Varmaz, Nermin

Article

GDPR vs. Big Data & AI in FinTechs

Vierteljahrshefte zur Wirtschaftsforschung

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Varmaz, Nermin (2020) : GDPR vs. Big Data & AI in FinTechs, Vierteljahrshefte zur Wirtschaftsforschung, ISSN 1861-1559, Duncker & Humblot, Berlin, Vol. 89, Iss. 4, pp. 55-71, https://doi.org/10.3790/vjh.89.4.55

This Version is available at:
https://hdl.handle.net/10419/270514

GDPR vs. Big Data & AI in FinTechs

NERMIN VARMAZ*

Nermin Varmaz, Anheuser-Busch InBev, nermin@varmaz.de

Summary: This article addresses the compliance of the use of Big Data and Artificial Intelligence (AI) by FinTechs with European data protection principles. FinTechs are increasingly replacing traditional credit institutions and are becoming more important in the provision of financial services, especially by using AI and Big Data. The ability to analyze a large amount of different personal data at high speed can provide insights into customer spending patterns, enable a better understanding of customers, or help predict investments and market changes. However, once personal data is involved, a collision with all basic data protection principles stipulated in the European General Data Protection Regulation (GDPR) arises, mostly due to the fact that Big Data and AI meet their overall objectives by processing vast data that lies beyond their initial processing purposes. The author shows that within this ratio, pseudonymization can prove to be a privacy-compliant and thus preferable alternative for the use of AI and Big Data while still enabling FinTechs to identify customer needs.

Zusammenfassung: Dieser Artikel befasst sich mit der Vereinbarkeit der Nutzung von Big Data und Künstlicher Intelligenz (KI) durch FinTechs mit den europäischen Datenschutzgrundsätzen. FinTechs ersetzen zunehmend traditionelle Kreditinstitute und gewinnen bei der Bereitstellung von Finanzdienstleistungen an Bedeutung, insbesondere durch die Nutzung von KI und Big Data. Die Fähigkeit, eine große Menge unterschiedlicher personenbezogener Daten in hoher Geschwindigkeit zu analysieren, kann Einblicke in das Ausgabeverhalten der Kunden geben, ein besseres Verständnis der Kunden ermöglichen oder helfen, Investitionen und Marktveränderungen vorherzusagen. Sobald jedoch personenbezogene Daten involviert sind, kommt es zu einer Kollision mit allen grundlegenden Datenschutzprinzipien, die in der europäischen Datenschutzgrundverordnung (DSGVO) festgelegt sind, vor allem aufgrund der Tatsache, dass Big Data und KI ihre übergeordneten Ziele durch die Verarbeitung großer Datenmengen erreichen, die über ihre ursprünglichen Verarbeitungszwecke hinausgehen. Der Autor zeigt, dass sich in diesem Verhältnis die Pseudonymisierung als datenschutzkonforme und damit vorzugswürdige Alternative für den Einsatz von KI und Big Data erweisen kann, die FinTechs dennoch in die Lage versetzt, Kundenbedürfnisse zu erkennen.

→ JEL classifications: K23, K29
→ Keywords: FinTech, GDPR, AI, pseudonymization, anonymization, regulation

* The author is legal counsel at Anheuser-Busch InBev. This article reflects the opinion of the author.

I Introduction

"G.D.P.R. (...) Makes Europe World's Leading Tech Watchdog"[1]

Europe as a pioneer for data protection principles and as a data privacy supervisor of globally operating tech companies – this could be one way of interpreting the headline in the New York Times on the day the General Data Protection Regulation[2] came into force. It is not only the choice of words that is astonishing, but also the perceived scope of a European regulation raised to relevance for the global tech market: the GDPR as a promise of more transparency for digital citizens; the GDPR as an act of empowerment in a world that is changing rapidly when it comes to digitization; the GDPR, of course, as means of regaining control of disruptive technologies. In the two years since it came into force, the GDPR has indeed proven to be the most comprehensive regulatory framework in information policy with an impact far beyond the European Union (hereinafter referred to as: E.U.).[3] According to a Fundamental Rights Survey, the GDPR is unmistakably gaining awareness: 69 % of the E.U. population above the age of 16 have heard about the GDPR and 71 % of people in the E.U. know about their national data protection authority.[4] This is also due to the relevance in everyday life – the GDPR regulates the legal conditions under which personal data may be processed. This is given enormous effect by the fact that the citizens of the E.U. have been granted comprehensive rights vis-à-vis data processing companies, a policy which has decisively turned the data sovereignty in favor of the citizens. These very citizens have already made extensive use of such rights – the awareness of their rights according to the GDPR and the issues surrounding data privacy is at an all-time high.[5]

The FinTech industry had no less disruptive tendencies when it entered the market for financial products. Young start-up companies in the field of cashless payment and FinTech have increasingly established themselves as competitors to conventional credit institutions.[6] The latter did not show much effort for digitization in recent years, when digitization was advancing at a considerable speed across all industries.[7] FinTech companies took up this challenge to enter the market with the digitalized supply of financial services. Cashless payment systems and FinTech solutions are replacing physical currencies to improve the transfer of funds from one party to another, using

1 Cf. *Satariano,* printed in the New York Times on May 25, 2018 in section A, page 1 of the New York edition, available online at https://www.nytimes.com/2018/05/24/technology/europe-gdpr-privacy.html, last accessed August 31, 2020.

2 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (hereinafter referred to as: GDPR).

3 The state of California, for example, has adopted a comprehensive set of regulations that is partly modelled on the GDPR (cf. CCPA – California Consumer Privacy Act). The fact that this jurisdiction has adopted at least some of the basic E.U. principles shows that the GDPR will have a lasting effect on global data protection.

4 Cf. *European Union Agency for Fundamental Rights;* Fundamental Rights Survey 2019. Data protection and technology, available online at: https://fra.europa.eu/en/publication/2020/fundamental-rights-survey-data-protection, last accessed August 31, 2020.

5 Cf. *European Commission,* Communication from the Commission to the European Parliament and the Council: Data protection as a pillar of citizens' empowerment, and the EU's approach to the digital transition – two years of application of the General Data Protection Regulation, COM(2020) 264 final, June 24, 2020, p. 8.

6 Cf. *Paul,* WPG 2016/2, 57.

7 Cf. *Dorfleitner/Hornuf,* FinTech und Datenschutz, 2019, p. 3.

technology to make financial activities more efficient.[8] In the long term, it may be safe to assume that FinTechs will gain a more important role and transform the financial industry.[9] Acting as a FinTech regularly requires the processing of personal data due to the daily customer relationship, which makes the GDPR applicable.[10] In an analysis published in 2019, 98 percent of FinTech companies' privacy policies stated that they process personal data.[11] Although currently not an essential component, the processing of personal data could be a necessary factor for improving the supply of services, especially in the context of disruptive technologies such as Big Data or Artificial Intelligence (hereinafter referred to as: AI).[12]

Yet, there is another way of interpreting above's headline in the New York Times – the GDPR as a regulatory framework blocking innovation and assuming sovereignty over data traffic on a global scale. On that note, the use of Big Data and AI appears to be opposed to the GDPR. The adoption of the GDPR substantiated that personal data is no longer an arbitrary economic asset[13], but an outflow of the right of informational self-determination initially developed by the German Federal Constitutional Court[14] and now embodied in a comprehensive regulatory framework that acknowledges data privacy as a fundamental right.[15] These regulations prohibit the processing of personal data unless permitted by law, arg. ex. art. 6 para. 1 GDPR.[16] In addition, there are extensive basic principles laid down in art. 5, which have an impact on the way personal data is processed. The question is whether these principles are compatible with Big Data and AI and with the expected expansion of the digital finance industry.[17] This is particularly important because, as will be shown below, the industry is convinced that more complete and comprehensive processing of personal data might enable it to provide better services to its customers.

This paper takes these developments as an opportunity to raise the question of an equitable balance between an innovative approach of FinTech companies on the one hand and the protection of the

8 Cf. *Schaffelhuber,* in: Kunschke/Schaffelhuber, FinTech: Grundlagen – Regulierung – Finanzierung – Case Studies, 2018, sect. I, chapter A, para. 1; *Sraders,* What Is Fintech? Uses and Examples in 2020, available online at: https://www.thestreet.com/technology/what-is-fintech-14885154, last accessed August 31, 2020. In this context, payment services such as PayPal have become well known.

9 Cf. *Dorfleitner/Hornuf,* loc. cit., p. 9. This is also supported by the fact that tech giants like Amazon, Apple, Facebook and Google have expanded their offerings to include digital financial services.

10 To be strictly distinguished from the processing of pure company data without reference to natural persons. The GDPR does not apply to this.

11 Cf. *Dorfleitner/Hornuf* loc.cit., p. 62.

12 Cf. *Basel Committee on Banking Supervision, Consultative Document, Sound Practices*, Implications of fintech developments for banks and bank supervisors, pp. 8 et seq, available online at: https://www.bis.org/bcbs/publ/d415.pdf, last accessed August 31, 2020.

13 Cf. for example *Paal/Hennemann,* NJW 2017, 1697. *Posner,* The right of privacy, Georgia Law Review 1978, 393, however arguing that privacy may have a negative impact on economic efficiency.

14 Cf. German Federal Constitutional Court, NJW 1984, 419, 421.

15 Recital 1 of the GDPR: "The protection of natural persons in relation to the processing of personal data is a fundamental right."

16 Hereafter, articles without legal reference are those of the GDPR.

17 Apart from the GDPR, the EU Payment Services Directive (PSD2) and the ePrivacy Directive also bear significant regulatory frameworks for FinTechs when processing personal data. This article shall however be limited to the GDPR as a critical regulation for the use of Big Data and AI.

interests of the data subjects on the other hand. The question is relevant because with the GDPR and the Big Data/AI applications, two elements that differ fundamentally in their core interests need to be reconciled, as outlined in section B and C. This is particularly true when deciding how to harmonize the maintenance of an accurate picture of customer needs with the interests of the parties concerned. While each approach may require compliance with the GDPR, it should ultimately create a position that leaves sufficient room for innovation and profitability. Section D argues that neither the preference of one side nor the other is a suitable solution for future development. Rather, the starting point must be a reduction in personal reference, assessed by the example of pseudonymization and anonymization.

2 The importance of Big Data & AI for FinTechs

The term FinTech is composed of the words "financial services" and "technology" and thus describes companies outside the traditional banking industry that use a technological approach to offer specialized services with regard to money or financial instruments.[18] The decisive factor for digitization is not just the further development of information technology as such, but the changed behavior of customers.[19] The terminology is used as a generic term to describe a payment system that does not rely on physical money and in which money is transferred electronically. It is a point-of-sale transaction system where an already identified customer can choose from different payment options to complete a transaction.

A very relevant field lies in B2C e-commerce.[20] This is not surprising. Most consumers buy goods through online marketplaces. Cashless payment systems enable instant transfers that accelerate transactions. Internet- or app-based service providers enable payment for goods or services in real time and thus a rapid implementation of the mutual performance obligations. There, the use of FinTechs is continuously increasing. In 2015, the share of online retail in overall retail was only 7 % – in 2019, the figure is already at 14 %.[21] This trend inevitably opens the door for new digital payment solutions, which is particularly important in the financial industry given the significantly lower amount transacted in cash.[22]

The terminology "FinTech" is often accompanied by the expectation that such companies will improve services in direct comparison to traditional credit institutions.[23] Although the link between

18 Cf. *BaFin*, Fintechs, in: Jahresbericht 2016 (available online at: https://www.bafin.de/DE/PublikationenDaten/Jahresbericht/Jahresbericht2016/Kapitel2/Kapitel2_5/Kapitel2_5_1/kapitel2_5_1_node.html, last accessed August 31, 2020; *Schaffelhuber,* in: Kunschke/Schaffelhuber, chapter A, para. 1 et seq.

19 Cf. *Dapp/Perthwieser,* in: Kunschke/Schaffelhuber, chapter B, para. 3.

20 Cf. *Negreiro,* European Parliamentary Research Service, The rise of e-commerce and the cashless society, 2020, p. 2.

21 Ibid.

22 For example, in 2018 cash transactions in Sweden accounted for only about 1 % of the GDP (vgl. *Riksbank*; Payments in Sweden 2019, available online at https://www.riksbank.se/en-gb/payments-cash/payments-in-sweden/payments-in-sweden-2019/the-payment-market-is-being-digitalised/, last accessed August 31, 2020). Although Sweden tends to be among the very advanced countries in the area of cashless payments, and cash transactions in the Euro countries average 11 % of GDP (ibid.), this should be seen as a first indication of a future trend in other regions as well.

23 Cf. *BaFin,* ibid.

financial services and technology is not new historically, the term FinTech is considered representative of the recently perceived interconnectedness of financial services and information technology.[24] In fact, the impact of these developments on the banking industry cannot be underestimated. Not only are FinTechs appropriating the business of traditional credit institutions, but traditional credit institutions, too, are adopting the very technologies that are used by FinTechs, which means that FinTechs can easily be considered "innovation drivers"[25].

What is indeed innovative – apart from the use of information technology – is the lack of an intermediary. With this so-called *cutting out the middleman* approach, peer-to-peer procedures are used instead of an intermediary, making the involvement of a credit institution redundant.[26] Ultimately, this stands out as a cost-saving measure. Although regulatory hurdles may be an important factor, disintermediation and the resulting cost savings are likely to be a driving force for FinTechs' progress in the financial sector.[27]

At the same time, the increased number of customers not only offers an increase in the processing of personal data, but also an opportunity to use that data in many ways to improve performance. Depending on the application system offered, the targeted customers for FinTech companies can also be natural persons. By referring to the personal data of the customers, FinTechs would be able to understand the customer's needs more efficiently and with less effort, and to expand or adapt their range of products and services based on this knowledge.

3 Methods of Big Data and AI

Why do FinTechs require large amounts of data? As with all start-ups entering the markets, competing against traditional institutions may be a lucrative long-term goal, but the first and utmost step is to build up a solid customer base. Once customers are attracted, it is key to develop a roadmap by which the company's goods and services can reach the targeted customers most efficiently. One way to accomplish this is to create customer profiles and screen the respective data. Linking that data can create valid insights, especially on open gaps or possible adjustment needs. The learnings produced by such processing operations touch one of the core interests of each enterprise: knowing and predicting your customers' needs. FinTechs, therefore, depend on a considerable amount of data that is not manageable manually by individual employees – hence the use of Big Data applications.

Big Data is a term for the structured and automated evaluation of data and data sets. It refers to quantities of data which, due to their size and complexity, cannot be evaluated using conventional methods of data analysis.[28] Accordingly, to the extent that there is a large volume and a wide variety

24 Cf. *Arner/Barberis/Buckley,* Georgetown Journal of International Law, 2016, v. 47 n. 4, p. 1345.

25 Cf. *Schaffelhuber,* loc.cit., p. 18.

26 Cf. *Basel Committee on Banking Supervision,* loc.cit., pp. 24 et seq.

27 Cf. *Lin,* Infinite Financial Intermediation, Wake Forest Law Review 2015, p. 643; *Schaffelhuber,* in: Kunschke/Schaffelhuber, chapter A, para. 10 et seq.; *Hopt,* in: Baumbach/Hopft, HGB, part II, sect. V (7) para. A/3a.

28 Cf. *Bachmann/Kemper/Gerzer,* Big Data – Fluch oder Segen?, 2014, p. 45 et seq.

of data as well as rapid processing speed (i. e. velocity) – which often is not possible when using conventional hardware and software –, all prerequisites for a classification as "Big Data" are fulfilled according to current opinion.[29] An essential feature is therefore the data mining of useful information on the basis of heterogeneous data sets. Historically, there have been many examples of mass evaluation of data.[30] The term however has only gained a broader public attention through the progressive digitization of modern industry. Today, almost all technical devices process, either specifically or *en passant*, a quantity of data that would not be manageable for an individual, both in relation to the device itself (functionalities, technical details) and in relation to the user (usage behavior, accesses, data transfers). From a technical point of view, Big Data evaluates data that has not yet been linked to any other data in order to create the needed insights.[31] This allows more reliable conclusions about the expected behavior of customers, which enables greater customer proximity as well as a better understanding of necessary improvements to the company's own services.[32]

AI does not differ much in its conceptual approach from the use of Big Data. In contrast to Big Data, AI only uses data that already exists in a system, especially since AI is largely bound to fixed data formats.[33] Moreover, AI does not necessarily use technologies with any ability to process large amounts of data in a very short time.[34] AI is a collection of technologies that combine data, algorithms, and computing power.[35] Technically, this is done by using algorithms in which each instruction or sequence of instructions solves a problem[36]; machine learning is used to determine and use certain patterns in data.[37] In sum, AI "revolutionized" the finance industry more to the extent that it improved precision levels, customer engagement and inquiry resolution periods. Instead of manual tasks and days of decision making, AI allows decisions to be made in seconds using all available knowledge.[38]

The growing volume of data as well as the technical possibility of efficient analysis and categorization of the various types of data offer opportunities for commercial utilization. One of these lies in fast and efficient insights on customer spending habits, client investments, market changes, and default risks. In other words, customer needs can be optimized, personalized, and correlated. This in turn can lead to innovation and quality competition. In this exact sense, the possible uses of Big Data and AI are manifold:

29 Cf. *Culik/Döpke,* ZD 2017, 226, 227 – the three "V's": volume, variety, and velocity.

30 Such as the "dragnet search" case prominent in Germany ("Rasterfahndung").

31 Cf. *Brandt,* in: Hauschka/Moosmayer/Lösler, Corporate Compliance, 2016, § 29 para. 164.

32 Ibid.

33 Cf. *Brandt,* loc. cit., para. 165.

34 Ibid.

35 Cf. *European Commission,* White Paper On Artificial Intelligence – A European approach to excellence and trust, COM(2020) 65 final, p. 2.

36 Cf. *Ory/Sorge,* NJW 2019, 710.

37 Ibid.

38 Cf. *Kebbel/Kaiser/Wassmer,* in: Kunschke/Schaffelhuber, loc. cit., p. 196.

- **Customer targeting** – One field of application is the so-called customer targeting, which covers a broad spectrum of customer needs. Goods or services are offered to customers in a targeted manner, whereby it can be seen in advance that those goods or services match an individual need.[39]
- **Personalizing** – In accordance with the interests pursued by the customer, certain offers and features can be displayed tailored to the customer's needs.[40]
- **Forecasting customer behavior** – Payment transaction data can make statements about the future behavior of customers.[41] This can represent a decisive competitive advantage. In addition to an extended use in customer targeting, it can also be used to predict, for example, imminent changes of customers to other providers.
- **Scoring** – The decision on an individual loan is made on the basis of a previous rating or scoring, which in turn is based on a comprehensive wealth of data.[42]
- **Profiling** – The term profiling refers to any type of automated processing of personal data for the purpose of evaluating, analyzing and predicting certain personal aspects, art. 4 no. 4 GDPR. Personal aspects related to the financial industry are in particular the economic situation of the individual, personal interests, reliability or usage-based behavior. With the help of profiling, automated evaluation processes can be created in which, by linking, evaluating and analyzing individual data, conclusions can be drawn about the data subject, thus also enabling forecasts.[43]
- **Tracking** – By means of tracking, the company undertakes the geographical tracing of a person or the tracking of communication behavior. As a rule, the aim is to create a profile of the customer that is as meaningful as possible.[44] Thus, FinTechs are enabled to approach their customers in a manner convenient to said customers, e. g. in accordance with the customer's usage patterns.

Irrespective of these versatile application possibilities, according to a survey conducted in Germany, only 22 percent of interviewed FinTech companies use Big Data analytics at all.[45] A business model originally based on Big Data is not evident among FinTech companies based in Germany.[46] One reason could be that in developed economies, aspects such as credit scoring do not rely on Big Data given the possibility to fall back on the information from the transactions made on a cus-

39 Tech giants like Amazon distribute their goods or services by using machine learning. By tracking customer behavior, e. g. by tracking visited pages or clicks, such a behavior pattern can be created in connection with the orders placed, according to which the respective customers can receive targeted offers (cf. *Camhi/Pandolph,* Business Insider, April 14, 2017, Machine learning driving innovation at Amazon, available online at: https://www.businessinsider.com/machine-learning-driving-innovation-at-amazon-2017-4?r=DE&IR=T, last ac-cessed August 31, 2020).

40 Cf. *Brandt,* loc. cit., para. 166.

41 Cf. *Brandt,* loc. cit., para. 164.

42 Cf. *Yan/Yu/Zhao,* How signaling and search costs affect information assymetry in P2P lending: The economics of big data, in: Financial Innovation, 1(1), 19e.

43 Cf. *Scholz,* Simitis/Hornung/Spiecker, Datenschutzrecht, art. 22, para. 22.

44 Cf. *Weichert,* ZD 2013, 255.

45 Cf. *Gimpel/Rau/Röglinger,* Wirtschaftsinformatik & Management, 2016(3), pp. 38 et seq.

46 Cf. *Dorfleitner/Hornuf,* loc.cit., p. 114.

tomer's personal account.[47] What is probably even more decisive, however, is that the strict data protection regulations and the resulting uncertainties in the assessment of compliance make possible Big Data applications appear largely unattractive in the member states of the E.U. from the outset. Most companies currently using big data applications in Germany use them primarily to improve their communication with customers rather than as a means of predicting customer behavior.[48] Yet, one can argue that the increasing use of these services shows a previously existing gap in the market.

4 Big Data and AI in the European data protection law

4.1 Structure and principles of the GDPR

The scope of application of the GDPR begins materially with the processing of personal data, art. 3 para. 1, which are defined in art. 4 no. 1 as all information relating to an identified or identifiable natural person. The range in which the GDPR applies is conceivably wide. A key driver for chasing compliance with the GDPR are its stipulated sanctions: Fines are turnover-related and can severely affect a FinTech company (e. g. two or four percent of the total worldwide annual turnover of the previous business year, art. 83 para. 4, 5).[49]

Actors of the GDPR are – as far as relevant for the subject of this study – the controller as the person responsible for the data processing on one side and the data subject on the other side. According to art. 4 no. 7, a controller is the body that decides on the purposes and means of processing personal data, i. e. the FinTech companies processing the data. The data subject is the person concerned by the processing, art. 1 no. 1.

Art. 5 explicitly names six principles by which the processing of personal data must be measured from a data protection perspective. The principles anchored in the GDPR are groundbreaking for the further understanding of the legal requirements for Big Data and AI.

4.1.1 Lawfulness, fairness, and transparency (art. 5 para. 1 lit. a)

The first mentioned principle is the principle of legality according to which any processing of personal data is prohibited unless there is a legal basis. Legal bases can be found in art. 6, which stipulates an exhaustive list.[50] The most frequently used legal basis is probably the consent of the data subject and the conclusion or execution of a contract. The legitimate interest as a "catch basin" mentioned in art. 6 para. 1 lit. f) also represents a frequently used basis in legal practice, but tends to appear unattractive due to the necessary comprehensive (and verifiable) balancing of interests in the fast-moving FinTech data traffic. The key takeaway of the principle of legality is therefore that a

47 Ibid.

48 Ibid.

49 Nevertheless, the turnover of FinTechs tends to be low compared to traditional banks, which can increase the probability of tolerating data protection risks.

50 Cf. *Albers/Veit*, in: BeckOK Datenschutzrecht, 32nd ed., art. 6 para. 15.

company processing personal data – *before* collecting the data – must determine a legal basis according to which the processing is carried out.

This is already the first major hurdle for the application of Big Data and AI. The underlying contractual relations between FinTech companies and customers always contains defined mutual performance obligations. A processing beyond this extent, i. e. both in terms of scope and duration, usually cannot be based on the initial contract (art. 6 para. 1 lit. b). Legal obligations (art. 6 para. 1 lit. c), such as laws or regulations that a responsible person must comply with, will most likely not provide for a basis for processing data on an oversized scale within the framework of Big Data and AI. Vital interests (art. 6 para. 1 lit. d) of the data subjects should not be affected in the FinTech sector, nor should the performance of tasks in the public interest (art. 6 para. 1 lit. e). In addition to the reasons mentioned above, recourse to a legitimate interest of the FinTech company (art. 6 para. 1 lit. f) tends to be problematic. The controller also has the burden of proof as to whether appropriate criteria for weighing up, such as reasonable expectations of the data subject or the foreseeability of extensive processing, have been taken into account.[51] Also in view of the right of objection laid down in art. 21 para. 1, stipulating a legitimate interest increases the amount of work in the argumentation.

The principle of fairness mentioned in art. 5 para. 1 lit. a) can be interpreted as the duty of the controller to take into account the interests of the data subjects.[52] In particular, data subjects must not be subject to any error as to what happens to their personal data and on what legal basis the processing takes place. This means that FinTech's must meet comprehensive information requirements. In addition, the fairness principle means that data subjects must not be confronted with obstacles once they want to assert their rights against the controller.[53] The data processing company must therefore be prepared to be able to react at any time to the exercise of data subjects' rights. Especially for big data users this means increased effort, because they have to fully explain the intended processing.

The transparency requirement, also laid down in art. 5 para. 1 lit. a), requires that data subjects be fully informed about the risks of processing their personal data. In particular, the information must be provided in a manner appropriate to the addressee.[54] As a result, art. 13, 14 list in detail the facts about which the controller must inform the data subject. This information is particularly important because, according to art. 13 para. 1, it must be fulfilled *before* or *at the time of the collection* of the data. From the point of view of the FinTech company, this effectively means a corresponding effort to provide adequate data protection notices including a determination of the intended use in advance.

51 Ibid., para. 52 et seq.

52 Cf. *Wolff*, in: Schantz/Wolff, Datenschutzrecht, 2017, para. 393.

53 Cf. *Schantz*, in: BeckOK Datenschutzrecht, art. 5, para. 8.

54 Cf. recital 39: "The principle of transparency requires that any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used."

4.1.2 Purpose limitation (art. 5 para. 1 lit. b)

Probably the most essential object of the GDPR is the so-called purpose limitation principle. According to art. 5 para. 1 lit. b), this principle states that personal data may only be processed for specified, clear and legitimate purposes on the one hand, and on the other hand that processing in a manner that is not in accordance with the initial purpose of collection is prohibited. As the controller has the power to define the purpose before starting its activities, one could argue that it has a certain freedom of action. Still, the controller needs to deal with this determination full-scope and upfront, which is why the purpose limitation principle is regarded as having an indication and warning function.[55]

The GDPR does not specify the required content of such purpose limitation. This may be rooted in the fact that it is up to a case-by-case examination to determine the circumstances, nature and scope of the data to be processed so that it can identify a certain degree of limitation. Given the meaning and purpose of the GDPR, adopting too general or consolidated approaches when determining the purpose obviously does not comply with this regulatory framework.

Processing data for additional, i. e. still arising, purposes is of paramount importance for the use of Big Data and AI. Yet, according to the GDPR, processing for purposes that are incompatible with the initial purpose of collection are non-compliant. Data controllers would often like to store the data once processed for a specific purpose for other purposes. Art. 6 para. 4 specifies in detail when the processing for additional purposes is compatible with the initial purpose of collection. This is the case either (i) if the data subject has consented, or (ii) if a legal provision of E.U. law or of the national law of a member state expressly permits this, in which case it must also be examined whether necessity and proportionality exist in accordance with the objectives mentioned in art. 23 para. 1.[56] The purpose of Big Data and AI, especially when combined, is to link together a large amount of personal data from different collection purposes in order to draw new conclusions. However, since, according to the aforementioned explanations, processing of personal data beyond the initial purpose of collection results in the alteration of said purpose, difficulties regularly remain in reconciling Big Data and AI with the originally communicated collection purposes. Moreover, there will generally be major hurdles to define the purposes sufficiently broadly to enable the desired use of Big Data and AI on the one hand and to make them sufficiently concrete to meet the transparency requirement on the other. It is precisely the non-existent predictability of possible further processing purposes that is one of the key aspects of Big Data. In order to be able to evaluate large amounts of data efficiently whilst not abandoning personal data, a controller will have to pre-define all purposes in a detailed and precise manner, make them transparent to the data subject, and obtain consent (see section III.1).

55 Cf. *Härting,* NJW 2015, 3284, 3286; *von Grafenstein,* DuD 2015, 789, 792.

56 Further processing for archiving purposes in the public interest, for scientific or historical research purposes or for statistical purposes, i. e. for non-commercial purposes, is also permissible, but not further relevant to the object of investigation.

4.1.3 Data minimization (art. 5 para. 1 lit. c)

The principle of data minimization mirrors the principle of proportionality common in German law, which is generally tested in three steps: (i) suitability – (ii) being limited to what is necessary – (iii) being adequate.[57] In the case of (i) suitability, the question arises as to whether the data processing is capable of achieving the initial purpose at all.[58] Big Data and AI should undoubtedly have the technical means to achieve any purpose and meet this requirement. In the case of (ii) being limited to what is necessary, the processing must take place in the most limited way that is still sufficient enough to achieve the purpose.[59] This means that the processing must not "overdo" beyond what is necessary. Big Data and AI will do so in any case if a less invasive alternative exists. FinTechs must therefore critically examine whether the purpose pursued in each case can be achieved in a less invasive manner, which in view of the almost unmanageable volume of data at Big Data and AI is likely to be fraught with major hurdles. Specifically, they would have to carry out a suitability test for each individual processing operation. The effort involved in this is likely to regularly cancel out the expected benefits.

The condition to (iii) be adequate requires an assessment of whether the data processing is proportionate.[60] In the proportionality test, the conflicting rights of the parties involved are weighed against each other in order to determine whose rights prevail in light of all the circumstances of the individual case. It can be assumed that the data processing is to be measured strictly against standards interpreted in favor of the data subject, since, as mentioned at the beginning, the GDPR already prohibits all processing whereas legal bases are considered exceptions, arg. ex art. 6 para. 1. The proportionality test can even make a processing operation that would usually be covered by a legal basis unlawful, for example if the scope of the processing is excessive from an objective point of view.[61] Assuming that Big Data is also used to collect data in case it is needed for any cases of usage that may arise, one will usually conclude that its use is not compatible with the adequacy requirement. Under the aforementioned conditions, the processing of personal data for hypothetical purposes is not allowed. This makes sense as the continuous referencing of data sets to the data subject can be far too indeterminate and unpredictable. Here, too, FinTechs will have to critically examine whether the data processing takes adequate account of data subjects' interests.

4.1.4 Accuracy (art. 5 para. 1 lit. d)

The principle of the accuracy of the processed data, which has been a defining feature of data protection law since the census ruling[62] of the German Federal Constitutional Court, requires the data controller to ensure that the data is accurate, always kept up to date and that appropriate

57 Cf. *Schantz,* loc.cit., para. 24 et seq.

58 Ibid.

59 Ibid.

60 Cf. *Frenzel,* in: Paal/Pauly, DS-GVO, 2018, art. 5, para. 35.

61 Cf. *Roßnagel,* in: Simitis/Hornung/Spiecker gen. Döhmann, Datenschutzrecht, 2019, art. 5, para. 119.

62 At that time, the Federal Constitutional Court already warned against personality images on which the data subject can only exert influence under the most difficult restrictions (cf. German Federal Constitutional Court, NJW 1984, 419, 421).

measures are taken to delete or correct incorrect data immediately.[63] This principle is of great importance for the use of AI and self-learning systems. It must be ensured that not only facts, but also value judgments do not start from faulty premises or are based on faulty conclusions. Accordingly, when using AI, the FinTech company has to ensure that decisions about customers are not based on inappropriate or unrepresentative data that can lead to false or discriminatory results.[64]

4.1.5 Storage limitation (art. 5 para. 1 lit. e)

The principle of storage limitation requires the implementation of appropriate retention periods. According to recital 39, sentence 8, the storage period must be limited to the absolutely necessary minimum. This is mirrored by the *right to be forgotten*, according to which a data subject can assume that its data will be deleted by the controller once the reason for collection has ceased to exist. The data controller must therefore determine the duration of the storage in advance and inform the data subject of the duration in accordance with art. 13, 14. If the consent of the data subject stipulates the legal basis, the principle of storage limitation means that the controller must regularly check the validity of the consent.

4.1.6 Integrity and confidentiality (art. 5 para. 1 lit. f)

Art. 5 para. 1 lit. f) refers to the integrity of the data, i. e. ensuring that the data does not get completely or partially deleted, destroyed or altered without authorization.[65] This requires an appropriate level of security through technical and organizational measures. According to recital 39, sentence 12, this includes ensuring that unauthorized persons have no access to the data or to the equipment with which the personal data is processed.[66] Thus, the FinTech company must be able to demonstrate an appropriate concept.

4.2 Consent according to GDPR

In view of the prospect of a lack of legal bases or incompatibility with the principles of data protection law, controllers are prone to use a data subject's consent as a basis of legitimacy in the hope being enabled to comprehensively process data, including possible changes of purpose.[67] This, however, is also faced with hurdles. Art. 7 requires:

- **Voluntary nature** – The person giving consent must act on the basis of a free decision, which must be an expression of individual self-determination.[68] The data subject must have a free choice and the option to refuse consent without fear of immediate disadvantages. Given that

63 Cf. ECJ, NVwZ 2009, 379.

64 Cf. *Data Protection Commission,* Hambacher Erklärung zur Künstlichen Intelligenz vom 03. April 2019, pp. 3 et seq.

65 Cf. *Frenzel,* loc.cit., para. 47.

66 Cf. European Court of Justice, NJW 2014, 2169.

67 Cf. recital 32 – a single consent may cover several processing purposes.

68 Cf. *Stemmer,* in: BeckOK Datenschutzrecht, art. 7, para. 37.

consent must not be tied to the principal services as set forth in art. 7 para. 4, the question will therefore arise as to whether the provision of consent is voluntary if the data subject or the customer assumes that it will not be able to obtain the desired main service of the FinTech company without granting such consent. One could argue that the data processing that is (also) in the interest of the data subject may stipulate voluntariness. The assessment of such interest is subject to a case-by-case examination and likely to be highly questionable in the context of Big Data and AI, since from the perspective of the data subject's horizon it is not the synergies generated for the controller that are decisive, but an actually improved service for the customer/data subject itself. This is ultimately for the FinTech to determine and prove.

- **Obtaining in advance** – Consent must have been obtained before the start of data processing. The responsible party must always determine or obtain the necessary legal basis before starting the processing activities.
- **Informedness** – As laid out, in accordance with the transparency requirement, the consenting party must be fully informed about the purpose and scope of the data processing. With regard to art. 7 para. 2, according to which special requirements exist when consent is combined for different situations, the requirement of being informed affects the need to provide a complete and exhaustive description of the desired data processing purposes in advance. The data subject must understand the meaning and scope of the decision in order to give an informed consent. This is a difficult balancing act for a FinTech. On the one hand, if the described purpose is extensive or difficult to understand, it may be questionable whether the data subject is sufficiently informed. Whether complex facts can be sufficiently perceived and understood by the data subject remains to be examined on a case-by-case basis. On the other hand, the Big Data/AI user might quickly reach its limits, since at the time of the first data collection it is not foreseeable for which purpose the combined data will be used.
- **Provability** – The existence of the consent must remain provable.
- **Visibility** – The consent must be emphasized separately to the data subject.

An equally high-maintenance aspect is the possibility of revocation according to art. 7 para. 3. Data subjects have the right to revoke a given consent at any time. Upon revocation, the legality of the data processing ceases *ex nunc*. Although this does not affect the lawfulness of the processing based on the consent until revocation, the FinTech must in any case stop the processing at the time of revocation. Besides the above reasons, this makes basing the processing on consent less attractive.

In the area of profiling and scoring, there is an additional regulatory requirement. According to art. 4 no. 4, profiling is any automated processing of personal data consisting in the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects relating to personal interest, reliability or behavior. Scoring is the determination of a score value in relation to a person, with the help of which, for example, credit applications are automatically accepted or rejected.[69] What both applications have in common is that the produced decisions are completely automated, i. e. created by the exclusive use of technology.

According to art. 22 para. 1, the data subject has the right to not be subject to a decision based solely on automated processing that produces legal effects or in a similar way constitutes a substantial

69 Cf. *Martini,* in: Paal/Pauly, art. 22, para. 24.

impairment. In fact, the law requires the use of appropriate mathematical or statistical procedures and appropriate technical and organizational measures to minimize the risks of errors or incorrect data. Effectively, this means that no legally impairing decisions are permitted when made without the intervention of a human being. Art. 22 para. 2 only provides for exceptions if the automated decision is necessary for the conclusion or performance of a contract, if an E.U. or national legal provision permits this or if there is an explicit consent of the data subject. The controller must critically weigh the concept of necessity according to art. 22 para. 2 lit. a) or obtain the concrete consent of the data subject, whereby the above statements on consent apply.[70] Moreover, the possibility of objection under art. 21 para. 1 is an additional aggravating obstacle.

4.3 Result

The use of Big Data and AI by FinTechs faces significant data protection hurdles. As shown, conflicts arise with all principles on which the GDPR is based. Recourse to the consent of the data subject does not help significantly. Not only does it bear strict requirements. A consent does not provide sufficient legal security for FinTechs since there is always the risk of revocation at any time. Art. 22 shows that European data protection law is generally skeptical about profiling and scoring for commercial purposes.

5 Pseudonymization and anonymization as mitigation?

The needs of the modern and digitized financial industry on the one hand and the European data protection law on the other show how different the interests under the guise of digital innovation can be. Is it necessary to allow for extensive interaction between FinTechs and their customers in order to fully understand any customer needs, or should the Big Data and AI approaches be completely abandoned on the territory of the EU/EEA due to widespread data protection concerns?

Neither the one nor the other extreme should take over the sole rule within this spectrum. As demonstrated, it is not possible for Big Data and AI to be applied to the processing of personal data without any restrictions whatsoever, solely on the grounds that they provide economic efficiency. Informational self-determination is – with good reasons – regarded as an overriding individual good. Nevertheless, the European Union should not lag behind in global competition for the most effective use of Big Data and AI in the financial market sector.[71] In a next step, it must therefore be clarified whether these conflicting perspectives can be reconciled. This is certainly worth striving for if the European Union's claim to become a global leader in digital innovation is based on consistent compliance with the GDPR.

70 E.U. or national regulations are not apparent for the commercial use of profiling and scoring.

71 Various approaches and position papers of the European Commission can be interpreted under this leitmotif, according to which the European Union needs to become a global leader in innovation in the data economy and its applications (cf. *European Commission,* White Paper On Artificial Intelligence – A European approach to excellence and trust, loc.cit.; Communication from the European Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data, COM(2020) 66 final; Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the safety and liability implications of Artificial Intelligence, the Internet of things and robotics, COM(2020) 64 final). Moreover, the President of the European Commission Ursula von der Leyen announced in her political guidelines a coordinated European approach on the better use of big data for innovation, availble online at: https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf, last accessed August 31, 2020.

As shown, the different disruptive applications have different conflicts with the GDPR: The biggest conflict for Big Data and AI is the existing personal reference due to an avalanche of legal consequences detrimental to FinTechs. Therefore, the question arises whether a reduction of the personal reference, this by means of pseudonymization or anonymization, allows the application of those technologies within the current legal framework.

5.1 The use of pseudonymization within Big Data and AI

Pseudonymization and anonymization are named several times in the GDPR. Already in recital 28, the use of pseudonymization is praised as a possibility that can reduce the risks for data subjects and support the controller in complying with its data protection obligations. The legislative will is thus evident – an increase in the use of pseudonymization is expressly desired. According to art. 4 no. 5, pseudonymization is the processing of personal data in such a manner that the personal data can no longer be assigned to a specific data subject without the need for additional information. This specifically requires separate storage of additional information as well as technical and organizational measures to ensure that the personal data is not assigned to an identified or identifiable natural person.

Despite these apparent advantages due to the reduction in the personal reference, there has been no significant increase in the use of pseudonymization at FinTechs since the GDPR came into force. Irrespective of the doubling of use cases observed in an analysis, the use of pseudonymization remains at a tendentially low level.[72] This is surprising given the possibility of reducing identifying features by separating data functionally. The information value of personal data is separated from its identity. Furthermore, additional information is required in order to link the information value with the identity. This in turn is only done under authorized conditions.[73] The data is therefore only "personal" for the person who actually has the additional knowledge, e. g. the assignment key. For those who do not have it, the data usually does not stipulate personal data at all. Identifying features can be replaced by other data, for example, whereby the references are stored in a separate manner and the original data stock remains unidentifiable. It is also conceivable to use hash values.[74]

The use of pseudonymization is supported by a whole range of data protection reasons:

- In light of privacy by design, pseudonymization represents the state of the art.[75] It can enforce dynamic protection of both direct and indirect identifiers. This is essential, since re-identification within Big Data is technically trivial and can be enabled at any time.
- Art. 25 para. 1 showcases pseudonymization as an appropriate technical and organizational measure in the context of Privacy by Design. Using pseudonyms, therefore, fulfills the requirement to provide suitable and appropriate technical and organizational measures. This also allows for greater flexibility when faced with data subjects' rights, since suitable

72 Cf. *Dorfleitner/Hornuf,* loc.cit., p. 81.

73 This principle can also be found in the CCPA as "de-identification."

74 Cf. *Schwartmann/Weiß,* Whitepaper zur Pseodonymisierung, 2017, p. 17 et seq.

75 Cf. recital 78.

technical and organizational measures can support the examination of a balance of interests.

- When examining the legality of changing the purpose, art. 6 para. 4 emphasizes pseudonymization as a suitable guarantee of the compatibility of the secondary purpose with the original processing purpose. Pseudonymization can represent a remediation measure for the purpose limitation requirement.
- Art. 32 et seq. emphasize that pseudonymization is a security measure that makes it seem unlikely that the rights and freedoms of natural persons can be violated. As a consequence, pseudonymization can reduce liability scenarios or obligations to report any incidences to the data protection authorities since the risks of data breaches can be significantly lower.

The use of pseudonyms can therefore resolve conflicts with the outlined basic principles of the GDPR. Problems of legality or purpose limitation do not arise if identification is largely prevented. In particular, the essential principle of data minimization is maintained. Pseudonymized data is regularly sufficient to achieve the desired purposes since the reference of the data to an individual person does not have to be essential.

The controller may not escape an ongoing check of the correctness, memory limitation and integrity. This, however, does not carry too much weight in view of the fact that pseudonymization can make it possible to fulfill the objectives of Big Data and AI to a large extent. In particular, linking data to create insights on open gaps or necessary adjustments often may not necessarily require any personal reference in order to predict customer needs. A FinTech can still generate learnings with aggregated data. Aggregation combines data to a group data set in such a manner that it is ultimately no longer possible to determine to whom individual data can be assigned within the data collective. This certainly offers protection against unintentional re-identification. For example, usage behavior could also be categorized into relevant customer groups, broken down by specific parameters such as age, income, or occupation. In light of the restrictions posed by the GDPR, the aspiration of FinTechs to generate comprehensive insights in order to better understand customer needs can be met by means of pseudonymization, at least better than without it. What is even more significant: They do not have to bear the risk of violating the regulation.

5.2 The use of anonymization within Big Data and AI

Anonymization changes personal data in such a way that the person behind the individual details can no longer be identified.[76] In contrast to pseudonymization, there is no separately stored information that enables identification. Although the content of a data set is retained, it no longer allows for assignment to a specific or identifiable person.[77] Since this is not a personal data, the GDPR does not apply at all. Similar as in pseudonymization, anonymization may produce insights by using aggregated data without any personal reference. This may work for insights on spending habits or forecasts based on a general comparative group. Anonymization may however not work for personalizing, scoring, and profiling matters since the complete abolishment of personal references could contradict these particular processing activities from the outset. Here, pseudonymization appears more as the workaround.

76 Cf. *Ernst*, in: Paal/Pauly, loc.cit., art. 4, para. 48.

77 Ibid.

More importantly, it is key to ensure that working with anonymous data is regularly checked to ensure that additional knowledge gained by linking different data does not allow a re-identification of the original anonymous data. Even if only one identification feature can be found, anonymization is no longer existent.[78] For the FinTech company, as the controller, this means that the prevention of re-identification in a mass data set must be ensured. This is likely to be difficult. Technically, it is relatively simple to re-identify a person since Big Data involves huge data sets that makes the re-identification of a person unavoidable.[79] Therefore, pseudonymization appears as a more suitable means as re-identification is only possible through separately stored additional information. It also ensures that a regulatory framework exists that expressly praises pseudonymization as a means of gaining compliance, as explained above. Should FinTechs have overcome the aforementioned hurdles within anonymization and ensured the prevention of re-identification – which always requires a case-by-case examination – the scope of application of the GDPR as well as the conflicts shown can be circumvented.

6 Summary

The technical capabilities of Big Data and AI allow significant benefits for FinTechs. They can determine customer needs more efficiently and less prone to errors. This can prove to be a decisive competitive advantage. Particularly in view of the European Union's recent efforts to become a global leader in the field of digitization, further discussion of Big Data and AI cannot be avoided.

That said, the processing of personal data in the context of Big Data and AI is made considerably more difficult by the GDPR. The right to informational self-determination has a superior rank, against which extensive processing, some of which even undefined by purpose, cannot be justified.

Pseudonymization and anonymization can enable customer insights. As far as the personal reference is reduced or completely excluded, the GDPR does not turn out to be a contrary opponent. The GDPR suggests pseudonymization in particular as a suitable means in many cases. However, it is predominantly the pseudonymization efforts that have the potential to provide an effective means for FinTechs to apply Big Data and AI while still complying with European data protection law. Anonymization in a Big Data context may still bear the risk of continuous re-identification.

78 It could however stipulate a pseudonymization.

79 Cf. Art.-29-Datenschutzgruppe, WP 216, Stellungnahme 5/2014 zu Anonymisierungstechniken, p. 9 et seq.