COMMERZBANK

A risk-based approach to Artificial Intelligence

15 February 2023

Commerzbank's White Paper on the European Artificial Intelligence Act

A. Abstract

Artificial Intelligence (AI) and Machine Learning (ML) are vital for the future of the banking industry to overcome the various challenges of the digital age and the additional (e.g. cyber-) risks arising from it. Implementing AI and ML in the highly regulated banking industry is complex. Here, a risk-based approach helps balancing the benefits of this technology against the regulatory hurdles surrounding it.

With this in mind, the paper at hand aims to demystify AI and ML for what it really is: controllable and non-magical! Such models are still trained and operated by humans.

B. Applications of AI in Banking

Customers nowadays expect good product recommendations, simple processes as well as fast response times and transactions. Banks are facing a disruption of their business model with a high level of digitalization and new competitors from FinTechs to BigTechs. At the same time, high and rising regulatory burdens and the low margin environment intensify the pressure on banks' profitability. Hence, the use of innovative solutions and smart forms of automation is crucial to succeed.

The availability of data, enhanced computing power and new methods to create insights from data can tremendously improve, accelerate and automate business processes even for tasks which until recently required human intelligence.

AI and ML offer novel opportunities from increasing customer satisfaction to helping banks with:

- lowering costs via increased automation and more efficient processes (e.g. document classification, automatic extraction of data in document processing, Talkbots and Chatbots),
- preventing or reducing losses from credit risks, fraud and cyber risks (like coordinated attacks) or optimizing capital allocation and steering via enhanced risk quantification,
- marketing activities and to increase revenues via targeted recommendations (e.g. Next-best-offer).

In addition, the information age has enabled criminals to upgrade their approaches to financial crime in general and fraud, terrorist financing, money laundering and cyber-attacks in particular. Only with AI and ML we can appropriately counteract in order to defend our industry as well as our society by establishing an equality of means ("weapons").

Commerzbank has established the division "Big Data & Advanced Analytics" as a Center of Competence for all AI and ML related matters which range from in-house AI modelling and implementation to advisory tasks. The Center of Competence also pays special attention to ML Governance and Trustworthy and Responsible AI. In light of the upcoming regulation (cf. section C) this white paper will outline Commerzbank's risk-based approach to ML governance.

C. Current Regulatory Environment

Entities with access to big data and cloud infrastructures have the best conditions to accelerate AI and ML. China, the United States and Europe are currently leaders in research and development of AI systems. The strength of Europe currently emerges out of joint research programs and other initiatives bringing together the decentralized actors in this field and supporting the participation in the development of open-access AI and ML models. Furthermore, with its high data protection standards Europe has set the groundwork to create an environment of trust. Trustworthiness is and will be important to enable the uptake of AI and to support the society.

At the moment, the European Union (EU) is in the middle of the legislative process to establish harmonized rules

COMMERZBANK

Definition according to AI Act: AI System

Artificial Intelligence System

"'Artificial intelligence system' (AI system) means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts;" (Proposal of the EU AI Act as per 6th Dec 2022, Article 3)

"A system that uses rules defined solely by natural persons to automatically execute operations should not be considered an AI system" (Proposal EU AI Act as per 6th Dec 2022, p.6 (6))

For a detailed definition of Machine Learning and Logic- and Knowledge Based Approaches please refer to p.6-7, 6a and 6b.

on AI, hereinafter: "AI Act"[1]. It aims to ensure smooth functioning of markets while at the same time meeting a high level of protection of public interests, such as health, safety, fundamental rights and Union values. The AI Act aims to regulate the AI components of IT systems regardless of the economic sector they are used in. It shall support and foster investment and innovation in AI and ultimately lead to safe, trustworthy and ethical AI adoption. The regulation is pioneering worldwide and will provide a quality seal for trustworthy AI made and used in Europe.

In early December 2022 the Council of the European Union adopted its common position[2]. Although the German government generally supports the Council's position, it still sees some need for improvement on certain aspects[3]. Once the European Parliament adopts its own position the 'trilogues' between European Council, Parliament and Commission can be entered. As of today, this is expected for the second quarter of 2023.

D. Definitions

Traditional programming codifies rules (e.g. if/then conditions). It freezes the modelled reality once and for all. Instead, ML aims to identify ("learn") these rules itself which can be too complex for traditional programming. ML models represent statistical input-output-relationships and they are very well suited to describe complex and non-linear relationships, which enables them to ideally generalize also to new, formerly unseen data. Often, their result provides probabilities for different possible outputs, which also require acceptance thresholds in order to utilize the output (e.g. to make a "yes or no"-decision).

Today efficient AI almost always relies on ML. AI involves the creation of an algorithm that uses data to model some aspects of the world. Model training is often based on data labelled by humans. Afterwards, the model is applied to new data to generate output such as content,

[1] All further refences are based on the General approach of the Council of the European Union as of 6th December 2022 "Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts"

[2] Council of the European Union: Press Release as of 6th Dec 2022

[3] Statement by Germany on the Proposal for an EU AI Act, 25th November 2022,; Remarks by Germany on the Proposal for an EU AI Act, 8th November 2022,

predictions, classifications, or recommendations with the goal to augment and support (human) activities, or ultimately also take certain decisions.

Our definition of AI is in alignment with the definition of an AI system according to the AI Act (see box above). Based on this definition we assume that ad-hoc analysis using ML techniques is out-of-scope of the AI Act.

E. Risk-Based Approach

AI has the potential to exceed capabilities of traditional models by far. It establishes a different level of modelling activities due to higher speed of processing large data sets and the ability for decision making etc. Hence, AI innovation leads to shifts in risk and return trade-offs, and successful AI implementations must always take the costs of risk mitigation into account. Among these are transparency vis-à-vis users and consistent performance of the systems (i.e. accuracy and robustness as well as cybersecurity).

The precise trade-offs in the choice of models and their implementation depend on the specifics of the situation:

- A high level of data quality is a prerequisite for robust development and implementation of traditional and ML models alike. Increasing the amount and spectrum of available data generally improves model results.
- In some cases, simpler data driven approaches or "classical" statistical methods can provide predictive power similar to more advanced AI models, but without some of the associated risks. Nevertheless, describing complex, non-linear relationships usually requires ML methods to be used.
- By entering a symbiosis between human and machine for labelling tasks, feedback loops and decision making, the efficiency and effectiveness of a process can be increased while at the same time risks can be mitigated.

Commerzbank pursues a risk-based approach. Systems containing AI are categorized into AI Risk Classes depending on a variety of factors such as business continuity, business criticality and model complexity.

In banking, modelling has been playing an essential role for decades and has been at the heart of many business processes and supporting activities. Many modelling activities are crucial for our business to succeed and do not affect any fundamental rights of persons. Additionally, ML methods vary significantly in their complexity and can be static (up to new releases) or based on continuous learning. Hence, Commerzbank's risk-based approach covers model complexity, impact range and business criticality, resulting in the following risk categories:

Class 1: Prohibited AI Systems[4]

Class 2: High-Risk AI Systems (according to AI Act Annex III **or** internally classified bank criticality)

Class 3: AI Systems with Transparency Obligations

Class 4: Low Risk AI Systems

Class 5: Ad-Hoc Analysis

Class 6: No AI

We appreciate that the proposal of the AI Act also lays down a risk-based approach and defines a risk methodology for high-risk AI systems (as per Article 6 and Annex III) that pose significant risks to health and safety or fundamental rights of persons. In order to address and mitigate these risks appropriately, high-risk AI systems (according to the AI Act) will have to comply with a set of horizontal mandatory requirements for trustworthy AI

[4] Prohibited is using Artificial Intelligence for manipulative, exploitative and social control practices. Further definitions and details can be found in Article 5 of the AI Act.

(Title III; chapter 2) and need to follow a conformity assessment procedure (Article 43) before they can be placed on the EU market. Moreover, they will be obliged to be registered in an EU database for high-risk AI Systems (Article 51).

Amongst others, the following requirements exist for high-risk AI systems that are listed according to Annex III of the AI Act:

- Risk Management System (Article 9),
- Data and data governance (Article 10),
- Technical documentation (Article 11),
- Record-keeping (Article 12),
- Transparency and provisions of information to users (Article 13),
- Human oversight (Article 14),
- Accuracy, robustness, and cybersecurity (Article 15).

The risk classification process of the AI Act needs to be fit and proper across different industries. Therefore, the methodology of classifying high-risk AI systems according to the AI Act needs to be simple and easy to understand. The Act provides a classification in form of a list of high-risk applications and hence results in a digital "yes or no"-decision.

Creditworthiness

The following banking function is labelled as a high-risk AI System according to subparagraph 5(b) of Annex III of the European Council's proposal:

"AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by providers that are micro and small-sized enterprises as defined in the Annex of Commission Recommendation 2003/361/EC for their own use"

While the intention is clear, i.e. preserving the people's ability to have access to loans, the expression "evaluate the creditworthiness" leaves room for interpretation.

During a customer's lifecycle we might process information related to a person's creditworthiness (such as lifecycle applications like information gathering or classification tools). These cases are often small, innovative ways to improve efficiency and effectiveness (leading to e.g. enhanced response times and lowered costs), but do not affect fundamental rights. Of course, these cases need to be validated according to their potential risks. Treating them on the same footing as high-risk AI Systems with the time and cost effort associated, would significantly reduce efficiency, innovation, and digitalization in the banking sector in general.

It is not this paper's intention to lower the standards, but to argue for a realistic, effective, and adequate implementation. We assume that this is about situations where there is specifically a "yes or no"-loan-granting-decision involved as one of the aims of the AI Act is to protect customers' ability to get a loan. Hence, we come to the conclusion that lifecycle applications are exactly what is remedied for with the exemption of AI components that have a "purely accessory" character (Article 6).

Machine Learning Governance

Risk management at Commerzbank follows the principle of "three lines of defence":

- Each unit (segments and functions) forms the first line of defence according to its operative responsibility and is directly responsible for identifying and managing risks in its own management area, while complying with the specified risk standards and policies.
- The second line of defence for each type of risk lays down standards for appropriate risk management procedures, monitors and ensures the application of such standards, and analyses and evaluates the risks.
- The third line of defence is carried out by internal audit.

This structure also applies to the management of risks arising from the implementation, deployment, and use of

AI. Most of the risks related to the use of AI models and systems in financial services are not new and are well-known from successfully handling traditional models in the past. Hence, we lever on existing knowledge and structures within the bank. Model Validation as second line of defence is supported if needed by various functions like Cyber Risk, Compliance and Operational Risk etc.

F. Certificates

Certificates and CE markings should increase trust in the solutions offered and consequently support and foster innovation and AI investments especially for small and medium-sized companies. This concept allows companies to buy certified third-party software ensuring the AI system's trustworthiness and fulfillment of requirements as laid down in the AI Act. While the intention is clearly in support of innovation, the concept has limitations:

- In general, the question arises as to what *exactly* is certified. The certification can only ever be a point-in-time snapshot or backwards oriented. It may not hold true beyond a certain time horizon in the future. A certification might adopt two different approaches:
  - 1.) Test of Design: Examination whether controls are designed properly and able to mitigate a defined risk.
  - 2.) Test of Effectiveness: Examination whether controls have been in place over a certain time period and worked effectively as required.
- There can be changes in the data over time or different results can occur in different situations. These and other circumstances could fundamentally change the explanatory power and up-to-dateness of the certification of an AI system. Hence, it is reasonable to certify the whole model lifecycle process incl. development, evaluation, deployment, monitoring and new model deployments etc. Yet a complex certificate cannot easily be understood by humans. Only a human understandable certificate is effective and can establish trust as intended.
- Third-party software certification only ever testifies on how well the model works in the designed environment. In other environments and with different data the model might work less efficient or not as intended. Consequently, the certification could be mistaken as a "carte blanche" to use the model regardless of the individual situation.
- As of today, AI inspection catalogues cannot be regarded as final and stable in the future. Therefore, certificates cannot be easily compared over time.
- Moreover, in the banking context, we need to ensure that our third-party providers fulfill the same high standards as we do. Responsibility can never be outsourced. It is questionable to what extent banks can rely on this testimony or will need to perform these kinds of evaluations themselves. Hence, it must be determined how banks can use certifications. Being able to rely on these is especially important for efficiently implementing state-of-the-art general purpose AI.

In summary, certificates might not yield the intended benefits but can quickly become costly and create redundant burdens for banks. It is important to note that credit scoring models are already audited by competent authorities which should be regarded as a certificate of high quality itself. Consequently, a harmonized standardization should highlight that this ongoing supervision fully meets certification requirements.

G. Trustworthy and Responsible AI

Compared to the situation a decade ago, today's computing power and rich data sources allow ML to extend its range of application, introducing the possibility to intrude deeply into people's sensitive and private areas of life. Hence ethical considerations need to be taken into account. Taking Environmental Social Governance (ESG)

seriously also means that paying special attention to the trustworthiness and responsibility of one's AI activities is an absolute must.

There are many aspects to consider when defining Trustworthy and Responsible AI. Most of these concepts are already well known and standard procedure when bringing software into production or processing data in general (cf. Prudential Requirements for IT (BAIT), General Data Protection Regulation (GDPR) etc.). Here, we want to focus on those three aspects with the most vivid debate in public with most room for interpretation: Transparency, explainability and fairness.

Transparency

Transparency is about being clear, open, and honest about how and why a person's data is being used. (cf. Article 13). In addition, information requirements on automated processing are laid out in Articles 13 and 14 of the GDPR. The data subject shall "have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." (Article 22 para 1 GDPR). A similar intention can be found in the AI Act which states that natural persons must be informed that they are interacting with an AI system (Article 52).

Explainability / Interpretability

The explainability of ML models addresses aspects ranging from the interpretability of input-output relations to the precise inner workings of models in mathematical terms. In their application, it essentially amounts to justifying the process from data and model selection to model validation and monitoring, thereby ensuring the control over its intended use.

The right amount and form of explanation cannot be determined without specifying the addressee and context. One could imagine publishing all AI Systems' source codes, however hardly any consumer will have the capacity nor training data to retrace these. Instead, the general public and specifically consumers need to receive the appropriate amount of information in a comprehensible manner allowing them to scrutinize decisions that have been made. On the other hand, for example for the purpose of audits, material needs to be complete and kept for certain time periods.

Valid concerns speak against disclosing detailed information on AI systems. Especially in the context of fighting fraud disclosing details about the tools used to find fraudulent activities might help circumventing them. This is obviously not a desirable result. Furthermore, business know-how has often been developed in-house with large investments in both time and money. Making such confidential business logic public can lead to significant losses of intellectual property and results into disadvantages over competitors.

Fairness

Ethics of AI and bias are among the most discussed topics around trustworthy AI in public and the use of this terminology varies largely.

Bias arising from algorithms describes a situation where there is a wrongful discriminatory judgement encoded into an algorithm. Since ML models learn patterns from past data, they "learn" these biases that prevail in the data through correlation. This bears the risk that the prevailing biases get even intensified through automation and usage.

Discrimination in this context means that someone is "demeaned, or treated as having lesser moral worth, in virtue of a socially salient characteristic"[5]. When we talk about fairness, we mean principles of "justice" where the most prominent principle is "equality" in a sense that discrimination is prevented, monitored and mitigated.

[5] Vredenburgh, K. (2022), 'Fairness' (p.3), in: Justin B. Bullock and others (eds), The Oxford Handbook of AI Governance (online edn, Oxford Academic, 14 Feb. 2022), https://doi.org/10.1093/oxfordhb/9780197579329.013.8 , accessed 10 Jan. 2023.

COMMERZBANK

Discrimination can be manifold and regardless of the designers' intentions it can enter the algorithm at various points in the modelling phase, like[6]:

- Choosing or formulating the task itself
- Systematic differences between a population sample vs. the whole population due to data selection or inaccurate data gathering
- Historic human judgment made on prejudices or established conventions
- Bias might be introduced through continuous learning where original training data was not biased but newly gathered training data is
- Manual thresholds used to convert predictions into decisions can lead to disparate treatments
- The application environment at model deployment can introduce bias

Fairness needs to be carefully considered and monitored in all stages of the model lifecycle with appropriate internal governance in place. It is also important to note that there can be different but similarly reasonable fairness measures where it is impossible to fulfill both at the same time. In these cases a detailed and context-oriented consideration has to be made.

Moreover, it is important to note that even when a potentially discriminating feature is not recorded or later deleted from the dataset, it is still mathematically possible that the model result is dependent on this feature. This is the case when the discriminating feature is correlated to the output. Hence, the model might still discriminate around this feature (e.g. Amazon Hiring Tool[7]). So, by just deleting the discriminating feature there is no chance to find and try to remove discrimination from the decision as long as correlation to other features exist.

It is important to point out that robust and trustworthy AI systems can also help to overcome subconscious biases human decision makers might have as it yields repeatable and traceable results.

At the same time differentiation is one of the banks oldest tasks: For risk purposes differentiation around statistically significant features such as income is necessary to estimate the customer's ability to repay debt. Banks take in deposits and use them to some extent to grant loans. Based on a large portfolio single actual credit defaults can be compensated. This form of differentiation protects the individual customers, the bank's long-term viability and subsequently financial market stability in general.

AI and ML model creation is an iterative process. Of course, transparency and fairness should be a pre-requisite. Nevertheless, lots of corresponding validation steps can only be completed during or after the modelling process. To ask for final fairness validations from the start without blocking innovation from the beginning is thus neither feasible nor meaningful.

H. Relationship with other laws and jurisdictions

From a regulatory point of view the development, training, evaluation, and deployment of AI systems needs to adhere to various requirements in addition to the upcoming AI Act. Among these are:

- Directive 2013/36/EU on access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms
- EU General Data Protection Regulation (GDPR)
- European Data Act

[6] Cf. Vredenburgh, Kate, 'Fairness' (p.3-5) cf. [7], and Zweig, K. (2019) "Ein Algorithmus hat kein Taktgefühl" p.208-220, München: Wilhelm Heyne Verlag

[7] IIF Bias and ethical implications in machine learning (p. 10)

- European Data Governance Act
- Supervisory Requirements for IT in Financial Institutions (BAIT)
- Minimum Requirements for Risk Management (MaRisk[8])
- EBA Guidelines on outsourcing arrangements
- Digital Operational Resilience Act (DORA)
- EBA Guidelines on ICT and security risk management
- Markets in Financial Instruments Directive, esp. relevant for algorithmic trading
- At least indirectly through many more regulation like laws on consumer protection, equality, anti-discrimination etc.

Other jurisdictions in the world are about to adopt AI-related regulation with different angles and priorities. Furthermore, some jurisdictions argue that most AI topics are already covered sufficiently by existing regulation and might only require to be detailed accordingly. These developments pose the risk of market fragmentation and put challenges to multi-national organizations in general.

I. Conclusion

We appreciate the general orientation the AI Act provides – an approach to a quality seal of AI made or used in Europe. Definitions are state-of-the-art and a risk-based approach is suitable. Since the Act shall apply for all various industries some banking specifics are omitted. This is especially important when considering the fine line between lifecycle customer management and credit scoring as well as the drawbacks additional certifications have in an already externally audited field of expertise like credit scoring. Here it is important to consider the ongoing model supervision by the competent authorities as equivalent to certifications. Additionally, the terms around AI Ethics like transparency, explainability and fairness are often misunderstood in public. Since the AI Act does not provide sufficient detail on how this shall be addressed from a practical perspective, we provided an idea of what it means and how it needs to be embedded in the banking context.

It is not this paper's intention to lower the standards, but to point to the already existing and effective regulation in the banking industry and to argue for a realistic, effective, and adequate implementation of the new requirements. Due to the entanglements of Artificial Intelligence with various other regulations as described above we call for a coherent harmonization of rules.

[8] Minimum Requirements for Risk Management (MaRisk) in the version of 16.08.2021 and Konsultation 06/2022 - Entwurf der MaRisk in the version of 26.09.2022 (German only)

Commerzbank AG
Zentrale
Kaiserplatz
Frankfurt am Main
www.commerzbank.de

Postanschrift
60261 Frankfurt am Main
Tel. + 49 69 136-20
info@commerzbank.com

Big Data & Advanced Analytics

Julia Sterling
Julia.sterling@commerzbank.com

Thomas Stadje
Thomas.stadje@commerzbank.com