms/our-work-programme/model-driven-machine-executable-regulatory-reporting, accessed on 12 January 2018.

[367] For example: Baer, 2016, Goldman Sachs Has Started Giving Away Its Most Valuable Software. In: The Wall Street Journal. Available online: https://www.wsj.com/articles/goldman-sachs-has-started-giving-away-its-most-valuable-software-1473242401, accessed on 12 January 2018. This enjoys a certain tradition within investment banking: RiskMetrics Group, for example, is a spin-off of J.P. Morgan & Co. from 1998.

[368] I.e. without effective use of economies of scale.

[369] I.e. to put it simply, firms use services from third-party providers and offer them under their own name.

[370] Cf. e.g. Bloomberg Application Portal (https://www.bloomberg.com/professional/product/application-portal-terminal-connect/, accessed on 30 January 2018)

[371] For example, Bloomberg (https://www.bloomberg.com/company/announcements/bloomberg-delivers-first-quantitative-model-for-calculating-liquidity-risk/, accessed on 30 January 2018).

As part of these developments, and due to economies of scale, there may be an increased concentration on incumbents within the groups of data, analytics and process providers. It is also conceivable that self-reinforcing data generation structures could have a positive impact on existing providers. In this way, capital market players would use the data provided to generate new data, which in turn would be used by data, analytics and process providers for their services.

5.4.3.2 Targeted increases in transparency

The use of BDAI could increase transparency for market participants, enabling them to improve their competitive positions and create new products.

In various areas of capital markets, participants are using BDAI in their attempts to secure greater transparency for themselves. The areas of sales, trading and asset management stand out in particular. Notably, more transparency could make it easier to create new products, e.g. by hedging events hitherto excluded from this possibility, and to "financialise" further societal and economic sectors, i.e. to make it possible to trade in and price ever more areas of life.

In terms of sales optimisation and/or situational awareness/market intelligence for sales and trading, it is observable that BDAI is increasingly used to generate transparency. Large sets of data are frequently evaluated in real time and machine learning is applied to support decision making. Increased transparency in sales and trading can also be seen in the use of AI as a pricing aid in bond trading. Such approaches are being implemented by incumbents themselves and in partnership with fintechs or by other suppliers.

In asset management, the pursuit of transparency can be seen for instance in the increasing use of new (and unstructured) data, e.g. specially conditioned web and social media data for predictive analytics. Given the generally increase in data availability, new transparency possibilities could open up. It is conceivable that linking up what is initially non-market-relevant data[372], such as information derived from the digital usage behaviour of corporate employees, could open up new possibilities for evaluating companies.

BDAI methods can also be used for internal planning and control. This could, for example, take the form of BDAI-supported asset liability management to control economic risk provisioning by banks, insurers and pension companies.[373] This approach could help to make better use of risk-mitigation techniques such as interest rate or longevity[374] hedges.

Furthermore, the greater transparency afforded by BDAI presents the opportunity of designing new products in the capital markets. Even now, financial indices are offered that are created with the aid of natural language processing. In this case, annual reports, for example, are evaluated in order to extract certain information about companies and then to add these companies to (new) indices.

---

[372] For the use of alternative data such as social media by asset managers, cf. Ram et al., 2017, When Silicon Valley came to Wall Street. In: Financial Times. Available online: https://www.ft.com/content/ba5dc7ca-b3ef-11e7-aa26-bb002965bce8, accessed on 14 January 2018.

[373] Cf. e.g. Schad et al., 2017, Big-data-powered, next-generation asset liability and liquidity risk management. In: IBM Banking Breakthroughs. Available online: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=ASW12435USEN, accessed on 14 January 2018.

[374] Hedging life expectancy risks.

BDAI also allows greater transparency for ever-smaller risks. With sufficient transparency, it would be possible to bundle these small risks together. In turn, this bundling could make it possible to operate volume businesses with these risks. Resulting products could open up new trading options. If this helps previously untraded risks to become tradable, new areas of life and of the economy could be financialised. One example of a financialisation of individual risks could be the large-scale securitisation of future income after completing a university degree.[375] This is an area in which data provided by bigtechs could find additional use. There is also the option of expanding financial services into virtual economies and/or to financialise their assets.[376]

From a macroeconomic perspective, financialisation could mean that the share of financial services providers in the economy's overall value creation could grow. At the same time, many areas in society and other economic sectors could be linked to the financial industry which previously were less associated with it.

Valuation frequencies for illiquid assets could increase due to BDAI. This could contribute to increased market dynamics.

There are signs of an increased use of BDAI in the valuation of illiquid assets. Some providers are already offering services for residential and commercial real estate that allow for the estimation of transaction prices by means of BDAI.[377] Unlike the regularly occurring valuations by experts, this approach could allow for more up-to-date and continuous valuation of these assets.

Furthermore, expanded estimations of macroeconomic parameters have existed for longer, at least in the academic world. These include the more up-to-date estimation of inflation rates by evaluating large sets of price data.[378] These types of data and processes could be used more frequently in the evaluation and creation of products in capital markets.

Consequently, assets that were previously subject to low observable levels of market value fluctuations may experience more pronounced and faster recognition of market value changes in the future. This would generate new volatilities in certain areas of the capital markets and thus increase the market dynamics.

---

[375] This is not a new idea in itself; it already exists on a smaller scale. Cf. Chaparro, 2017, Investors are paying college students' tuition — but they want a share of future income in return. In: Business Insider. Available online: http://www.businessinsider.de/income-share-agreements-help-students-pay-for-college-loan-alternative-2017-3?r=US&IR=T, accessed on 18 January 2018.

[376] There are marketplaces for virtual assets (e.g. for objects in computer games) where these can be traded for money. The overall costs of virtual assets in games are not a trivial matter; cf. e.g. Moore, 2014, A closer look at the 22-hour Eve Online battle that cost gamers over £180,000. In: Wired. Available online: http://www.wired.co.uk/article/eve-online-battle, accessed on 30 January 2018.

[377] This involves BDAI-based further development of automated valuation methods.
[378] Cf. the "The Billion Prices Project" by the MIT Sloan Management School (http://www.thebillionpricesproject.com/, accessed on 30 January 2018).

5.4.4 Use cases

Currently, various use cases can be observed in the capital markets; a selection are detailed below. Despite the complex and varied landscape of players and markets as described above, these specific use cases can be assigned to specific areas of use of BDAI:

Customer interface: algorithmic trading for order execution by institutional investors. As described in the market analysis, BDAI allows for an increase in automation and "algorithmisation" and thus for increased machine/machine interaction in trading. This use case (cf. Chapter 5.4.4.1) shows that algorithmic trading generates a complex ecosystem of data and algorithms. Economic and regulatory discussions of the implications are already quite advanced.

Core processes: integrated analysis and portfolio management platforms. This use case (cf. Chapter 5.4.4.2) shows how players could make use of economies of scale that arise from data, analytics and process providers and which opportunities and risks may arise for markets and users. For example, the benefits that arise from powerful tools are explored, but also possible risks that might be posed by the widespread use of just a few models or platforms.

New business models: crowdsourcing of investment algorithms. This use case (cf. Chapter 5.4.4.3) shows how new algorithms can be used to break up value chains in capital markets. Here a classic component of core processes – the development of investment algorithms – is externalised. The use case also offers a perspective on possible cooperation structures that BDAI might usher into the capital markets.

5.4.4.1 Algorithmic trading for order execution by institutional investors

1. Introduction

Because of their business models, institutional investors such as insurers, asset managers or pension funds sometimes move large volumes of securities at short notice. For example, in 2016, trading in ETFs accounted for almost 30 percent of US stock exchange trading volume by value,[379] and ETF block trading[380] exceeded US\$3,000 trillion in 2016 (approx. 80 percent of which was in equity ETFs).[381] To achieve the highest possible process reliability while maintaining effective and efficient execution, large orders in particular are executed using algorithms.[382] In October 2017, for instance, over 10 million equity transactions were executed on Deutsche Börse through the electronic order book.[383] Volumes such as these could not be handled manually.[384]

---

[379] Wigglesworth, 2017, ETFs are eating the US Stock market. In: Financial Times. Available online: https://www.ft.com/content/6dabad28-e19c-11e6-9645-c9357a75844a, accessed on 27 December 2017.

[380] Here, a block trade is a transaction involving more than 10,000 shares or an amount in excess of US\$200,000.

[381] Lin, 2017, Trends in ETF Block Trading. In: Credit Suisse – Trading Strategy.

[382] An algorithm might however still be found on the other side of a smaller, non-algorithmic trade order.

[383] Cf. Federation of European Security Exchanges, 2017, European Trading Statistics – European Electronic Order Book Equity Trading. October 2017. The number of accounts only for the actual transactions – not all of the orders that were entered, their changes and cancellations.

[384] One algorithmic market maker has, in its own words, an infrastructure of more than 30PB hard drive space and 300 TB RAM (cf. https://www.xtxmarkets.com/about, accessed on 29 January 2018, a quantitative-driven electronic market-maker).

This use case examines the BDAI-supported interaction of different market participants at a capital market customer interface. Institutional investors use algorithms to execute orders for liquid asset classes such as equities, foreign currencies and exchange-traded derivatives.[385] (As a matter of classification: execution algorithms or order execution using algorithms are often not considered part of high-frequency trading.[386]) Counterparties are frequently other institutional investors, market makers and banks, i.e. the trade often occurs between institutional market participants.

This use case exclusively considers algorithmic trading for order execution by institutional investors who are the customers of such offers from investment banks for instance. These transactions mainly aim to reduce transaction costs and respectively market impact[387] (unlike algorithmic trading aimed at exploiting price patterns or arbitrage opportunities).[388] Retail investors would currently only utilise algorithms such as these indirectly when their orders are bundled (e.g. by a bank and an asset manager).

Algorithmic trade execution involves processing large data volumes and streams in real time; accordingly, it is developed using large data sets and is therefore a classic example of a BDAI application (with methods initially inspired by signal processing approaches).

Computerised trading, particularly for executing orders, has existed since the 1970s[389] and is now widely used. In some markets, over 50 percent of some products are already traded completely – i.e. both selling and buying – using algorithms.[390] So, BDAI applications are already highly relevant to the market.

2. Application

The following aims to exemplify the use. An institutional investor enters certain parameters for an order into the screen of a trading system and the underlying algorithm addresses different trading venues, for example. When a request reaches a trading venue, it might trigger other algorithms of capital market participants that are also active in that marketplace (e.g. market making algorithms). This interaction of algorithms is an important aspect within the trading venues, and the mutual dependencies are already creating something like independently-operating ecosystems.[391]

---

[385] The most advanced algorithms are generally found in markets with central limit order books, cf. Bank for International Settlements, 2016, Electronic trading in fixed income markets. Markets Committee, p. 7.

[386] For example, high frequency trading (HFT) is frequently associated with (passive) market making or directional and arbitrage strategies, cf. Deutsche Bundesbank, 2016, Bedeutung und Wirkung des Hochfrequenzhandels am deutschen Kapitalmarkt. In: Monatsbericht Oktober 2016, pp. 37–61; cf. Bank for International Settlements, 2016, Electronic trading in fixed income markets. Markets Committee, p. 8; cf. Bank for International Settlements, 2011, High-frequency trading in the foreign exchange market. Markets Committee, pp. 3–6. The International Organization of Securities Commissions places HFT usually in Proprietary Trading Firms (or desks), cf. International Organization of Securities Commissions, 2011, Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency. Consultation Report CR02/11 of the Technical Committee of the International Organization of Securities Commissions, p. 21.

[387] Market impact here refers to the effect that orders and transactions have on the price of a security. For example, a major buy order can cause the price to rise.

[388] Cf. Bank for International Settlements, 2016, Electronic trading in fixed income markets. Markets Committee, p. 8. The BIS groups automated trading into three categories: the use of algorithms for trade execution, market making and for directional, relative value and arbitrage strategies, whereby HFT is particularly associated with market making.

[389] Computer-based since 1977, cf. Farmer et al., 2011, An ecological perspective on the future of computer trading, In: The Future of Computer Trading in Financial Markets – Foresight Driver Review – DR6, Government Office for Science, p. 7.

[390] Federal Register, Vol. 80, No 242, at 78826.

[391] Cf. Farmer et al., 2011, An ecological perspective on the future of computer trading, In: The Future of Computer Trading in Financial Markets – Foresight Driver Review – DR6, Government Office for Science.

Configuration options for trade orders

Institutional investors have a choice between many algorithms, strategies, parameters[392] and order types on the providers' platforms, which often extend well beyond the spectrum of functionalities offered by the trading venues (in the simplest case, the choice of the provider is the first or only parameter to be selected).[393] Such execution algorithms are widely used and are offered, for instance, by many investment banks.

Figure 25 (stylised example): The customer selects and determines the details for the execution mechanism made available by a provider (e.g. from a bank or a trading platform). Selecting the execution mechanism becomes an integral element of the intended trade.

---
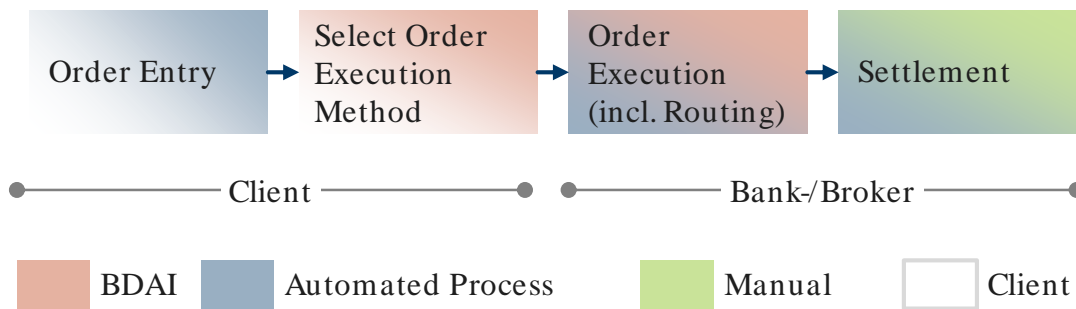
[392] Examples include: How much of the entire order is disclosed in each case; The maximum share of the volume offered in the order books that will be taken; Parallel or serial requests to trading platforms. Also, the choice of algorithm could even involve BDAI – either directly by the client or by the provider offering such functions.

[393] Some trading venues offer their own algorithms or more complex order types (basically encapsulated algorithms).

3. Technology and data

Algorithmic trading is always a technological package solution. The trading venues provide the necessary market infrastructure and the connection to other market participants.

Some providers already offer institutional investors AI-based algorithms, although usage of true AI-based algorithms is still not a dominant phenomenon[394] in trade execution for institutional investors. Instead, statistical and heuristic methods are widely used.[395]

Specialised algorithms anticipate algorithmic behaviour

The algorithms used have differing levels of complexity and many variations. Some market participants try to exploit the smallest market fluctuations using algorithms that are not aimed (just) at order execution. This could in turn result in institutional investors using algorithms that anticipate this behaviour and in response use "counter-algorithms" tailored for this.[396] Another (simple) example would be "iceberg" orders[397]. Here, market participants attempt to avoid showing the quantity of their total order but certain algorithms detect these iceberg orders and exploit them. Simpler algorithms such as iceberg orders are frequently offered by the trading venues themselves.

Development of algorithms using big data from order books and trading

The algorithms are developed on the basis of large sets of historic data and can be considered a subset of BDAI. These algorithms are typically developed, tested and applied with the help of trade and order book data. For example, order book data includes at any point in time prices for buy and sell offers along with the corresponding volumes as well as prices and volumes for buy and sell offers not at the best price in each respective case.

4. Opportunities

It is not retail customers that mainly and regularly use these algorithms but rather, for example, by asset managers. If these users are able to process large trading orders more efficiently, they could reduce their costs or the transaction costs incurred by the products, which could in turn benefit retail customers.

Algorithms optimise execution of large volumes

Very large trading orders tend to be more difficult to execute. The reason for this is that there are few counteroffers for the entire volume for these transactions, so that the principal has to accept a price premium or discount due to the market impact. To avoid this, algorithms are used to split the large orders into a great many smaller orders and to execute these at the prevailing market price.

Institutional investors utilise price benchmarks, for example, such as volume-weighted prices to trade larger volumes without themselves (unduly) impacting the market. These benchmarks are the anchor point for some algorithms used for order execution.

---

[394] Cf. e.g. Noonan, 2017, JPMorgan develops robot to execute trades. In: Financial Times. Available online: https://www.ft.com/content/16b8ffb6-7161-11e7-aca6-c6bd07df1a3c, accessed on 10 December 2017.

[395] Machine learning/artificial intelligence already plays an important role in the development (and increasingly in the selection) of such algorithms. The development of algorithms, following chapter III, generally requires significantly more memory and computing capacity than the application does, so that elastic environments could represent a solution here (cf. Chapter 4.2). The difference between production and development occurs, among other things, because during development the data is analysed or used many times in a short period of time.

[396] Cf. Tse et al., 2012, AES Analysis. High Frequency Trading – Measurement, Detection and Response. In: Credit Suisse –Trading Strategy. Available online: https://edge.credit-suisse.com/edge/Public/Bulletin/Servefile.aspx?FileID=23285&m=923226224, accessed on 13 November 2017.

[397] Can be available directly as an order type or created by an algorithm.

Economic and process efficiency and effectiveness

The clear advantages of algorithmic trading explain the widespread use of algorithms in executing orders by institutional investors. From a technical perspective, the focus is on efficient processing, i.e. high processing speed, avoidance of human error, fast response to market changes, the ability to simultaneously address several trading venues and the associated accuracy of the trade execution. From an economic point of view, algorithmic trading makes it possible to effectively achieve a number of business objectives: The market impact can be reduced for specific trading orders and the trading intent can be disguised, thus allowing costs to be optimised. Whether total liquidity is increased as a result of algorithms as a whole – particularly through HFT – in all or in which particular situations or is increased in terms of individual trading venues is still the subject of some scientific debate.[398]

Lower barriers to entry and improved basis for monitoring

From the perspective of the overall market, the availability of efficient algorithmic execution gives smaller players a better chance of becoming active, as it might not be necessary to have proprietary algorithms to execute orders, thus reducing the initial investment. In this way, the widespread use of algorithms supports greater market efficiency by removing barriers to entry into the market. As transactions are executed completely electronically, they are already currently being recorded, which allows them to be evaluated from a supervisory and regulatory perspective. Records could also be kept by the provider on the respective state of the algorithm (e.g. key statistical figures on the functionality of the algorithm), making it possible to reconstruct the trading process using the algorithm's "intent".[399]

---

[398] It is not scientifically conclusively proven that algorithmic trading increases liquidity overall (cf. e.g. also Bank for International Settlements, 2016, Electronic trading in fixed income markets. Markets Committee.). The German Central Bank monthly report for October 2016 indicates, for instance, that HFT contributes to better liquidity during calm market phases, while during highly volatile phases it can produce temporary reductions (cf. Deutsche Bundesbank, 2016, Bedeutung und Wirkung des Hochfrequenzhandels am deutschen Kapitalmarkt. In: Monatsbericht Oktober 2016, pp. 37–61). Another study by the ESMA also focuses on HFT and the possibility that liquidity is overestimated in fragmented markets (European Securities and Markets Authority, 2016, Order duplication and liquidity measurement in EU equity markets. In: ESMA Economic Report No. 1, 6 June 2016). For further discussion, cf. also Farmer et al., 2011, An ecological perspective on the future of computer trading, In: The Future of Computer Trading in Financial Markets – Foresight Driver Review – DR6, Government Office for Science.

[399] The question of whether this happens traditionally in a database or version control system, or if even blockchain technologies are potentially used, remains open. From a technological perspective, saving the relevant data is not a problem per se. For instance, using blockchain to create an audit trail for AIs is discussed by Corea, 2017, The convergence of AI and Blockchain: what's the deal? In: Hackernoon. Available online: https://hackernoon.com/the-convergence-of-ai-and-blockchain-whats-the-deal-60c618e3accc, accessed on 5 December 2017. When it comes to saving trading and order data, HFT is subject to clear requirements based on MiFID II.

5. Risks

Added costs could occur as a result of large market movements or errors. Even if losses resulting from erroneous algorithms may not be charged to the consumer, they may still potentially result in indirect costs, such as increased risk costs, which could affect the price of investment products.

Possible impact of algorithms and their interaction

Market movements triggered or exacerbated by algorithms can directly change the market value of investments.[400]

Algorithms can be "exploited" by third parties. An illustrative example of this is called "algo sniffing". For example, an algorithm targets VWAP[401] algorithms in order to buy shares faster than them and to then profit from a potential price increase as a result of the subsequent buying by the VWAP algorithm.

Error and loss control could be more difficult

Due to the speed and efficiency of the algorithms, erroneous algorithms or the incorrect set-up or deployment of algorithms can result in heavy losses before the source of the problem can be identified and isolated.[402] Furthermore, the complexity of the algorithms and the trading environment can also make it more difficult to detect errors (and any related damages).

In addition, there is strong pressure to adapt in the investment management industry. Because many companies use algorithms, other companies can be forced to use such algorithms, too, as simple execution methods might no longer be effective when they encounter algorithms.

Algorithms can exacerbate short-term market trends. The afore-mentioned volume-weighted price algorithms are constructed in such a way that transactions are triggered in the vicinity of the volume-weighted price. When a larger transaction is executed in the market it triggers further transactions at the same price level, which can exacerbate a price trend. [403]

As a rule, algorithms are not only passive market participants but can be utilised to (manipulatively) change market conditions in order to benefit from these changes (one example would be momentum ignition, i.e. causing or intensifying a trend by placing certain orders or series of orders[404] in order to then benefit from these trends). Such practices are by no means limited to execution algorithms and are more related to directional strategies.[405]

---

[400] Presumably less as a result of the execution algorithms alone but rather through interaction with algorithms with different objectives.

[401] Volume-weighted average price, VWAP

[402] E.g. cf. Philips, 2012, Knight Shows How to Lose $440 Million in 30 Minutes. In: Bloomberg. Available online: https://www.bloomberg.com/news/articles/2012-08-02/knight-shows-how-to-lose-440-million-in-30-minutes, accessed on 10 December 2017.

[403] Cf. Farmer et al., 2011, An ecological perspective on the future of computer trading. In: The Future of Computer Trading in Financial Markets – Foresight Driver Review – DR6, Government Office for Science, p. 8.

[404] Cf. European Securities and Markets Authority, 2012, Guidelines. Systems and controls in an automated trading environment for trading platforms, investment firms and competent authorities, p. 17.

[405] For example, momentum ignition is discussed as an HFT Strategy (together with quote stuffing and layering) in Tse et al., 2012, AES Analysis. High Frequency Trading – The Good, The Bad, The Regulation. In: Credit Suisse – Trading Strategy. https://edge.credit-suisse.com/edge/Public/Bulletin/Servefile.aspx?FileID=23284&m=1815212669, accessed on 15 November 2017.

Faster speeds in trading could result in greater volatility[406] and a less stable (or potentially rapidly changing) liquidity situations[407]. As a result of greater market speed, market participants face growing pressure to exploit the smallest and most fleeting market movements with the highest effort possible. As a consequence, even a small change can quickly have a major impact.

6. Supervisory and regulatory issues

The aspects depicted in this use case are of supervisory and regulatory significance. Some aspects of algorithmic trading are already currently subject to regulation, for instance access to markets and high-frequency trading.[408,409] In future, supervisory and regulatory authorities will face the question of how to handle increasingly powerful algorithms.

To ensure market integrity overall, one technical solution would be to further expand the safety mechanisms (such as fail-safes, circuit breakers and speed bumps), which are already successfully employed[410] on or required of platforms or providers.[411,412] One example of this would be potential interruptions to the trading process in the event of severe volatility,[413] acting as a circuit breaker when the execution price moves too far outside of a set price corridor by halting ongoing trade and switching to an auction trade model. These measures also include upstream processes such as plausibility checks upon order input (e.g. used by Deutsche Börse).

---

[406] Cf. for example Farmer et al., 2011, An ecological perspective on the future of computer trading, In: The Future of Computer Trading in Financial Markets – Foresight Driver Review – DR6, Government Office for Science, p. 9.

[407] Cf. Tse et al., 2012, AES Analysis. High Frequency Trading – Measurement, Detection and Response. In: Credit Suisse – Trading Strategy. Available online: https://edge.credit-suisse.com/edge/Public/Bulletin/Servefile.aspx?FileID=23285&m=923226224, accessed on 13 November 2017.

[408] In 2012, the European Securities and Markets Authority published its Guidelines for Systems and Controls (European Securities and Markets Authority, 2012, Guidelines. Systems and controls in an automated trading environment for trading platforms, investment firms and competent authorities).

[409] In Germany, for instance, in the High-Frequency Trading Act (specifically in section 33(1a) of the German Securities Trading Act (Wertpapierhandelsgesetz, or WpHG) and according to the Second Financial Markets Amendment Act (Zweites Finanzmarktnovellierungsgesetz, or FiMaNoG), section 69(2) WpHG), in the EU by MiFID II (e.g. Art. 17, 18 and 48) and in the US in SEC Rule 15c3–5.

[410] Cf. Linton, 2011, What has happened to UK equity market quality in the last decade? An analysis of the daily data. In: The Future of Computer Trading in Financial Markets – Foresight Driver Review – DR1, Government Office for Science, for example p. 27.

[411] E.g. in Germany already with section 24(2a) of the German Stock Exchange Act (Börsengesetz, or BörsG) or also by MiFID II (Art. 48 (5) in conjunction with Art. 18 (5)). Some safety mechanisms such as these have also been employed on US stock exchanges for quite some time, for example (cf., for example, International Organization of Securities Commissions, 1992, Coordination between Cash and Derivatives Markets. Report of the Technical Committee on Contract Design of Derivatives Products on Stock Indices and Measures to Minimize Market Disruption; cf. International Organization of Securities Commissions, 2011, Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency. Consultation Report CR02/11 of the Technical Committee of the International Organization of Securities Commissions.

[412] Measures such as these are not to be considered a one-size-fits-all approach – they must be able to meet specific national or regional requirements and asset class requirements (this is already evident in the variation of current measures).

Expansion of these safety mechanisms could also mean using BDAI for calibration (the European Securities and Markets Authority focuses here on a "statistically supported methodology", cf. European Securities and Markets Authority, 2017, Guidelines.) Calibration of circuit breakers and publication of trading halts under MiFID II, p. 7. Any expansion could also include the use of BDAI in designing and placement of these measures.

The question could generally be raised as to whether such measures could in turn be applied to other areas – those not directly related to trading – in a highly interconnected financial market.

[413] E.g. established since the beginning of Deutsche Börse's XETRA trading.

Fail-safes can be provider-specific – the provider stops the algorithms when certain trading parameters (such as volume or number of tickets) are exceeded[414] – as well as systemic (outside of the German context and only of limited transferability, but worth mentioning here by means of illustration: self-help between trading venues[415]).

Speed bumps are yet another such safety technology, i.e. an intentional delay of trading or quotation information.[416] Speed bumps can block certain algorithmic approaches and are already being used by some trading venues. Furthermore, some market participants in algorithmic trading would favour trading venues that employ such a structure.[417]

A different direction is being taken by those considering the creation of synthetic test and simulation environments where both algorithms as well as policy measures[418] could be tested.[419] For example, MiFID II is introducing testing environments for trading venues on regulated markets, multilateral trading facilities (MTF) and organised trading facilities (OTF)[420] where the trading participants are required to test their algorithms[421]. Supervisory authorities could also use test environments both for issues transcending single venues and for policy issues.

The following example illustrates this approach. In preparation for the decimalisation of the NASDAQ (i.e. conversion of trading from fractions to hundredths), a study was conducted to examine the effects of this conversion using a simulation. It became evident, among other things, that changing price granularity also affected possible trading strategies.[422]

---

[414] MiFID II includes numerous regulations governing or affecting algorithmic trading.

[415] In connection with US trading venues, "self-help" refers to an alert which is issued when a trading venue notices a glitch (e.g. a delay) with another trading venue. The trading venue that has been identified is then bypassed by the order routing system (by unilateral action of the identifying trading venue).

[416] The idea of speed bumps as a temporary delay in trading has been around for some time; cf. International Organization of Securities Commissions, 1992, Coordination between Cash and Derivatives Markets. Report of the Technical Committee on Contract Design of Derivatives Products on Stock Indices and Measures to Minimize Market Disruption, p. 8. One of its current interpretations as implementing a delay across the board is newer (cf. Levine, 2016, Speed Bumps Are the Hot New Thing for Exchanges. In: Bloomberg. Available online: https://www.bloomberg.com/view/articles/2016-08-31/speed-bumps-are-the-hot-new-thing-for-exchanges, accessed on 9 January 2018).

[417] Cf. for instance risk.net, 2017, Currencies flow market-maker of the year: XTX Markets. Available online: https://www.risk.net/risk-magazine/analysis/2479841/currencies-flow-market-maker-of-the-year-xtx-markets, accessed on 10 December 2017.

[418] Cf. Farmer et al., 2011, An ecological perspective on the future of computer trading. In: The Future of Computer Trading in Financial Markets – Foresight Driver Review – DR6, Government Office for Science, p. 21; cf. also for an example Baptista et al., 2016, Macroprudential policy in an agent-based model of the UK housing market. In: Bank of England, Staff Working Paper No. 619.

[419] Also to prevent certain measures from changing markets in unwelcome ways, e.g. trade shifting from on exchange to OTC.

[420] Art. 48 (6) MiFID II in conjunction with Art. 18 (5) MiFID II, and in Germany in section 26d(2) BörsG .

[421] Testing is subject to more specific guidelines (Art. 5–7 del. Regulation (EU) 2017/589 (MiFID II level II)). In addition to conformity checks (Art. 6), a securities firm is required, for example, to use clear methodologies for developing and testing algorithms (Art. 5).

[422] Cf. Darley et al. 2007, A NASDAQ market simulation: Insights on a major market from the science of complex adaptive systems. New Jersey et al.: World Scientific Publishing.

Discovering and describing issues such as market manipulation in very large volumes of data is a challenge that could require that supervisory authorities also utilise BDAI. Supervisory authorities use trade monitoring reports, for instance, to prepare internal forecasts, analyses or to identify patterns.[423] Increased use of statistical or BDAI methods could serve to enhance the ability of the supervisory authorities to detect anomalies,[424] but further information is needed in addition to the algorithm to better assess the aims of market participants (e.g. coincidental side effects or intent).

Chapter VI revisits the supervisory and regulatory implications of issues mentioned.

5.4.4.2 Integrated analysis and portfolio management platforms

1. Introduction

The use of analysis (tools) in institutional portfolio management has a long history: Graham and Dodd are an early example of quant investing.[425] Tools and platforms[426] are very widely used today.

This use case shows how BDAI is used in portfolio management and administration at an institutional level, i.e. in a core process of asset management. Some simpler asset allocation models used in a retail setting are not considered here.[427]

Traditional risk models long in use for equities (like factor models[428] for instance) had – and still have – a high data depth. Still used today, these models map equity performance on underlying factors such as value, growth and momentum. However, because increasing volumes of and better data are available, some of the current models, approaches and developments capture more extensive assets and data pools. This has led to enormous growth in the use of models (and associated technologies) – originating from equity portfolio management.[429] Now, in addition to liquid assets such as equities, the underlying models also cover more complex and partially less liquid assets such as mortgage-backed securities.

In equity portfolio management, the first commercially available factor models became available in the 1970s[430], forming the basis for some of the earlier platforms. Generally, analysis platforms are not a new phenomenon, but integration of the different functions and tasks of portfolio management and administration has increased significantly in commercially available systems.

---

[423] For instance, to estimate the anticipated number of reports, to quickly detect changes in the market structure. (The possibility, generally, for supervisory authorities to employ predictions of relevant events/reports is mentioned in Hunt, 2017, From Maps to Apps: the Power of Machine Learning and Artificial Intelligence for Regulators. Speech, Beesley Lecture Series on regulatory economics, 19 October 2017).

[424] Conceptually not unlike using statistical analysis for tax audits.

[425] Graham & Dodd, 1934, Security Analysis, York, PA: Whittlesey House.

[426] In this use case, we use "platform" in reference to systems used by asset managers to analyse, optimise or manage and administrate portfolios. From the provider side, these systems can be platforms in a strict sense (e.g. hosted services) as well as systems installed directly at the asset manager.

[427] These models are also data based, though often with less data depth.

[428] Based on a factor analysis (in the broader sense) that attempts to explain high-dimensional data (such as the return on stock in a market) using a small number of variables (the factors), models can be produced regarding the risk and/or return characteristics (factor models). One example as a "prototype" is the Fama–French three-factor model for stock returns. Cf., e.g., Fama et al., 1992, The Cross-Section of Expected Stock Returns. In: Journal of Finance. Vol. XLVII, no. 2, June 1992, pp. 427–465).

[429] Cf. Financial Times, 2018, The irreversible rise of the investing machines. In: FT View. Available online: https://www.ft.com/content/e1def550-f600-11e7-8715-e94187b3017e, accessed on 29 January 2018.

[430] Cf. Bender et al., 2013, Foundations of Factor Investing, MSCI Research Insight, p. 2.

2. Application

Some platforms cover the entire work flow – from attribution, analysis, optimisation, trading, risk and flow management[431] all the way through to compliance and settlement functions – in highly integrated systems (depending on the process, also across portfolios). The scope of functions depends on the respective asset manager. Some asset managers, for instance, do not have their own trading desks but rather only prepare trade orders that they then passed on to banks, brokers, etc. Some platforms allow complete end-to-end use; certain operational functions and tasks are then largely shifted to the providers.

Example: portfolio optimisation

Portfolio optimisation represents a typical example of the relationship between these different functionalities. As an example, a portfolio composition maximising returns for a given risk would be calculated (mean variance optimisation, dating back to H. Markowitz[432], though current models and techniques can offer certain refinements, for instance by resampling[433]). This optimisation involves several models of several parties, since models are used both for risk and return. In this example, the platforms tend to provide the risk models only, while the return models would remain in the domain of the asset managers and be created by them.

At the same time, greater integration within the platform allows better feedback loops and workflows between the stages of analysis, optimisation, trading and trade processing, and also including risk and flow management. For instance, current market impact[434] data could flow directly into the optimisation process, allowing transaction costs to be better reflected in changes to the weighting within the portfolio. There are numerous commercial platforms and tools with different levels of integration and analysis and model depths available in the market, whereby different providers have different emphases in their products, e.g. operational or analytical aspects.

3. Technology and data

Analytically, many methods from the BDAI spectrum are represented, but deep learning seems to be under-represented in traditional asset management. [435] By virtue of their age, numerous models originate in time series analysis and traditional statistics; methods such as clustering, however, are now considered to be part of unsupervised learning, so that in this sense certain elements of machine learning have been present for a long time.

The data used in the platforms is based on both market data and additional data such as classifications and key corporate data.[436] Examples of these include industry sectors and debt ratios. Furthermore, various asset classes also use data from other asset classes. A typical example of this is the use of equity data in valuing corporate bonds. Put simply, equity can be considered to represent a call option on the corporate value whereas bond investors would hold a short put position. Equity volatility can and is used in this way for approximating the volatility of the company's value.

---

[431] The management of inflows and outflows.

[432] Markowitz, 1952, Portfolio Selection. In: The Journal of Finance, Volume 7 (1), pp. 77–91.

[433] Simply put, the repetition of optimisation steps using statistical variations of input parameters.

[434] For a discussion of the market impact, see also the capital markets use case "Algorithmic Trading for Order Execution by Institutional Investors".

[435] Cf. Marriage, 2017, Fund managers deny AI threatens jobs. In: Financial Times. Available online: https://www.ft.com/content/bd26af40-7dd9-11e7-ab01-a13271d1ee9c, accessed on 29 January 2018.

[436] Cf. for example BARRA: United States Equity. Version 3 (E3), Risk Model Handbook, 1998, p. 73.

There has been a trend in recent years to consider more exotic data – such as aerial or satellite imagery[437], i.e. unstructured data. For oil storage tanks of a certain design, for instance, the fill level can be estimated by measuring the shadow the tank wall casts on the tank's floating roof (this is already being used and offered commercially). Another experimental application is the analysis of face and facial expression of central bank governors with a view to predicting policy changes taken by the central banks.[438] In both cases, BDAI is used to analyse unstructured data, such as satellite imagery and pictures of people. This enables a significant broadening of the data spectrum.

Not all of these data and models are already integrated into existing platforms, but the platform approach makes it easier to provide new data or models to many users at the same time as it requires only the creation of a single interface. At the same time, BDAI makes it possible to analyse new data extensively and quickly.

4. Opportunities

Because these platforms are used by institutional asset managers, consumers will benefit only indirectly – for example from better portfolio management – for instance as a result of better risk management, lower operational risks or more effective use of opportunities for additional excess returns.

Platforms could lower market barriers and risks for smaller players

Through such platforms, smaller participants could gain access to powerful analysis and work flow tools that they would be unable to build for themselves. In this way, platforms could help to create a level playing field for all participants by supporting the smaller ones in their analytical and process-related capabilities. In effect, the platforms could – where applicable – improve large sections of the value chain by providing more comprehensive tools for the work steps and making it easier to allow feedback between these work steps (for example, analysis, optimisation, risk management, trading, processing), reduce operational risks and, as a result, help smaller players' position in the market. The option of end-to-end solutions also supports this, because a participant on its own would potentially only need to provide a few, select components of the value chain.

Reducing the number of interfaces – for example between portfolio optimisation, order preparation, trading connections and risk management – results in fewer operational risks because it reduces the number of times information is transferred between systems and standardises the flow of information. At the same time, the processes as a whole could be tracked using log data so that, for instance, analysis or optimisation results can be verified afterwards.

In principle, the more widespread use of analysis tools also leads to the expansion of systematic risk management in the market.

---

[437] Cf. Finley, 2015, How AI can calculate oil surplus … from space. In: Wired. Available online: https://www.wired.com/2015/03/orbital-insight/, accessed on 29 January 2018.

[438] Cf. Uetake, 2017, In Kuroda's face – researchers find ways to predict central bank changes, In: Reuters Business News (online). Available online: https://www.reuters.com/article/us-japan-boj-face/in-kurodas-face-researchers-find-ways-to-predict-central-bank-changes-idUSKBN1CP0GS, accessed on 11 December 2017.

5. Risks

Consumers are also indirectly affected by risk. Any errors or misuse of the systems or data could affect the prices of asset management products.

When using less transparent or non-transparent models, there is a danger that errors are discovered at a late stage or models are inaccurately assessed – also, such errors and their impact could spread faster in highly integrated systems. For in-house models, however, such errors could also remain undiscovered for long periods of time.[439] When using end-to-end platform solutions, it might also be difficult to discontinue their use and shift to a new provider or to bring the processes back in house; at the same time, there could potentially be a strong operational dependency on platform providers.

In contrast to the level playing field described above, major differences in access to such platforms could also exacerbate the advantages or disadvantages of individual participants in the market.

Possible monocultures in risk assessment

In extreme cases, market players using the same risk models or platforms could produce monocultures with potential systemic risk implications. One example of this is would be the market-wide use of copula[440] models for CDOs, the weaknesses of which became evident during the financial crisis[441]. On the other hand, commercial factor models have been in use for a long time without causing obvious market dislocations (portfolio managers use models in very different ways and advertise this accordingly).[442]

A wider data spectrum also means more room for data error or manipulation. (This could potentially be compensated for by a larger data spectrum being used to create greater differentiation between the models in portfolio management and in risk management.)

---

[439] Cf. e.g. Securities and Exchange Commission, 2011, SEC Charges AXA Rosenberg Entities for Concealing Error in Quantitative Investment Model. Washington, D.C., 3 February 2011. Available online: https://www.sec.gov/news/press/2011/2011-37.htm, accessed on 29 January 2018.

[440] A copula function can be used to determine the correlation of different individual probability distributions.

[441] Cf. for example Brigo et al., 2010, Credit Models and the Crisis, or: How I learned to stop worrying and love the CDOs. Available online: http://arxiv.org/abs/0912.5427, accessed on 30 January 2018

[442] The next analogue could be a discussion of the anti-competitive effects of passive investments. Here the discussion is about the effects of passive investments on the market and corporate strategy; cf. for example Levine, 2015, Index Funds May Work a Little Too Well. Bloomberg (online). Available online: https://www.bloomberg.com/view/articles/2015-07-22/index-funds-may-work-a-little-too-well, accessed on 29 January 2018.

6. Supervisory and regulatory issues

This use case highlights various supervisory and regulatory issues. For example, this section raises several questions that might in principle – i.e. beyond the bounds of asset management – also occur in connection with other (integrated) analysis and management systems.

Systemic risks in model and data use?

One example is the collection and detection of systemic risks that could potentially arise from market-wide use of a single or only a few models or tools[443], particularly the necessity of techniques for assessing these risks, e.g. by asking which data needs to be collected to understand the landscape of the platforms or the model and data usage within this landscape. In order to gain an understanding of the effects of large-model cultures in the market, the market simulations presented in the use case "Algorithmic trading for order execution by institutional investors" could also play a role.

Critical transparency in the analysis of securities?

The use of certain data sources capable of bringing companies' "inside" data to the outside and the use of certain internal data held by the asset managers touch on the issue of the boundaries between public and private information.[444] For example, an aerial photo showing a parking lot not visible from the outside could make it possible to draw conclusions about the company's business position by analysing the vehicles.

New relevant players?

Another potential question would be how supervisory and regulatory authorities should view analytics providers (or non-transparent tools) that are highly integrated into the business. One question could be whether or not the loss of a provider could be absorbed or compensated. For end-to-end platforms in particular, dependencies on only a few providers could arise which might be difficult to untangle, dependencies representing operational meta-risks if many users or large asset volumes were to be managed through only a few such platforms.

Chapter VI addresses the above-mentioned issues in greater detail.

---

[443] Not unlike the risks of monoculture such as those in agriculture, in other words the danger that one error or problem could destabilise the entire system.

[444] Cf. also Thornhill, 2017, The big data scramble still relies on people, In: Financial Times. Available online: https://www.ft.com/content/358fab32-0317-11e7-ace0-1ce02ef0def9, accessed on 29 January 2018.

5.4.4.3 Crowdsourcing of investment algorithms

1. Introduction

Crowdsourcing of investment algorithms – i.e. using the internet to outsource the development of these algorithms to developers outside of the investment firms – could offer an opportunity, based on the use of BDAI, to harmonise two conflicting interests: increasing returns or performance while lowering operating costs.

Crowdsourcing investment algorithms offer a way to use recent developments in BDAI for the creation of new business models in asset management, particularly in the hedge fund segment[445], by isolating what has until now been a traditional core competence.

Conventional development of investment algorithms

In investment firms, investment algorithms are usually developed by full-time specialists[446] on the basis of proprietary or expensive data, such as market data. Both the data set used to develop the algorithms as well as the algorithms themselves are essential to the foundation of the business for these market participants and are therefore protected accordingly.

From the perspective of some market participants, however, being limited to using full-time specialists is more of a disadvantage because it means that only (relatively) few algorithms can be generated for investment purposes. Crowdsourcing algorithms could be a solution to this problem.

Innovation: development of investment algorithms via crowdsourcing

The idea behind this is that as the number of available algorithms grows so too does the probability of finding successful algorithms or expanding portfolio effects to algorithms.

At present, crowdsourcing investment algorithms are mainly used by hedge funds, primarily in liquid asset classes such as equities and exchange-traded derivatives.

Two innovations can be observed here:

- Analytics: new methods and applications make it possible to work with encrypted data and algorithms. Utilising new encryption techniques, users can develop algorithms in encrypted data, and hedge funds, for instance, can test encrypted algorithms to determine their profitability and risk profiles.

- Infrastructure and hardware: the current data infrastructure and hardware make it possible to transport large volumes of data quickly and securely. Furthermore it can handle the analysis, backtesting and observation of a large universe of algorithms and the resulting portfolios of algorithms. In addition, the speed of today's processors makes it possible to utilise encrypted data and algorithms in ongoing applications and for development purposes.

Hedge funds that use algorithms provided by crowd-developers currently represent only a very small portion of the market. And yet because hedge funds act extremely opportunistically[447] it is conceivable that – if all technical barriers can be successfully overcome and appropriate results could be achieved – more hedge funds with such products would appear in the market.

---

[445] The crowdsourcing model is also used by platforms that connect investors and algorithm developers or provide investment signals. That means that such platforms collect algorithms and show their performance.

[446] These specialists are sometimes referred to as "quants", although this description is not always unique.

[447] New, successful business models are therefore quickly copied by other market participants.

Figure 26 (stylised): Breakup of the traditional (alpha[448]) process chain in investment management
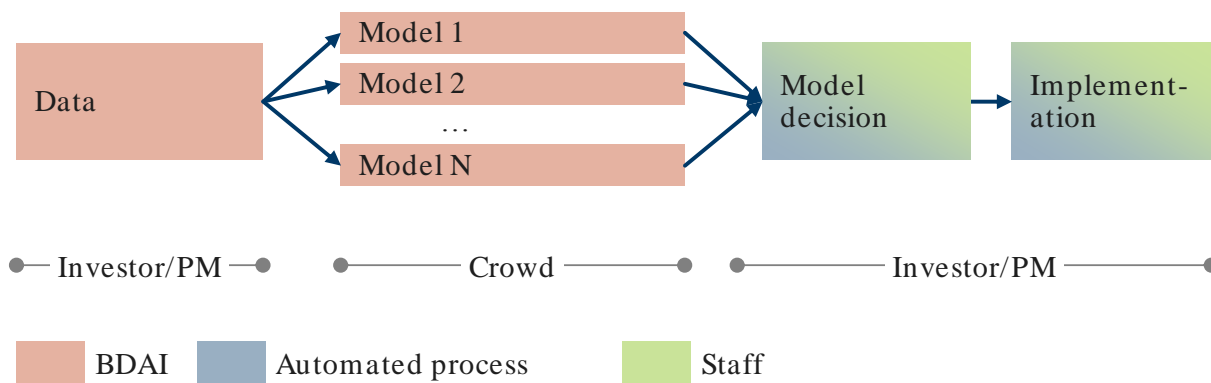
2. Application

Using various quantitative criteria, investment algorithms – unlike, for instance, algorithms for executing trading orders (cf. use case entitled "Algorithmic Trading for Order Execution by Institutional Investors") – make automated investment decisions or, if applicable, recommendations.[449] One example of this would be algorithms that identify arbitrage opportunities, exploiting price differences between different markets.

Specifically, crowdsourcing of investment algorithms is realised as follows: hedge funds or platforms provide developers with (sometimes encrypted) data. The developers design algorithms for investment decisions based on this data which are then made available to hedge funds or platforms. The hedge funds/platforms evaluate the algorithms automatically[450] to thus identify the most promising ones. Hedge funds or investors in turn use the identified algorithms to make investments, and, if successful, the developers share in the profits or are remunerated accordingly (for instance with a cryptocurrency or crypto tokens).

Investment algorithms are already crowdsourced by some specialist start-up companies and new participants (mainly in the US for investable forms), for instance by hedge funds, investment firms, platforms and specialist providers.

---

[448] Alpha refers to a portfolio's excess return compared to the market.

[449] I.e. the execution itself would not be done automatically. In addition, decision frequency also varies widely (from very short periods up to days and months).

[450] This is also an area in which BDAI or algorithms are used to assess the algorithms' probability of success. False positives, in other words the erroneous assumption that an algorithm is successful, are a relevant statistical problem given the large number of algorithms to be tested, cf. Craib et al., 2017, Numeraire: A Cryptographic Token for Coordinating Machine Intelligence and Preventing Overfitting. Available online: https://numer.ai/whitepaper.pdf, accessed on 20 February 2017.

3. Technology and data

Whether or not this approach catches on depends not only on investment performance but also on technological feasibility. In addition to mere provision of the infrastructure, some potentially relevant aspects in the design of such schemes can be observed:

- *Encryption*: for a large number of external developers to be in a position to develop algorithms, hedge funds must provide them with the corresponding data. As this data is in some cases proprietary or sharing[451] would require additional and costly licences, encryption might offer one option for handling the intellectual property rights.[452] From a technological and conceptual point of view, this type of encryption is challenging because developers need to be provided with the information they need to develop the algorithms, but this information cannot be identical to nor should it be possible to trace it back to the unencrypted information. In addition, the encryption technology must allow for the fast encryption and decryption of large volumes of data. Currently, this problem is partially solved. Encryption can – possibly in conjunction with the anonymity of the developer (see below) – create a high degree of structural opacity. At present, the data used consists mainly of structured data such as equity and futures prices or volumes.

- *Contractual security*: in order to give as many developers as possible the opportunity to participate, developers should have the option of remaining anonymous. The algorithms themselves are sometimes encrypted in order to ensure that the developers retain the rights and thereby also the basis for sharing in the profits.[453] This ensures a separation between ownership of the data and ownership of the algorithm. In addition, the developers have to have sufficient certainty that, if their algorithm is successful, they will actually share in the profit. To ensure that this occurs, some participants are using DAO and blockchain technologies[454] coupled with cryptocurrencies or crypto tokens. The separation between the authority over the data and the algorithm (possibly in conjunction with anonymity) observed here demonstrates that there could also be technological solutions to complex problems in the protection of data or intangible assets.

---

[451] For example, if developers should be able to work independently with the data.

[452] Cf. Craib, 2018, Encrypted Data for Efficient Markets, In: Medium.com. Available online: https://medium.com/numerai/encrypted-data-for-efficient-markets-fffbe9743ba8, accessed on 29 January 2018.

[453] An example of this is given in Quantiacs, 2017. Available online: https://quantiacs.com/Tutorials/MatlabSubmitting.aspx, accessed on 29 January 2018.

[454] A DAO (decentralised autonomous organisation) is a combination of smart contracts (i.e. programs that act as contracts) and blockchain, used – broadly – as a ledger. For example, if the algorithm is successful, the developers are directly remunerated in a cryptocurrency through the smart contracts. For an example of this cf. Craib, 2017, An AI Hedge Fund Goes Live On Ethereum, In: Medium.com. Available online: https://medium.com/numerai/an-ai-hedge-fund-goes-live-on-ethereum-a80470c6b681, accessed on 29 January 2018.

4. Opportunities

By expanding the base of developers of investment algorithms (private persons, in particular), more people will gain access to another aspect of the capital market. Retail customers could benefit from potentially better performance and lower costs of funds and asset managers, should such models be successful and be accessible to retail customers.

Broader base of algorithms

It might become possible for investment firms to make more successful investment decisions because there is a significantly larger pool of algorithms to choose from (one of the providers mentions over 190,000 machine learning algorithms[455]). A larger inflow of algorithms could also mean that investment approaches would be able to adjust quickly to new market phases or environments and that, in this way, performance would become more robust.

Stronger diversification of the market

There are also opportunities from the perspective of the overall market because the greater range of algorithms could lead to better diversification, producing a more efficient and stable market. Moreover, establishing new business models could provide for a fund market that is more diversified and thus more robust.

5. Risks

Opacity for customers

The opacity – which characterises some these hedge funds and platform models – poses a risk to the consumer. For example, models similar to pump-and-dump[456] schemes could make a comeback in automated form (particularly among platform models where the operator or the algorithm developer could maintain that the other is responsible). Because the investable models (hedge funds) presented are currently not available to retail investors, the risks are limited to a small group of investors.

For the investors and investment firms that use crowdsourcing for investment algorithms, challenges in identifying errors could lead to losses. In turn, controlling the algorithm universe could require higher costs and greater effort, which would offset the model's advantage.

---

[455] Cf. Craib, 2016, Super Intelligence for the Stock Market. In: Medium.com. Available online: https://medium.com/numerai/invisible-super-intelligence-for-the-stock-market-3c64b57b244c, accessed on 29 January 2018.

[456] The misleading inflation ("pumping up") of stock prices at low volumes and price levels in order to subsequently sell the stock to investors. In this context, this could mean for example: an actively implemented algorithm is "successful" by buying stock with low liquidity; this greatly increases the price which then in turn improves the performance of the algorithm. When more investors step in, the algorithm developers sell more stocks at a high profit (before the price collapses when demand slows).

Possible obstacles for (macro) economic risk evaluation

The crowdsourcing approach is generally agnostic and purely results-oriented. In other words, the hedge funds, platforms and investors providing the entire infrastructure might in some circumstances only review the vast number of generated algorithms for profitability – or rather evaluate the performance and risk characteristics – and then invest based on the most promising algorithms. Investors (or portfolio managers) would then make their investment decisions on the basis of statistical analyses[457] and simulations/market tests of the algorithms.[458] Risks could therefore no longer be assessed on a directly (macro) economic basis. For the same reason, errors or vulnerabilities of these algorithms – both prior to the investment decision as well as problems arising during the investment – could be difficult for the portfolio managers to remedy individually, yet could be identified and handled as part of portfolio management, for example.

Market effects could also occur when similar algorithms with similar mechanisms (and similar vulnerabilities) are used by a greater number of market participants. This would happen if a large number of market participants were to use the same crowd. The effect would be exacerbated if erroneous algorithms had been intentionally provided to the investors by that crowd.

6. Supervisory and regulatory issues

The aspects depicted in this use case highlight various supervisory and regulatory implications. First of all, the underlying approach of presenting numerous algorithms is already familiar from platforms using signal providers (or algorithms). However, it occurs here as described above in a more technologically sophisticated and expanded form. Further complications could arise as a result of the use of certain cryptographic methods. Potentially, developers of signals or algorithms might not know that they were developing financial signals, and it would therefore be more difficult for the regulatory authority to discover[459] if they work on the assumption that developers intentionally create financial signals. This could occur, for instance, if the signal provider itself could be regulated, which could be the case in connection with platforms or commercial signal providers.[460]

Maintaining an overview despite a lack of transparency

For supervisory and regulatory authorities, the question remains as to how to handle a proliferation of investment firms acting agnostically.[461] The discussion follows the use case "Algorithmic trading for order execution by institutional investors" – here, too, the potential interdependence between different algorithms and market participants represents a challenge.

As a rule, encryption and/or anonymisation generate transparency problems for the supervisory authorities and may impede effective and appropriate regulation, although they are a welcome development from a data protection standpoint. In particular, the emerging possibilities (e.g. using homomorphic encryption) to separate the authority over the data from the authority over the algorithms could find wider use in data protection or in maintaining data sovereignty in complex data exchanges between different parties (cf. Chapter 3.5.2)).

---

[457] This could potentially also include information on the genesis of the algorithms.

[458] This is not a new development in the investment area per se.

[459] Regardless of any additional anonymity.

[460] Cf., for example, the BaFin explanation of Signal Following und Social Trading. Available online: https://www.bafin.de/EN/Aufsicht/FinTech/Signalgebung/signalgebung_artikel_en.html, accessed on 9 January 2018.

[461] Not unlike certain discussions relating to passive investments.

Legal evaluation of contractual relationships on market entry in Germany

As hedge funds in the USA are currently using crowdsourcing for investment algorithms, their appearance in Germany or Europe would call for close observation, for instance, of how the legal relationship between the asset manager and the algorithm developers is structured (there is currently some mention of a licence for the algorithm[462], in another model the developer might even remain anonymous[463]). As a consequence, consideration should be given as to whether these relationships could be considered outsourcing and to what extent these models could meet the requirements for that.

To the extent that purely blockchain assets in the structures referenced here are used as speculative or investment objects (assets such as certain cryptocurrencies or crypto tokens that do not form part of the "real" capital market), it would not be necessary to link the hedge fund to the traditional capital market.[464] These funds – in contrast to funds that are still active in the "real" capital market – would potentially be unrecognisable from the outside. How could such structures be detected and classified legally?

Similarly, in systems with distributed consensus (e.g. blockchain) it could be difficult to assign responsibility as decisions might not be attributable to individual participants.

Mirroring of crowdsourcing by regulatory and supervisory authorities?

In addition, there might at least be a theoretical possibility of the supervisory and regulatory authorities using crowdsourcing themselves. In such a scenario, encrypted[465] data could be provided to users via an online platform (anonymously). Identifying vulnerabilities or anomalies would in turn result in financial compensation for the (anonymous) data scientists. In this way, the business model described above would in a sense be mirrored on the supervisory and regulatory side. That this is being tested in practice can be seen in the example of the Department for Homeland Security in the US, which conducted a public data science competition with the goal of helping the Transportation Security Administration improve the recognition of threats in images.[466]

Chapter VI addresses the supervisory and regulatory implications and key questions in more detail.

---

[462] E.g. Quantopian-Website. Available online: https://www.quantopian.com/allocation, accessed on 19 November 2017.

[463] Cf. Metz, 2016, 7,500 Faceless Coders Paid in Bitcoin Built a Hedge Fund's Brain. In: Wired. Available online: https://www.wired.com/2016/12/7500-faceless-coders-paid-bitcoin-built-hedge-funds-brain/, accessed on 17 January 2018.

[464] For example, these speculative transactions may not have been carried out on normal markets or trading venues but rather on markets potentially not subject to regulation or supervision.

[465] Or other data adequately protected in another way.

[466] Cf. Department of Homeland Security, 2017, Passenger Screening Algorithm Challenge. Available online https://www.challenge.gov/challenge/passenger-screening-algorithm-challenge/, accessed on 29 January 2018.

VI. Supervisory and regulatory implications

6.1 Introduction

Innovative developments require supervisory/regulatory attention at an early stage | The preceding chapters have looked in detail at the possible impact of BDAI technology on the financial services industry. New findings can be generated by combining and analysing the massive amounts of data available. In the financial services sector, value is created from data mainly through product and process innovations. The market analyses and use cases showed that changes and enhancements to products and processes are driven by both existing and new players. As new players enter the market and the dependency on external BDAI providers grows, new market structures could emerge. This is a gradual, dynamic process still in its early stages in the traditional banking and insurance business. But on the capital market, it has already gained greater relevance in practice. For supervisory and regulatory authorities, the task is to understand and accommodate these potentially profound changes. Therefore, both the risks and opportunities associated with BDAI innovations are described in this report.

Implications of product and process innovations and of the emergence of new players and market structures | In order to determine the implications of BDAI from a supervisory perspective, we have examined the main effects of this phenomenon for the three areas described above: product innovations, process innovations, and the resulting emergence of new players and market structures. Each individual BDAI effect is examined to determine whether it requires a supervisory and/or regulatory response. However, national and international supervisory authorities still need to extensively discuss how exactly to respond in each case. This chapter presents the potential supervisory and regulatory implications as strategic key questions that must be jointly addressed at both national and international level.

Holistic approach | Supervisory and regulatory implications are derived for the areas of financial stability (at the macroprudential level) and market supervision, firm supervision (at the microprudential level) and collective consumer protection. This chapter is organised along three main topics. Here, we examine the potential supervisory and regulatory implications for the financial system as a whole – i.e. for banks, insurance companies and the capital markets[467]. However, BDAI developments are also relevant beyond the financial system and raise overarching social and political questions that are already widely debated. Accordingly, the implications for financial supervision and regulation should also be considered within the context of these overarching debates. The key issues affecting the financial system as a whole are briefly addressed in the last section of this chapter.

---

[467] In this study, the term "capital markets" includes – inter alia – investment services enterprises, investment funds, asset management companies and operators of trading venues (stock exchanges).

6.2 Supervisory and regulatory implications

Figure 27: Overview of the implications for supervision and regulation

Goals	Supervisory and regulatory issues
Financial stability and market supervision	1 Emergence of new business models and entities
	2 Connecting markets and market participants
	3 Using technology to limit undesirable developments
	4 Redefining and addressing systemic importance
Firm supervision	1 BDAI governance
	2 Fighting financial crime and preventing conduct violations
	3 Internal models subject to supervisory approval
	4 Handling information security risks
Collective consumer protection	1 Risk of discrimination
	2 Consumer sovereignty

6.2.1 Financial stability and market supervision

Opinion

Big data and artificial intelligence are attributed with the potential for tremendous impact – such as the power to create new business models in the financial market. Supervisors have to investigate the critical issue of whether current supervisory legislation adequately covers these business models, and whether authorisation is required. If, however, these models fall through the cracks but still harbour material risks, then regulatory adjustments may be required.

Dr. Thorsten Pötzsch

Chief Executive Director of
Resolution Directorate
at BaFin

BDAI could also attract new, systemically important providers. It is possible that companies with data-driven business models quickly become more systemically important due to their scalability and reach. And yet systemic importance could also be secured if central data or platform providers make identical or very similar structures for processes or algorithms available for a whole variety of market players. The phenomenon of BDAI puts to the test the old definition of systemic importance and, as a corollary, how supervisory and regulatory authorities deal with it.

The use of BDAI for the detection of money laundering is still in the trial stages. However, it is quite conceivable that it will prove to have far-reaching effects here, too. At present, closer inspection of automatically generated pointers flagging a suspected case of money laundering often proves these cases not worthy of further investigation. BDAI processes could help to raise the quality of the hits and independently identify previously undetected patterns. However, banks and insurance companies must first of all be in a position to work with such processes. This learning curve is steep but BaFin believes that companies will do all they can to succeed.

6.2.1.1 Emergence of new business models and companies

Identifying and addressing regulatory gaps at an early stage | The market analyses showed that new players are emerging as a result of BDAI-based innovations (see Chapter V). This may further drive the disaggregation of the value chain, particularly if existing businesses cooperate with new, specialised providers. And this effect is intensified by technical solutions such as application programming interfaces (APIs). The phenomenon of BDAI gives rise to new types of business models and market participants that are not yet adequately covered by the current regulatory framework. It is vital that these are identified and that the range of firms and providers to be supervised is expanded accordingly.

The following key questions can be derived:

- Which business models that are not adequately covered by the current regulatory framework need to be observed at the moment or are to be expected in the near future?

    o Which analysis methods, e.g. from market research, could help to address such business models at an early stage?

- Should assessments of the medium- to long-term solvency of existing market participants reflect growing competitive and margin pressures to a greater extent?

6.2.1.2 Connecting markets and market participants

BDAI can lead to greater connectivity between markets and market participants | Generally speaking, all of the market analyses and various use cases on banks, insurance companies and the capital markets indicate that markets and market participants will be more connected than ever before. This interconnectedness can arise both indirectly, e.g. if the same models, data or platforms are used, and directly through new contractual and trade relationships made necessary in the first place by BDAI usage. The disaggregation of the value chain could also result in a growing number of new market participants that, in turn, are also connected with each other in a multitude of ways.

Maintaining transparency and monitoring new structural relationships | Greater interconnectedness could result in greater complexity in the market, for instance, if a market participant's formerly internal processes are distributed among several market participants. New risks could also arise at the interfaces to and between market participants. And as these risks are no longer within the organisational structures of supervised firms, these firms may not be able to completely identify or manage these risks. The changing structure of this dynamic market and the resulting risks must therefore be evaluated and addressed from a supervisory and regulatory point of view.

The following key questions can be derived:

- How can we ensure that the structure of this dynamic market and the resulting risks are transparent in the long term?

    o Could supervisory authorities use graph analysis or topological methods, for instance, in the long term in order to identify market structures?

- Could such findings be used to calibrate macroprudential buffers, e.g. by directly taking into account the degree of interconnectedness in the same way as SIFIs[468] are determined?

- What can be done to identify risks that are outside the organisational structure of supervised market participants and that can only be partially identified or managed by both the market participants themselves and the supervisory authorities (e.g. risks resulting from a dependency on external ratings)?

6.2.1.3 Using technology to limit undesirable developments

BDAI-induced connectivity increases the risk of domino effects | Even outside trading venues, closely interconnected systems are susceptible to the rapid and uncontrolled spread of disruptions. This is relevant particularly with respect to interconnectedness, as discussed above (see Chapter 6.2.1.2) because this can amplify domino effects.

Applying technological safeguards from trading to other areas | In the use case entitled "Algorithmic trading to execute orders for institutional investors", we discussed technological safeguards, such as volatility interruptions, which are already widely used in trading venues and algorithmic trading. In terms of design, such safeguards are not limited to trading venues only and can be structurally expanded. This raises the question of whether such safeguards would be necessary and could even be usefully applied within a BDAI context, even outside of trading venues. For example, decoupling mechanisms for data streams could be considered, as data supplies will become a matter of greater importance as a result of BDAI.

The following key questions can be derived:

- Would safeguards for limiting domino effects in the context of BDAI be necessary and useful even outside of trading venues?

- How could existing protective measures, such as volatility interruptions, speed bumps and circuit breakers, be optimised or improved specifically for this purpose? How could innovative protective measures be advanced, for instance, through experiments conducted in test environments?

---

[468] SIFI stands for "systemically important financial institution".

6.2.1.4 Redefining and addressing systemic importance

BDAI enables the emergence of new systemically important players | The market analyses (see Chapter V) have shown that BDAI could foster the emergence of systemically important players. For example, providers with data-driven business models could quickly become more systemically important due to their scalability and reach. However, systemic importance could also arise if key data providers or platform providers offer numerous market participants identical or very similar infrastructures or databases for processes or algorithms. Systemic importance could also emerge structurally from the interactions between different market participants.

Redefining and addressing systemic importance | Traditional supervisory methods for determining systemic importance might be unsuitable or would have to be supplemented in this case. In addition, it is necessary to examine whether current risk mitigation strategies would also be effective with new systemically important companies and structures. This raises the question of whether and how the banking- and insurance-based concept of systemic importance needs to be redefined in order to keep pace with new business models and market structures.

The following key questions can be derived:

■ Does the concept of systemic importance need to be redefined as market structures are changing? If so, how can this be done?

      o   Could the methods used to identify structures, such as topological methods as mentioned in the previous chapter, be used here as well?

      o   Do current risk mitigation strategies need to be readjusted in order to take into account potentially new systemically important companies and structures?

6.2.2 Firm supervision[469]

Opinion

A strong trend towards process modernisation is emerging in the insurance industry. Those technologies offering the greatest potential include big data and artificial intelligence, as this report clearly shows.

Dr. Frank Grund

Chief Executive Director of Insurance and Pension Funds Supervision at BaFin

Artificial intelligence is often mentioned in connection with dark processing. However, in many cases, it appears to be more an elaboration of rule-based algorithms that classify claims – well in some cases, not so well in others – and regulate or prepare decisions accordingly. The challenge for the industry will be to develop real intelligence in the sense of independent cognitive power, i.e. the solution to what has, to date, been an unknown and complex problem in decision-making.

At present, big data is primarily of importance in risk-based differentiation of insurance premiums. This is not a new phenomenon; it is simply that advances in digitalization have attracted greater public awareness. Think, for instance, of the variety of tariff features in motor insurance or occupational disability insurance.

Public interest also draws attention to the limitations of big data. Ever greater price differentiation is sometimes considered socially undesirable. In addition, there are also concerns regarding data protection and moral aspects.

However, over time, these and other technologies will impact increasing numbers of core processes in the insurance sector. This report gives several examples. Policyholders should also reap benefits from this development because communication with the insurance company becomes easier, core processes are speeded up and cost savings are secured in the medium term.

Testing the new technologies should reveal whether digitalisation has the potential to fundamentally change the business model. BaFin's insurance supervision has recognised the growing importance of BDAI and is actively seeking to develop the necessary know-how.

---

[469] Here, the term "firm supervision" generally refers to firms subject to ongoing supervision, i.e. banks, financial service providers, payment institutions, insurance companies and asset management companies. The implications for each supervisory regime are intrinsically different.

Opinion

Financial transaction data is a valuable resource in the world of big data and artificial intelligence. It is not surprising, therefore, that it attracts the interest of companies outside the financial sector. In the world of finance, however, trust is an equally important resource, particularly with regard to how companies handle personal data. It remains to be seen which players will be able to gain and retain that trust.
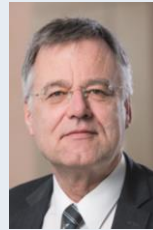
Raimund Röseler

Chief Executive Director of
Banking Supervision
at BaFin

Established banks possess both of these resources, trust and financial data. The ideal position to be in, one might think. And yet, they may still be left behind. Before the traditional world of banking can reap the benefits of big data and artificial intelligence, some institutions will have to replace their legacy IT systems. Time is of the essence and competition is fierce. Bigtechs could edge their way into the financial sector. Not only do they boast huge volumes of data, but they also have the modern technology needed to exploit it.

The only way for banks to survive in this fully digitised environment is to invest heavily in technology – and in people. Banks must realise how crucial this is, that this is all about equipping the traditional business model for the future. The culture has to change – yet without sacrificing consumer trust.

It must also be crystal clear that artificial intelligence is and will only ever be a tool. Senior management will always be responsible for all the decisions made. You can outsource a lot of things, but not responsibility! BDAI will not change that.

6.2.2.1 BDAI governance

Senior management remains responsible – embedding BDAI within a proper business organisation |
BDAI will create additional opportunities for automating standard market processes – from chatbots to
liquidity management. One example is the progression of automation in the lending business at banks or
claims management at insurance companies (see Chapters 5.2.1 and 5.3.1). Further automation is aimed at
increasing efficiency and effectiveness – and thus decreasing costs. However, this does not mean that
responsibility for the results of BDAI-supported processes is to be shifted to machines. When designing
(partially) automated processes, it is therefore important to ensure that they are embedded in an effective,
appropriate and proper business organisation. Ultimately, responsibility for automated processes is to
remain with the senior management of the supervised firm. Appropriate documentation is required to
ensure this.

**No black box excuses – explainability/traceability of models is necessary and can improve the
analysis process** | It is the responsibility of supervised firms to guarantee the explainability/traceability[470] of
BDAI-based decisions. In particular, chapter 3.5.2 details how new approaches could provide at least some
insight into how models work and the reasons behind decisions, even in the case of highly complex models,
thereby preventing models from being categorised purely as black boxes. Supervisory and regulatory
authorities will therefore not accept any models presented as an unexplainable black box. In addition, a
better understanding of models provides an opportunity to improve the analysis process – allowing, for
instance, the responsible units in the supervised firm to identify overfitting and data bias (see Chapter 3.3.1).

Continuing to develop existing concepts within the context of machine-based governance | It may be
necessary to further develop established governance concepts, such as the principle of dual control, and to
apply these to automated processes.[471] For example, one could think about introducing special safeguards
for certain particularly risky BDAI applications, safeguards already used in other technological applications.
In aviation, for instance, speed is measured using several independent systems – the back-up would be an
algorithm.

Supplementing existing documentation requirements by monitoring results | Due to the complexity of
the applications, also resulting from the type and scope of data used, consideration should be given as to
whether the process results should also be monitored in the future, in addition to the documentation
requirements described above. For example, this could be done by evaluating the results produced by an
algorithm in a test scenario set by the supervisory authorities.

---

[470] Explainability describes the ability to determine the main factors influencing a specific individual decision that has been reached
by a system (see Chapter 3.5.2).
[471] "*Wie gehen wir beispielsweise mit Vorgaben zum Vier-Augen-Prinzip um, wenn bei voll digitalisierten Prozessen nicht einmal
ein menschliches Auge auf einzelne Geschäfte schau*t?", speech by Raimund Röseler, BaFin's Chief Executive Director of Banking
Supervision, at BaFin's press conference on 3 May 2018; available online (only in German):
https://www.bafin.de/SharedDocs/Veroeffentlichungen/DE/Reden/re_180503_jahrespressekonferenz_p.html; accessed on
7 May 2018.

The following key questions can be derived:

- Does the scope of existing supervisory practices and the corresponding legal requirements governing proper business organisation need to be expanded as the use of BDAI increases?

  - Could additional technical safeguards such as those used in aviation be necessary and suitable for the appropriate management of particularly risky BDAI applications as part of a proper business organisation?

  - Which BDAI applications would be considered for such special treatment? Should, for instance, a chatbot and a model for liquidity management be treated differently?

- Is it necessary to extend the scope of existing requirements governing the evaluation of processes for BDAI-driven applications beyond the documentation requirements that currently apply?

  - Does it make sense, for instance, to also focus on examining actual results when complex processes are evaluated, required documentation aside?

- How can a minimum standard be established to govern the explainability/traceability of the algorithms used, potentially with different levels depending on the areas in which they are each used?

- Is it necessary to define additional eligibility requirements for senior management as the use of BDAI increases?

6.2.2.2 Fighting financial crime and preventing conduct violations

Exploiting BDAI potential in compliance processes | BDAI can improve the detection rate of anomalies and patterns, and thus increase the efficiency and effectiveness of compliance processes, such as money laundering detection or fraud prevention. This is explained in detail under "Compliance" (see Chapter 5.2.6.2). In addition, BDAI can also play a role in monitoring employee conduct.

Preventing criminals from turning to less advanced firms | Should the use of BDAI technologies result in far more efficient ways to detect money laundering, criminals could potentially turn to firms that are less advanced in this area. It is therefore necessary to monitor whether this will materialise.

Defining supervisory requirements for the explainability and effectiveness of algorithms for detecting financial crimes | In addition, supervisory and regulatory authorities need to discuss whether it is necessary to define specific minimum standards for the explainability/traceability and effectiveness of the methods used. The results of the algorithms would have to be traceable to the extent that the supervisory authorities can monitor them and that the competent authorities, e.g. law enforcement agencies, can make use of them. In addition to performing comparisons with specific minimum standards that may need to be developed, supervisory authorities could also determine the effectiveness of algorithms by benchmarking them against the algorithms of other providers.

The following key questions can be derived:

- What steps can be taken to prevent undesirable or criminal activity from moving to firms where money-laundering detection processes are less developed in terms of BDAI?

- What explainability and documentation requirements must algorithms meet in order to ensure that their results can be used for official sanctions and intervention measures aimed at preventing or prosecuting crimes?

- Should general standards be defined for the effectiveness of the methods used when applying BDAI technologies, particularly in order to identify money laundering activities?

6.2.2.3 Internal models subject to supervisory approval

Possible uses of BDAI in models requiring supervisory approval | Supervised firms may also wish to use BDAI methods in models that are subject to approval by the supervisory authority, for example, the internal ratings-based approach for banks. This is likely to be the case if the use of BDAI leads to more precise results than the combination of methods and data currently used. In this way, BDAI could potentially be used to improve current methods, for instance, those used to optimise the assessment of individual counterparty default risks. In addition, a better portfolio view could possibly alter the weighting of individual risks as concentration risks could be addressed more appropriately, for instance, by analysing the degree of interconnectedness between borrowers.

Actual effects of BDAI use on capital requirements are likely to be limited | It is worth noting that there are currently no BDAI applications observed or authorised in the algorithms of such models. It should also be noted that the authorised models in existence today are highly advanced, for instance, with regard to the scope of the data used, including external data. In the banking sector, it should also be borne in mind that the output floor approved as part of the finalisation of the Basel III reforms would additionally limit the actual effects on capital requirements.

Defining prerequisites for BDAI use in models requiring supervisory approval | The use of BDAI models would always be subject to a corresponding approval on a case-by-case basis. Beyond the individual case, the question could be asked as to whether all BDAI methods are equally suited for use in models that require supervisory approval, or whether there are methods that should be ruled out per se. Furthermore, it is necessary to examine whether existing legal (minimum) standards for the data used and for BDAI model transparency are sufficient or whether additional requirements would be necessary. The following should also be discussed: What degree of added complexity would be acceptable for an improvement in forecast quality? In the case of dynamic BDAI models, the question arises as to what general modifications would constitute a model change from a supervisory point of view, which banks or insurers (e.g. as part of the model change policy for insurance companies) would then have to report to the supervisory authority and secure approval for where necessary.

The following key questions can be derived:

- How would the use of BDAI change model development, monitoring and approval process requirements, particularly with respect to the increasingly dynamic pace of change for data and algorithms?

- From a supervisory perspective, what general modifications would constitute a model change that would need to be reported by supervised firms and would potentially be subject to approval?

- Does the scope of the existing legal (minimum) requirements governing the explainability of the models and data need to be extended to cover the use of BDAI?

- Are all BDAI methods equally suitable? How can this be determined?

- Could an increased use of data help reduce the algorithmic complexity of models while still improving explainability?

6.2.2.4 Handling information security risks

Information security risks increase as a result of greater use of BDAI | The growing complexity caused by BDAI presents new challenges in managing information security risks. Together with ever larger data volumes, the disaggregation of value chains supported by BDAI creates a larger attack surface while simultaneously reducing each individual provider's ability to control the data used and distributed. In addition, a new phenomenon has arisen: attacks on certain BDAI algorithms through data manipulation, e.g. in the form of adversarial or poisoning attacks. This leads to risks that are very different to those encountered in traditional online banking, resulting in a rise in both operational and reputational risks.

BDAI and encryption systems for mitigating information security risks | Although the use of BDAI can result in an increase in information security risks, BDAI can also be used to mitigate such risks, for example, to analyse and detect danger. For instance, BDAI could be used to identify irregularities in user behaviour in online banking, which could indicate potential improper use. Certain encryption systems that allow BDAI methods to be used directly on encrypted data could also be used to reinforce resilience against such risks.

The following key questions can be derived:

- Which specific standards – also with regard to algorithm-specific risks – are appropriate and suitable for mitigating information security risks?

- Should existing supervisory principle-based requirements or rule-based control measures be adapted for BDAI-specific matters?

- Which specific BDAI or encryption systems could be suitable for preventing information security risks?

6.2.3 Collective consumer protection[472]

Opinion

Consumers should view big data and artificial intelligence (BDAI) with a healthy dose of scepticism. Of course, tailored products, customised and faster services have their advantages. But most consumers underestimate the price they have to pay for them. They actually pay with their own data and reveal a lot about themselves – in many cases without really being able to tell exactly who will use this information and to what ends.

Elisabeth Roegele

Chief Executive Director of Securities Supervision/ Asset Management at BaFin

BDAI makes it possible, for instance, to relatively accurately predict the behaviour and the needs of consumers – especially if financial data is combined with other data, which some companies, such as bigtechs, have in abundance. Therefore, any company using BDAI methods can exploit these to evaluate what consumers are willing and able to pay and then charge accordingly. Then there is also the risk of unjustified discrimination: All the more reason to raise consumer awareness of this issue. BaFin is doing its best to play its part. We take the legal mandate to keep our consumers informed very seriously.

Financial market players are well advised to deal responsibly with any data their customers supply – even when that data is submitted formally. They are ultimately responsible for ensuring the privacy and informational self-determination of their customers. Data protection authorities are responsible for monitoring the implementation of data protection requirements; however, if data protection violations become more frequent, this could also have implications for BaFin.

---

[472] See Chapter 4.1.

6.2.3.1 Risk of discrimination

BDAI gives firms more and more information about their customers | The quality (zoom in), scope (big data) and predictive strength (predictive analytics) of the information obtained about customers is increasing. This makes it possible to assess risks with growing precision (credit default risk, illness etc.) and to engage in targeted customer contact (segment-of-one). In addition, by evaluating and linking personal data, BDAI makes it possible to segment customer groups with higher granularity.[473] Firms can use this as differentiating factors in product design and in pricing. Overall, the use of BDAI could lead to an increase in power and information asymmetries – to the disadvantage of customers. For example, information asymmetries might arise if consumers do not know the value of their data ("paying with data") and are therefore unaware of the actual price of a product or service.

More possibilities to extract the consumer surplus[474] | From a consumer protection perspective, firms are faced with the dubious incentive to unilaterally extract the consumer surplus, i.e. to take advantage of customers' higher willingness and ability to pay. This could potentially be of particular interest to providers who do not primarily offer financial services but use the knowledge they have gained about customers in their core business, for instance in online retail. By linking financial transaction and behavioural data with the data made available through online services on preferences, needs and wants, such providers will be able to generate a comprehensive view of the individual customer's ability and willingness to pay. This is particularly critical in situations where consumers urgently need a product or service, such as a loan, but do not actually have the choice of several different offers. Overall, consumers have to be made more aware of how their data may be used and how important or valuable it is.

BDAI-induced selection can limit access to financial services for individual consumers | The analysis of BDAI in the insurance sector has also shown that, under certain conditions, a more precise assessment of the risk exposure for individual policyholders could result in increased selection. Although the use of BDAI will not always result in a greater differentiation of risks, and hence of premiums (cf. Chapter 5.3.4),[475] nonetheless, selection mechanisms that are only feasible as a result of BDAI could disproportionately limit individual consumers' access to certain financial services. The situation can be particularly precarious if consumers are disadvantaged by having access to a narrower range of products but are unaware that this is caused by the personal data they have supplied.

Providers must prevent discrimination | Furthermore, the risk of discrimination could increase. Algorithms could be based on features for which differentiation is prohibited by law. When programming algorithms and when checking the generated results, providers have to ensure that individual consumers are not discriminated against. This raises the question as to what monitoring and transparency mechanisms might be helpful here.

Discrimination by approximation of prohibited characteristics | There is also a risk that customer segments could be differentiated on the basis of false assumptions or false conclusions made by algorithms on the basis of these assumptions, and that customers may in fact be discriminated against – even if this is unintentional. In the EU, for instance, private health insurance companies are not allowed to use a person's gender as a differentiating factor. However, by evaluating biographical information that can correlate with gender, it is possible that gender may be used as a differentiating factor in practice. Therefore, algorithms must be programmed in such a way that legal particularities are adequately taken into consideration. As a result, firms must ensure that adequate monitoring and transparency mechanisms are in place to prevent their models from drawing such false or unauthorised conclusions.

---

[473] Cf. Chapter V, Market analyses: banks and insurers.

[474] Cf. explanatory notes in footnote 63.

[475] In segments currently deemed as having low volatility in expected claims, further differentiation will not be possible or necessary.

The following key questions can be derived:

▪ How should supervisory and regulatory authorities respond to companies using BDAI to extract the most consumer surplus?

  o How could consumers be made more aware of the significance and value of their financial data?

▪ How can a holistic supervisory and regulatory approach guarantee that, in addition to the advantages of better risk assessment thanks to BDAI, those customer groups "filtered out by an algorithm" still have sufficient access to affordable financial products?

▪ How can supervisory authorities ensure that customers who are unwilling or unable to provide more data than is legally required will continue to have access to financial services in the future?

▪ What monitoring and transparency mechanisms could help financial services providers prevent the discrimination of groups of consumers?

  o How could existing methods to prevent discrimination be applied to (partially) automated processes?

  o What technical measures should financial services providers take to prevent discrimination, e.g. discrimination based on unauthorised differentiating factors (cf. Chapter 3.5.5 on non-discriminatory data analysis)?

  o How can BDAI algorithms be prevented from unintentionally using characteristics – that the financial services provider is unaware of or is legally prohibited from requesting – for de facto discrimination by approximating such characteristics?

6.2.3.2 Consumer sovereignty

Ensuring trust in the financial market – cooperation with data protection authorities could increase | Even when BDAI is widely used, measures must be taken to ensure that consumers can be confident that the financial market handles their data in a fair and correct manner. In this context, guaranteeing privacy and informational self-determination are important prerequisites for long-term trust. If customer data is misused, the trust in individual firms suffers damage. This loss of trust can also spread to entire market segments. Data protection authorities are indeed responsible for monitoring the implementation of data protection requirements; however, if data protection violations become more frequent, this could also have implications for financial supervisors. Thus, cooperation between financial supervisors and data protection authorities could become increasingly important.

Consumer sovereignty as a catalyst for trust in BDAI innovations | BDAI applications make the use of data a more complex issue – something that also endangers consumer self-determination. Consumers must be allowed to decide for themselves whether they would like to use BDAI-based financial services and, if necessary, permit the use of their data in such cases. Consumers can only make a sovereign decision if they are adequately informed about the potential reach and consequences of the use of their data, if they are given reliable options for controlling how their data is used, and have actual freedom of choice. Providers are responsible for ensuring these prerequisites are met. Financial services providers are also required to enforce the customer rights granted to them under the EU-GDPR, such as the right to delete or change personal data. In terms of adequate consumer information, it is not enough to simply provide the consumer with incomprehensible and complex terms and conditions – most of which are usually accepted although they have not been read. Companies that want to maintain consumer trust, must provide consumers with understandable information about how their data will be used.

Pressure (perceived or actual) to release data could grow | Surveys (such as the Factsheet Data Protection Eurobarometer) show that many consumers feel they no longer have control over their data, for example in the case of online services (cf. Chapter IV). Many also state that they often see no alternative to providing the data if they want to use certain services. If the use of BDAI should require more data in the future than already requested for financial services, it might result in greater perceived or de facto pressure on consumers to provide their data, for instance in order to secure a loan on reasonable terms or to be able to take out insurance at affordable premiums.

Guaranteeing freedom of choice by providing less data-intensive products | In this respect, the question is how supervisory and regulatory authorities should address the issue of adequate freedom of choice in this age of BDAI. One option would be for supervisory and regulatory authorities to ensure that sufficient alternatives in the form of conventional financial services and/or services that are economical with personal data continue to be offered in order to prevent consumers from being virtually coerced into releasing data. How exactly "economical with personal data" and "conventional" are to be defined in this context and which financial services are to be considered are matters to be discussed.

Technical options for using BDAI with anonymised data | Overall, many BDAI innovations can also be created using sufficiently pseudonymous or anonymised data, so that technical data protection measures (e.g. privacy-preserving data mining) can enable innovations even without further recourse to personal data. A "privacy by design" concept can further bolster consumer trust in BDAI innovations. Furthermore, other innovative approaches could be tested and implemented by service providers using BDAI to ensure that the levels of trust are maintained in how personal data is handled. The use of blockchain or distributed ledger technologies would be conceivable examples of such approaches.

The following key questions can be derived:

- How can technical data protection measures – such as privacy-preserving data mining – best contribute to reinforcing consumer trust while exploiting the full potential of BDAI?

- Should supervisory and regulatory authorities ensure financial services that are economical with personal data/conventional are offered as alternatives? How should "economical with personal data/conventional" be defined in this context and which financial services are to be considered?

6.3 Effects on society as a whole

Opinion

In times of increasing digitalisation, big data and artificial intelligence, supervisory authorities such as BaFin are also faced with the challenge of keeping up with the latest advancements in technology. This is particularly important considering the rapid developments in these areas.

Beatrice Freiwald

Chief Executive Director of
Internal Administration
and Legal Affairs
at BaFin

Supervisory authorities cannot allow themselves to be vulnerable here, and it will become increasingly important for them to develop and build on their BDAI expertise both when dealing with companies and providers in the financial market and in the context of supervisory processes. BaFin must therefore provide its existing employees with relevant (further) training and gain as much knowledge on BDAI as possible when recruiting new staff. It is not always easy for public authorities to attract and retain the right talent. However, BaFin has a lot to offer as an employer: varied and challenging tasks in an international context alongside the opportunity to assume responsibility at an early stage, contribute ideas and help to shape the development of financial supervision.

Supervisory authorities, too, will have to invest in BDAI-supported processes. BDAI has the potential to improve the quality of traditional supervisory processes, allowing patterns to be detected faster, for example.

The general shift towards digital processes and systems is set to bring improvements in effectiveness and efficiency, particularly in standardised processes. However, financial supervision is and will continue to be a highly flexible process that focuses on the assessment of complex issues. And high-quality supervision is characterised by careful assessments involving difficult discretionary decisions. An increase in standardisation is often associated with a decrease in flexibility – whether and how BDAI will change this remains to be seen.

As already outlined, the use of BDAI affects society as a whole and raises important questions outside the areas of financial supervision and regulation. For example, it could lead to a structural transformation of the economy, involving a significant reduction in jobs. This could in turn affect supervised firms. Accordingly, the potential supervisory and regulatory implications derived in the previous chapter are to be considered while taking into account the effects of BDAI on society as a whole.

Therefore, the section below briefly addresses the main societal topics associated with the use of BDAI in the financial services sector.

Figure 28: Embedding the effects of BDAI on supervision and regulation into a societal context
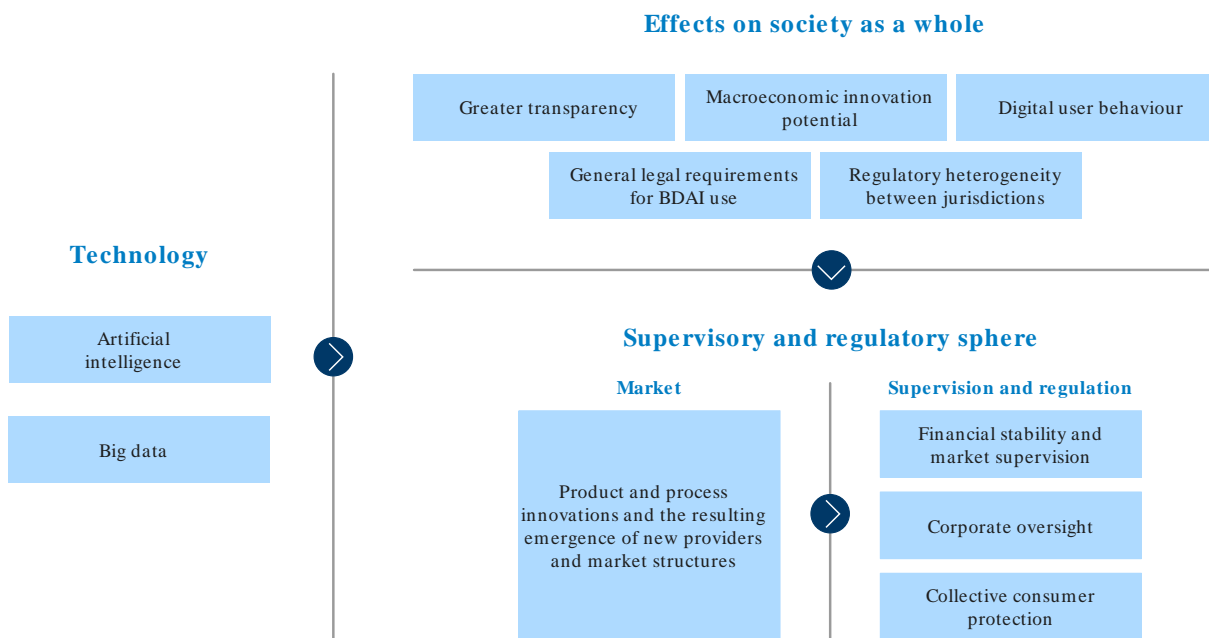
Greater transparency | BDAI can contribute to greater transparency in terms of both the market and its participants (firms and consumers)[476]. As a result, providers can increase customer value with new and innovative offerings as BDAI allows them to gain a better understanding of the economic situation and customer preferences. On the other hand, BDAI gives firms the opportunity to extract more of the consumer surplus through individualised and situational pricing, as discussed above. In a market with sufficient competition, however, this effect could be offset by (standardised) comparison options and a broad range of alternative offers, provided these are independent and made available to consumers, e.g. via platforms.

Macroeconomic innovation potential | BDAI allows for gains in efficiency and effectiveness in many areas of business and society. The associated innovation potential in processes and products could also boost the real economy and produce positive external effects. In this way, greater availability and better analysis of data could contribute to mitigating risks and damage to society as a whole. One example in this regard would be the telematics products increasingly offered by insurance companies. The potential prevention of damage associated with telematics rates can prevent accidents (in motor vehicle insurance) and thereby contribute to reducing risk and claims overall.

Digital user behaviour | Digital user behaviour is often analysed by accessing personal data or data that allows conclusions to be drawn about individual consumer characteristics. This data often constitutes a key component of BDAI-based innovations and services. At the same time, the information derived from digital user behaviour can be used to influence the individual behaviour and limit a consumer's freedom. The increasing amount and greater diversity of the data generated and provided can make it more difficult for many consumers to evaluate what data a company has in its possession and who else has access to this data. Consumers could feel compelled to change their behaviour in response.

---

[476] This does not apply equally to all market participants, since it depends on whether they have access to BDAI applications that create greater transparency. It is possible that transparency for the consumer could tend to decrease as the number of providers increases if the consumers do not have access to the corresponding BDAI tools that create transparency.

General legal requirements governing the use of BDAI | How and where algorithms can and should be used and which requirements should apply to their developers and users are key elements that are being discussed at a political and legislative level. There are already a variety of initiatives at both the national and European level.[477] Regardless of the issues specific to the financial markets, the general requirements governing BDAI usage generated by these initiatives would also have to be taken into consideration by the financial markets. One example of the definition of legal requirements for BDAI usage could be the provision of evidence in legal proceedings which deal with algorithm-based decisions and activities. Admissibility of evidence will face new challenges due to the highly complex nature of algorithm-based methods. This applies in particular when written evidence is required.

Regulatory heterogeneity between jurisdictions | While the effects of BDAI-based business models seem to be similar in various countries, regulatory and supervisory practices may still differ from one country to another. This increases the risk of supervisory arbitrage. The phenomenon of BDAI makes it all the more important for countries to agree on equivalent regulation and supervisory practices along the lines of "same business, same risks, same rules". This is something that must be ensured in future for every sphere in which BDAI is used – especially for new market structures. At the same time, the distinctive features of each national market must be sufficiently taken into account, both now and in the future.

---

[477] For example, the European Parliament called on the European Commission to propose common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots and their subcategories and to introduce a European system of registration: cf. resolution of the European Parliament of 16 February 2017, with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). The Advisory Council for Consumer Affairs, an advisory body to the Federal Ministry of Justice and for Consumer Protection recommends legal regulation of algorithms to comply with the specifications of consumer protection laws, data privacy laws, anti-discrimination laws and digital security; cf. Digitale Souveränität. Report on Consumer Issues by the Advisory Council for Consumer Affairs (2017).

VII. Appendix

7.1 In-depth description of machine learning

Machine learning will be described in the following from the perspective of content, i.e. based on the learning task. For this purpose, the most important learning tasks, their differences and areas of application will be explained.

> **Running example**
>
> To better understand the individual algorithms, methods and the learning tasks, a simplified running example will be used in the side notes. It allows for different learning tasks and thus the application of different methods. The sample task describes possible issues that might be of interest in relation to car insurance.
>
> The first issue concerns the prevention of contract cancellation. The task is to predict whether a client is going to cancel his contract, is considering cancelling his contract but can be persuaded not to do so by the company taking appropriate action, or is tending towards not cancelling the contract. In other words, historical client decisions are to be used to predict one of the three above-mentioned categories for new situations. For the model, the factors that can potentially influence the client and make him cancel the contract must be established first. For reasons of simplification, the example only considers the no claims bonus period in years and the vehicle mileage per year. However, factors which initially appear insignificant such as weather data and Twitter feeds can also influence a client's decision.
>
> The second issue involves the prediction of the expected amount of an insurance damages claim in euros. In contrast to the first issue, a numeric variable is to be modelled in this case rather than fixed categories. For reasons of simplicity, only one feature is taken into account for the prediction, namely the age of the vehicles in years.

7.1.1 Supervised learning

Supervised learning methods are often used to classify examples in certain categories or classes (detection of money laundering, classifying creditworthiness and detection of objects and persons in images). Such tasks are termed classification tasks. Figure 29 illustrates such a task for the first question in the running example, where the contract cancellation class is to be predicted for a car insurance client. A suitable method thus initially learns the classification based on the data points with known labels (circles) and then predicts the labels for unknown data (squares).

Figure 29: Illustration of the classification of new data points (squares) on the basis of labels for known data points (circles) for the sample dataset using the example prediction of contract cancellation class
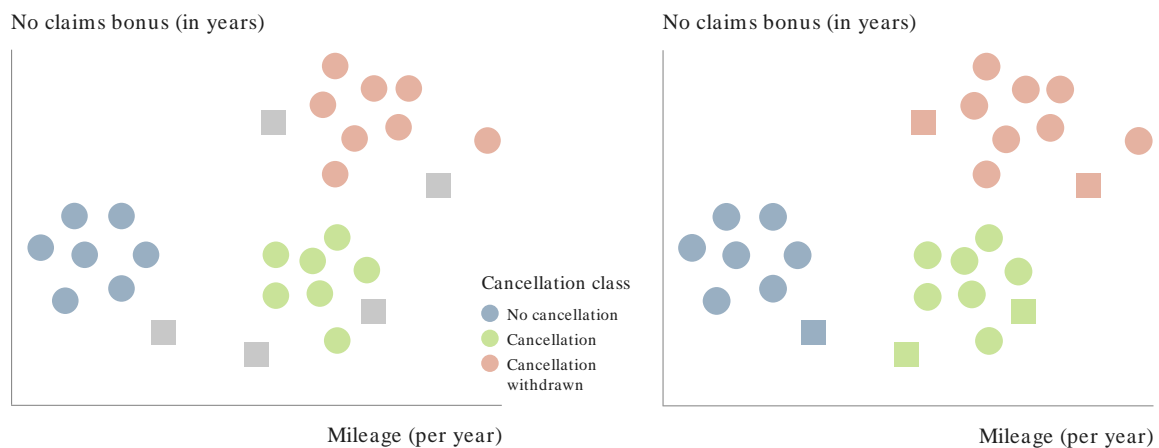
Known classification methods include decision trees, support vector machines, feed-forward neural networks and logistic regression, for example, which will be explained in more detail within side notes. The regression methods of linear regression and support vector regression will also be explained.

The respective assumptions of the methods must be checked and the types of features must be examined in order to select the right learning method. While a linear regression can only consider and predict numerical attributes, for example, a variance analysis can also use a mixture of numerical and categorical attributes for estimates.
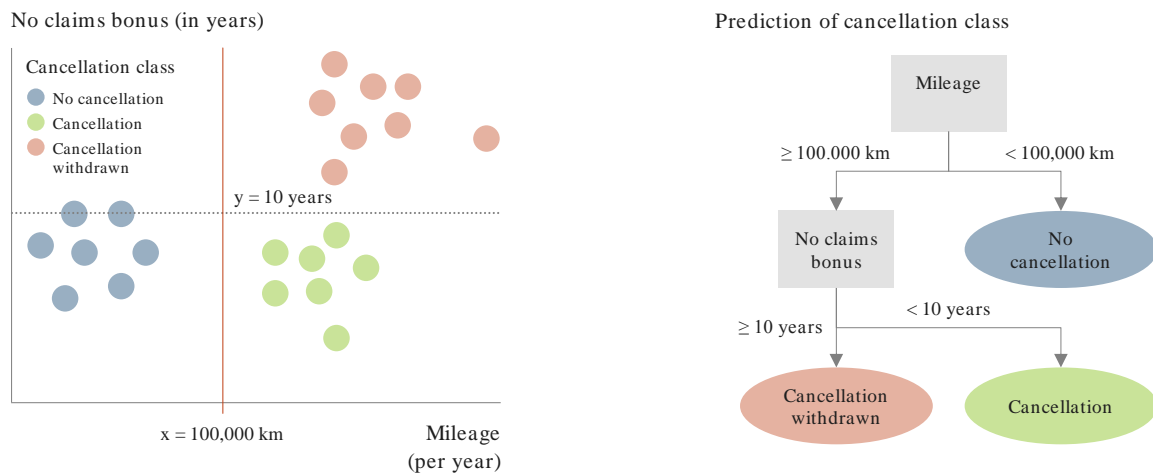
Side note: Decision tree learning

One well-known example of a supervised learning method that can be used for classification is decision tree learning[478]. A decision tree provides a series of decision questions that can be used to answer a question. To obtain the answer, the decision process starts from a root node, from which the tree is then followed through. At each node, the decision is taken as to which path is to be followed on the tree depending on the characteristic of a specific feature. Finally, a leaf node is reached that contains the answer to the question.

Decision tree learning involves learning the best possible decision tree automatically from labelled data. For this purpose, all the possible decisions are tested for a dataset, whereby for nominal features individual features are compared with a fixed value (e.g. "Weekday = Monday") and for numeric features individual features are compared with fixed threshold values (e.g. "Stock exchange price < €250"). Which decision is best to make the label predictable is then evaluated using a statistical measure. Subsequently, a sub-decision tree for fine-tuning the forecast is learned recursively on the basis of the resulting partial datasets.

---

[478] Quinlan 1993, C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco.

If the dataset above is considered again, the task of predicting the contract cancellation class on the basis of the no claims bonus in years and mileage per year can be answered using a decision tree. As shown in Figure 30, a tree with two rules would suffice to classify a new data point clearly in one of the three contract cancellation classes. The first rule asks whether the mileage per year is higher or lower than 100,000km. If it is lower, the class "no contract cancellation" applies. If it is higher, the no claims bonus limit of 10 years is decisive as to whether the client can be expected to cancel the contract or can be persuaded not to cancel.

Figure 30: Possible decision tree with respective rules that could be used for the example in order to predict contract cancellation status on the basis of mileage and no claims bonus

Numerous further developments are possible taking the basic algorithm of decision tree learning as a basis. Decision trees can be adapted to solve regression tasks, too. Due to their simplicity and efficiency, decision trees often serve as a basis for ensemble methods which combine several simple methods in order to be able to learn more complex or more robust models. One example of this is a random forest which initially generates many trees, which are not correlated if possible, by pulling elements randomly from the training set, and on this basis the tree is then prepared using the randomly selected combinations of features. The mean value of the individual tree predictions is taken as the basis for the prediction. Rule learning is closely related to decision tree learning. In this case, numerous individual if-then rules are learned instead of trees.

Side note: Support vector machine

The support vector machine[479] (SVM) is a frequently used method of classification. The concept is to find a decision boundary between the classes in such a way that the margin to each of the classes is as wide as possible. This makes the classification less susceptible to noise. Since this method is difficult to explain and illustrate for the division of three classes, it is assumed here for the sake of simplicity that only two contract cancellation classes are to be predicted. For this purpose, clients who cancel their contract and clients who withdraw their cancellation are merged in one class, namely clients who try to cancel their contract. If these data are to be classified using an SVM, as shown in Figure 31, the method would place the margin between the two groups as shown.

---

[479] Cortes et al., 1995, Support–vector networks. In: Machine Learning Journal, 20: pp. 273–297.

No claims bonus (in years)

Cancellation class
● No cancellation
● Attempted
  cancellation

Mileage (per year)
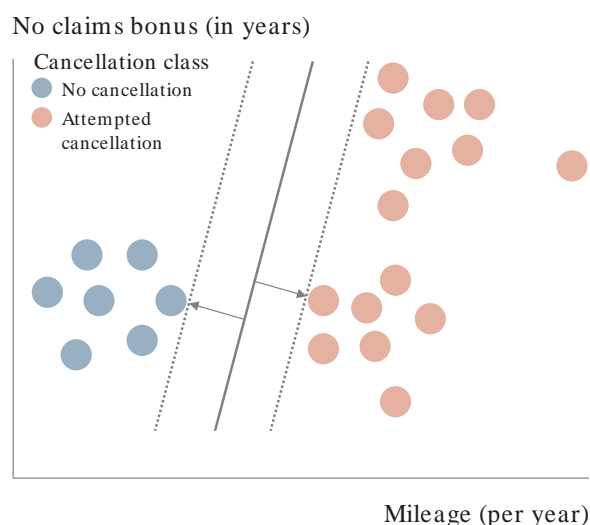
In the simplest version of the SVM the data are classified using a linear function – also referred to in mathematics as a hyperplane. If the objects cannot be divided linearly, the kernel trick is applied, where a transformation function is used to transform the data objects into a higher-dimensional space where they can then be separated by a linear hyperplane. This approach becomes efficient by virtue of this transformation not having to be computed explicitly. Reverse transformation results in a non-linear dividing plane. The advantages of this method are high flexibility and a very good theoretical base.[480]

Side note: Neural networks

(Artificial) neural networks[481] are one currently extremely popular class of algorithms. The first papers on neural networks were published as long ago as the 1950s[482] and are inspired by the functional aspects of the human brain. The smallest unit of a neural network, the neuron, is represented mathematically by a function which multiplies a vector of numeric inputs by linear weights and then transforms these using a non-linear activation function. Individual types of neurons are distinguished on the basis of numerous possible activation functions. The weights vector corresponds to the model to be learnt by the neuron. A neural network is only created by individual neurons being connected in a variety of different ways, i.e. the outputs of individual neurons in turn serve as inputs for further neurons.

The way the neurons are connected is known as the network architecture. This can have different levels of depth and complexity. One of the simplest neural networks is the single-layer feed-forward network, where the information received via the input layer is processed to the output layer via a hidden layer (i.e. a number of neurons with the same input but different weights). This results in classification output. The training of the neural network, i.e. adjustment of the individual weights, is carried out by means of mathematical optimisation. Whereas training the network requires a lot of time-consuming computing, the application of the networks can take place in real time. Figure 32 illustrates one possible use of a single-layer feed-forward network for the classification of the clients cancelling their contract in the sample dataset.

---

[480] Vapnik 1998, Statistical Learning Theory. John Wiley & Sons, Inc.

[481] Bishop, 1995, Neural networks for pattern recognition. Clarendon Press.

[482] Rosenblatt, 1958, The perceptron - a probabilistic model for information storage and organization in the brain. Psychological Review, 65, pp. 386-408.

The learning process uses the method of backpropagation, where the output decisions are compared with the actual, correct answers in order to compute output errors. These error values are then inputted to the network again and used by the algorithm to tune the weights at each node. After several runs, such a network converges in such a way that errors in prediction are low.

Compared with many other methods, neural networks can even make complex decisions and thus find decision boundaries which are not subject to any specific structural conditions. In addition, neural networks can be trained efficiently even on the basis of very large quantities of data. Their disadvantages include low model transparency.

Figure 32: Possible use of a single-layer feed-forward neural network to classify contract cancellation for new clients for the sample dataset
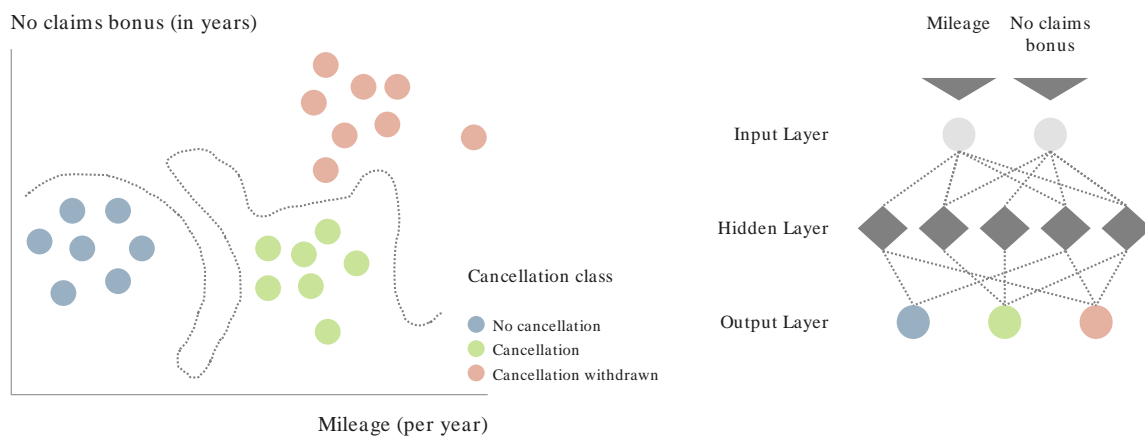
Side note: Logistic regression

Despite its name, logistic regression is in fact a method of classification. However, in contrast with other methods, it is used not only to learn to predict the class but also the probability of an example belonging to a certain class. The term logistic regression is used because technically a regression approach is used. The values of the features are multiplied by a coefficient and added to form a score. A mathematical function uses this score to calculate the probability of the label value. If the features to be explained are classes in themselves, e.g. different professions, the result is one score per class.

Thanks to its simplicity and high degree of transparency, logistic regression is often used in the financial industry, for instance in credit rating models (probability of non-payment) and in fraud detection. Experts often make manual corrections – particularly for cases which only occur rarely in the data or for specific upward or downward revaluation in the interest of the business. Despite the undisputed advantages of this method, the quality of the logistic regression model is often significantly lower than that of the non-linear models described above.

Side note: Linear regression

One important application for regression tasks is the forecasting of risks, product demand or optimum prices. Linear regression is a method very frequently used for this purpose. Here, a label y is explained by a linear model based on one or more input features. A distinction is made between a single-variate model with only one input feature x and a multivariate model with several input features x_1, x_2, ..., x_n. A single-variate model is described using the equation y = m*x + b, where m describes the degree of the influence of x on y and b is a constant. For the above example and the second issue, the insurance damage in euros could thus be modelled on the age of the vehicle in years, as shown in Figure 33.

The advantages of linear regression are its simplicity and the fact that non-linear dependencies have been proven to be approximated well by a suitable linear function particularly for high-dimensional data. The disadvantage is the lack of flexibility when the data is distributed differently.

Figure 33: Possible result of a linear regression for the sample dataset which models the insurance damage depending on the age of the vehicle
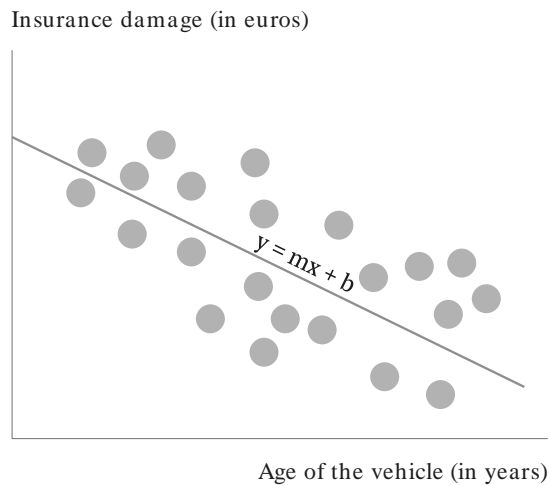
Side note: Support vector regression

Figure 34: Possible result of a support vector regression for the sample dataset. The maximum permitted error results in an upper and lower bound within which most examples are located
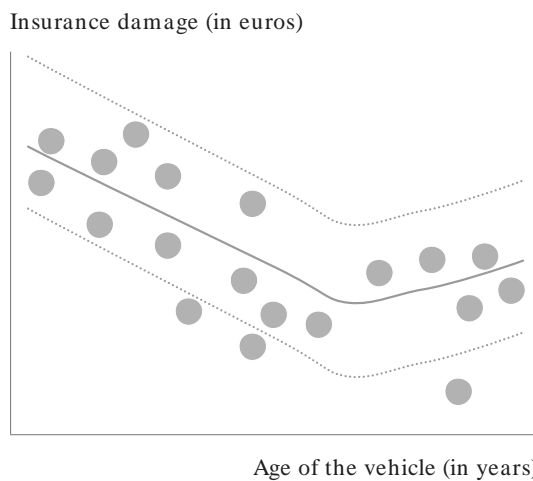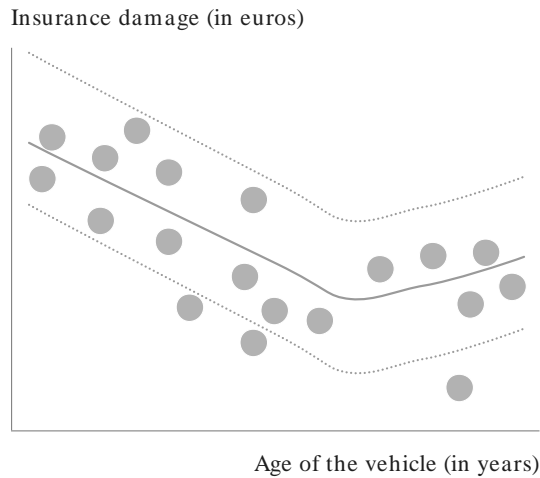

Support vector regression[483] extends the idea of the support vector machine to regression. Rather than finding a linear function which separates two classes with as wide a margin as possible, the very similar mathematical problem is solved by finding as simple a function as possible that matches the label as exactly as possible with one specified maximum permitted error. The advantage of support vector regression is that a non-linear function can be learned very easily using the kernel trick described above. In addition, the underlying theory of both support vector regression and support vector machines has been well researched. The disadvantage, however, is a high runtime when applied to large data volumes.

---

[483] Drucker et al., 1997, Support Vector Regression Machines. In: Advances in Neural Information Processing Systems 9, pp. 155-161.

Insurance damage (in euros)

Age of the vehicle (in years)

7.1.2 Unsupervised learning

Unsupervised learning uses different methods to extract intrinsic patterns from the data. In contrast to supervised learning, there is no specific question that can be coded in a concrete target feature. Two of the most important learning tasks in unsupervised learning are clustering and anomaly detection.

7.1.2.1 Side note: K-means clustering

One of the best-known clustering methods is k-means clustering, where a sample of data points is divided into a given number of k clusters. Clusters are defined by the centre (centroid) and a distance function. A point belongs to the cluster whose centroid is nearest. Initially, a random number k of cluster centres is chosen and the individual data points are then assigned to the cluster centre they are nearest to. Then the actual centres of the resulting clusters are computed and the data points are assigned to the nearest centre again. This step is repeated as often as necessary until the distribution of points in the k clusters no longer changes.

For the above example, Figure 36 illustrates two possible clustering results for k = 2 (left) and k = 3 (right) when no claims bonus in years (y-axis) and mileage per year (x-axis) are considered:

No claims bonus (in years)   No claims bonus (in years)

k = 2   k = 3

Mileage (per year)   Mileage (per year)
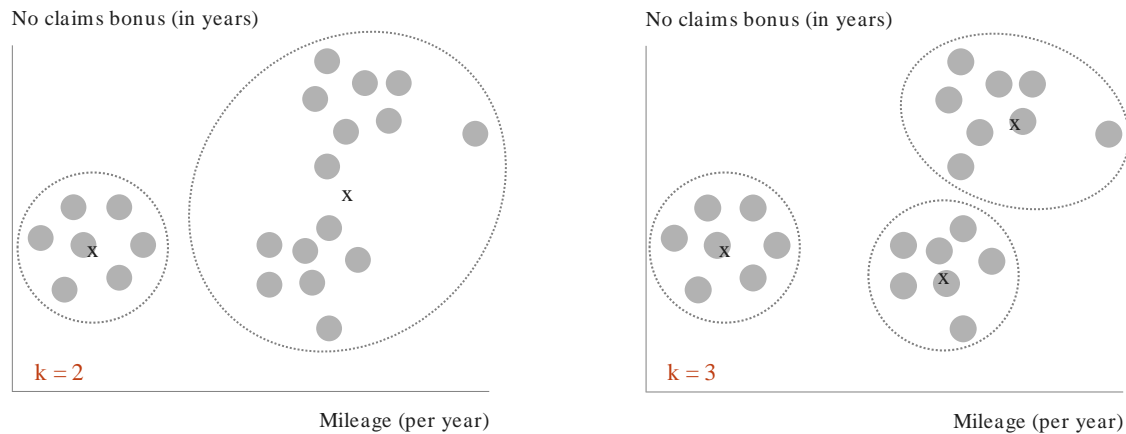
K-means clustering is a very effective method and is often used. The disadvantage, however, is that the results are heavily dependent on the features and parameters selected by the data scientist, which permits a whole variety of different results. In addition, this method cannot cleanly separate overlapping groups.

7.1.2.2 Side note: anomaly detection

One prominent approach to anomaly detection relies on detecting the structure of most of the data first – by clustering and simple statistical distribution assumptions, for example – and then finding points which do not follow this structure. They thus set up a hypothesis as to what is "normal" and belongs together and what, by contrast, deviates from the so-defined normality. A second approach is to directly define functions that evaluate the degree of deviation of a data point.

In the k-nearest-neighbour method for anomaly detection, for instance, the distance between one example and its nearest neighbour (or in general the distance to the k-nearest-neighbour) is defined as the degree of deviation. In Figure 37 the marked point is an anomaly on the basis of both approaches: it neither matches the pattern of the three clusters nor is it near to its nearest neighbour.

The disadvantage of anomaly detection is that the results are often not intuitive. Anomaly detection seems to be a simple task for humans for a simple two-dimensional case, but in fact it presents an extremely complicated problem for higher dimensions.

Figure 37: Possible outlier in the sample dataset (marked in red) which could be identified by anomaly detection

No claims bonus (in years)

Mileage (per year)
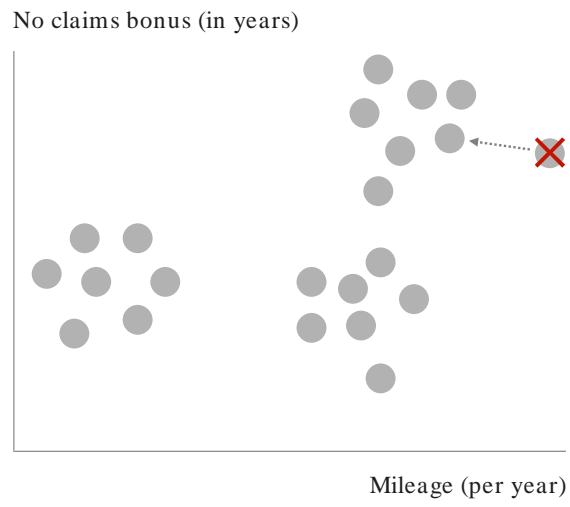
7.2 Summary of the papers of the European supervisory authorities and the Financial Stability Board

This study follows a series of publications on the topics of big data, advanced analytics and artificial intelligence in the financial services industry that were recently issued by leading institutions in the regulatory and supervisory area. The reports by the Financial Stability Board, the Joint Committee of the European Supervisory Authorities, the Basel Committee on Banking Supervision and the European Banking Authority are of particular note in that regard. Moreover, the topic of big data in the financial sector has also been addressed by the European institutions (the European Parliament and the EU Commission).[484] The contents of these publications are summarised in the following section.

Table 1: Reports on Big Data and Artificial Intelligence in the supervisory and regulatory area

Institution	Report title	Publication
Joint Committee of the European Supervisory Authorities	Final Report on Big Data	March 2018
Financial Stability Board	Artificial intelligence and machine learning in financial services. Market developments and financial stability implications	November 2017
Basel Committee on Banking Supervision	Sound Practices: Implications of fintech developments for banks and bank supervisors (consultation paper)	August 2017
European Banking Authority	Discussion Paper on the EBA's approach to financial technology (FinTech)	August 2017
Financial Stability Board	Financial Stability Implications from FinTech. Supervisory and Regulatory Issues that Merit Authorities' Attention	June 2017

Joint Committee: Final Report on Big Data (March 2018)

The Joint Committee of the European Supervisory Authorities (ESAs)[485] published a report on the use of big data.[486] The report examines stakeholder feedback on a discussion paper on the use of big data by financial institutions, published by the Joint Committee in December 2016.[487] This report evaluates the potential benefits and risks associated with how financial institutions use big data – these aremainly identical to those mentioned in the discussion paper. The report does not recommend any legislative action. The reason being that the existing legal requirements in the relevant segments (data protection, internet security, consumer protection and sector-specific financial legislation) already form a substantial framework for minimising risk.

---

[484] Cf. European Parliament, 2016, Report on FinTech: The influence of technology on the future of the financial sector (2016/2243(INI)); European Commission, 2017, Consultation Document: FinTech: A more competitive and innovative European financial sector; European Commission, 2017, Assessment of current and future impact of Big Data on Financial Services. Available online: https://ec.europa.eu/info/sites/info/files/file_import/1606-big-data-on-financial-services_en_0.pdf, accessed on 16 November 2017; EBA, 2017, Discussion Paper on the EBA's approach to financial technology (FinTech), pp. 9 ff. Available online: https://www.eba.europa.eu/-/eba-publishes-a-discussion-paper-on-its-approach-to-fintech, accessed on 16 November 2017.

[485] The European Banking Authority (EBA), the European Insurance and Occupational Pensions Authority (EIOPA) and the European Securities and Markets Authority (ESMA).

[486] Cf. Joint Committee of the European Supervisory Authorities, 2018, Final Report on Big Data. Available online: https://esas-joint-committee.europa.eu/Publications/Reports/Final%20Report%20on%20Big%20Data.pdf, accessed on 25 April 2018.

[487] Cf. European Securities and Markets Authority, 2016, Discussion Paper on the Use of Big Data by Financial Institutions. Available online: https://www.esma.europa.eu/press-news/consultations/joint-committee-discussion-paper-use-big-data-financial-institutions, accessed on 7 November 2017.

This framework is further supported by other important legislation (e.g. IDD[488], MiFID II, PSD2 and GDPR). The ESAs should monitor and evaluate the impact of these new regulations.

To keep pace with the development of big data, the supervisory authorities should continue to work hand in hand to ensure adherence to all the relevant regulations. Moreover, the ESAs suggest that financial institutions implement good practices for big data usage (robust big data processes and algorithms, consumer protection and disclosure of how big data is used).

The final report was published along with a fact sheet ("Use of Big Data by financial institutions")[489] summarising in simple language the main messages in order to raise consumer awareness of the opportunities and risks of big data.

Financial Stability Board (FSB): Artificial intelligence and machine learning in financial services. Market developments and financial stability implications (November 2017)

In November 2017, the Financial Stability Board published a report on artificial intelligence and machine learning in the context of financial services.[490] The analysis and assessment are focused on the implications for financial stability.[491] The report looks at use cases in financial institutions in order to discuss the new technologies. From the point of view of the FSB, the rapidly expanding areas of application include fraud detection, capital optimisation and portfolio management. At the same time, the report emphasises that currently available data is not yet sufficient to conclusively assess the impact of AI and machine learning.

The FSB states that AI and ML can be used to improve efficiency within the financial system. In addition, regulatory technology (regtech) and supervision technology (suptech) could help to improve regulatory compliance. Network effects and economies of scale could, however, result in system-relevant players emerging outside of the regulated sector. New data correlations give rise to new forms of interdependence between the financial market and financial institutions, producing unexpected challenges. From a regulatory and supervisory perspective, the FSB also emphasises the difficulties that could arise through non-transparent models for interpretability and supervisory assessment. Corresponding competences and skills would therefore be required in order to understand and monitor the new, complex AI and ML approaches. From the FSB's perspective, it is also crucial to comply with relevant requirements in the areas of data protection, conduct risks and cybersecurity.

---

[488] Insurance Distribution Directive.

[489] Cf. European Securities and Markets Authority, 2018, Use of Big Data by financial institutions. Available online: https://www.esma.europa.eu/sites/default/files/library/big_data_factsheet_de.pdf, accessed on 25 April 2018.

[490] Cf. Financial Stability Board, 2017, Artificial intelligence and machine learning in financial services. Market developments and financial stability implications. Available online: http://www.fsb.org/2017/11/fsb-considers-financial-stability-implications-of-artificial-intelligence-and-machine-learning/, accessed on 6 November 2017.

[491] The paper also includes two appendices in which the FSB addresses "Legal issues around AI and machine learning" and "AI ethics". Here, the topics of data protection and data ownership are also examined in the context of artificial intelligence and machine learning.

Basel Committee on Banking Supervision (BCBS): Sound Practices: Implications of fintech developments for banks and bank supervisors (August 2017)

In light of the increasing investments in fintechs and the resulting changes that are expected in the financial sector, the BCBS has set up a task force to examine the importance of fintechs for the supervisory authorities and banks. The BCBS presented its results in its consultation document "Sound Practices"[492]. In this document, five scenarios are identified for the development of the banking market, from which the BCBS derives recommendations for supervisory and regulatory authorities. The scenarios are based on the assumption that it will become increasingly difficult for banks to maintain their business models against the backdrop of technological change and changing customer expectations, with competition for customer relationships set to become the key issue. The BCBS assumes that the impact of the current technological changes will be more disruptive than that of earlier innovations. As central factors, the BCBS identified the rapid acceptance of fintech services and lower entry barriers.

European Banking Authority (EBA): Discussion Paper on the EBA's approach to financial technology (fintech) (August 2017)

The European Banking Authority also published a consultation paper on the topic of fintech in 2017.[493] In this paper, the EBA presents the results of a mapping exercise of the European fintech sector that it conducted in cooperation with national supervisory authorities. On the basis of the results, the EBA determined a number of areas where it deems further analyses and follow-up work necessary. The EBA suggests the following priority areas in its consultation paper: authorisation and sandboxing regimes; prudential risks for credit institutions, payment institutions and electronic money institutions; the impact of fintech on the business models of these institutions; consumer protection and retail business conduct; the impact of fintech on the resolution of financial firms; and the impact of fintech on anti-money laundering and countering the financing of terrorism.

Financial Stability Board: Financial Stability Implications from Fintech. Supervisory and Regulatory Issues that Merit Authorities' Attention (June 2017)[494]

Currently – according to the provisional conclusion – fintech innovations do not pose a risk to financial stability. However, the FSB does highlight ten issues that regulatory and supervisory authorities should draw attention to. The FSB also emphasises the importance of developing a better understanding of how business models for start-ups and established companies will actually be influenced by fintech.

An overarching topic in the FSB report is international cooperation. The FSB specifies three priority areas for action here in terms of regulation and supervision: managing operational risks from third-party service providers, mitigating cyber risks and monitoring macrofinancial risks. Other relevant topics according to the FSB are open questions regarding cross-border issues, limited traceability of big data models, agility and flexibility of the regulatory framework, cooperation between the private and the public sector and expanding communication between the relevant supervisory and regulatory authorities. The FSB also highlights the need to build staff capacity in new areas of required expertise in order to be able to adequately address the developments triggered by fintech. Lastly, the FSB suggests studying the implications of digital currencies.

---

[492] Cf. Bank for International Settlements, 2017, Sound Practices: Implications of fintech developments for banks and bank supervisors. Basel Committee on Banking Supervision. Available online: https://www.bis.org/bcbs/publ/d415.htm, accessed on 16 November 2017.

[493] Cf. European Banking Authority, 2017, Discussion Paper on the EBA's approach to financial technology (FinTech). Available online: https://www.eba.europa.eu/-/eba-publishes-a-discussion-paper-on-its-approach-to-fintech, accessed on 16 November 2017.

[494] Financial Stability Board, 2017, Financial Stability Implications from FinTech. Supervisory and Regulatory Issues that Merit Authorities' Attention. Available online: http://www.fsb.org/wp-content/uploads/R270617.pdf, last accessed on 7 November 2017.