COMMERZBANK

Ein risikobasierter Ansatz zu Künstlicher Intelligenz

15 Februar 2023

White Paper der Commerzbank zur europäischen KI-Verordnung (AI Act)

COMMERZBANK

A. Einführung

Künstliche Intelligenz (KI) und Maschinelles Lernen (ML) sind für die Zukunft der Bankenbranche von entscheidender Bedeutung, um den vielfältigen Herausforderungen des digitalen Zeitalters und den daraus entstehenden zusätzlichen (Cyber-) Risiken zu begegnen. Die Implementierung von KI und ML in der stark regulierten Bankenbranche ist jedoch komplex. Deswegen braucht es einen risikobasierten Ansatz, der die Vorteile dieser Technologien mit den verbundenen regulatorischen Hürden abwägt.

Vor diesem Hintergrund zielt das vorliegende White Paper darauf ab, KI und ML zu entmystifizieren, auf das was sie wirklich sind: Kontrollierbar und nicht-magisch! Solche Modelle werden von Menschen trainiert und betrieben.

B. Anwendungen von KI im Bankwesen

Kunden erwarten heute gute Produktempfehlungen, einfache Prozesse sowie schnelle Reaktionszeiten und Transaktionen. Die Bankenbranche befindet sich im Umbruch durch die Notwendigkeit einer effizienten Digitalisierung der Prozesse sowie das Auftreten von neuen Wettbewerbern von FinTechs bis hin zu BigTechs. Gleichzeitig verstärken hohe und steigende regulatorische Anforderungen und ein Umfeld mit niedrigen Margen den Druck auf die Rentabilität. Daher ist der Einsatz innovativer Lösungen und intelligenter Automatisierungsformen zunehmend entscheidend für den langfristigen Erfolg.

Die Verfügbarkeit von Daten, die verbesserte Rechenleistung und neue Methoden, um Erkenntnisse aus diesen Daten zu gewinnen, können Geschäftsprozesse enorm verbessern, beschleunigen und automatisieren. Das gilt auch für Aufgaben, die bis vor kurzem menschliche Intelligenz erforderten.

AI und ML bieten zahlreiche neue Möglichkeiten, von der Steigerung der Kundenzufriedenheit bis hin zur Unterstützung der Banken bei:

- der Kostensenkung durch erhöhte Automatisierung und effizientere Prozesse (z. B. Dokumentenklassifizierung, automatische Extraktion von Daten in der Dokumentenverarbeitung, Talk- und Chatbots),
- der Vermeidung oder Reduzierung von Verlusten durch Kreditrisiken, Betrug und Cyberrisiken (wie z.B. koordinierte Angriffe) oder der Optimierung der Kapitalallokation und Steuerung durch verbesserte Risikoquantifizierung,
- Marketingaktivitäten und zielgerichtete Empfehlungen (z. B. Next-Best-Offer).

Das Informationszeitalter ermöglicht es Kriminellen ihre Aktivitäten rund um Finanzkriminalität im Allgemeinen und Betrug, Terrorismusfinanzierung, Geldwäsche und Cyber-Angriffe im Speziellen zu verbessern. Nur mit KI und ML können wir dieser Entwicklung angemessen entgegentreten, um unsere Industrie und unsere Gesellschaft durch die Schaffung einer Gleichheit der Mittel („Waffen“) zu verteidigen.

Die Commerzbank hat den Geschäftsbereich „Big Data & Advanced Analytics“ als Kompetenzzentrum für alle KI- und ML-bezogenen Themen etabliert. Dessen Aufgaben reichen von der internen KI-Modellierung und -Implementierung bis hin zu Beratungsaufgaben. Ein besonderes Augenmerk legt das Center of Competence auch auf ML Governance sowie vertrauenswürdige und verantwortungsvolle KI. In Anbetracht der bevorstehenden Regulierung (vgl. Abschnitt C) soll dieses White Paper den risikobasierten Ansatz der Commerzbank für ML-Governance erläutern.

C. Aktuelles regulatorisches Umfeld

Unternehmen mit Zugriff auf Big Data und Cloud-Infrastrukturen verfügen derzeit über die besten Voraussetzungen, KI und ML zu entwickeln. China, die Vereinigten Staaten und Europa sind derzeit führend in der Forschung und Entwicklung von KI-Systemen. Die Stärke Europas ergibt sich dabei aus gemeinsamen Forschungsprogrammen und anderen Initiativen, die die

dezentralen Akteure in diesem Bereich zusammenbringen und die Beteiligung an der Entwicklung von Open-Access-KI- und ML-Modellen unterstützen. Darüber hinaus hat Europa mit seinen hohen Datenschutzstandards den Grundstein für ein Umfeld des Vertrauens gelegt. Vertrauenswürdigkeit ist und wird auch zukünftig wichtig sein, um die Akzeptanz von KI zu ermöglichen und die Gesellschaft in diesem Prozess zu unterstützen.

Derzeit befindet sich die Europäische Union (EU) mitten im Gesetzgebungsprozess zur Festlegung harmonisierter Regeln für die Verwendung und Entwicklung von KI, im Folgenden „AI Act“[1] genannt. Dieser soll ein reibungsloses Funktionieren der Märkte gewährleisten und gleichzeitig ein hohes Schutzniveau für öffentliche Interessen wie Gesundheit, Sicherheit, Grundrechte und Werte der Union gewährleisten. Der AI Act zielt darauf ab, die KI-Komponenten von IT-Systemen unabhängig vom Wirtschaftszweig, in dem sie eingesetzt werden, zu regulieren. Weiterhin soll er Investitionen und Innovationen in KI unterstützen und fördern, um damit letztlich zu einer sicheren, vertrauenswürdigen und ethischen KI-Einführung führen. Die Verordnung ist weltweit wegweisend und wird ein Gütesiegel für vertrauenswürdige KI, die in Europa hergestellt und verwendet wird, etablieren.

Anfang Dezember 2022 nahm der Rat der Europäischen Union seinen „General Approach“[2] an. Obwohl die deutsche Regierung den Standpunkt des Rates generell unterstützt, sieht sie nach wie vor in einigen Punkten Verbesserungsbedarf[3]. Sobald das Europäische Parlament seine eigene Position eingenommen hat, kann der „Trilog“ zwischen dem Europäischen Rat, dem Parlament und der Kommission aufgenommen werden. Dies wird Stand heute für das zweite Quartal 2023 erwartet.

D. Definitionen

Traditionelle Programmierung kodifiziert Regeln (z.B. Wenn-/Dann-Bedingungen). Sie friert die modellierte Realität ein. Demgegenüber zielt ML darauf ab, diese Regeln selbst zu identifizieren (zu „lernen“), die für die traditionelle Programmierung zu komplex sein können. ML-Modelle stellen statistische Input-Output-Beziehungen dar und eignen sich sehr gut zur Beschreibung komplexer und nicht-linearer Beziehungen, was es ihnen ermöglicht, auch auf neue Daten oder Veränderungen der Realität zu reagieren. Oft bietet das Ergebnis Wahrscheinlichkeiten für verschiedene mögliche Outputs, die Akzeptanzschwellen erfordern, um den jeweiligen Output zu nutzen (z.B. um eine „Ja oder Nein“-Entscheidung zu treffen).

Effiziente KI fußt heute fast immer auf ML. KI beinhaltet die Erstellung eines Algorithmus, der Daten verwendet, um bestimmte Aspekte der Welt zu modellieren. Das Modelltraining basiert häufig auf Daten, die von Menschen gekennzeichnet werden. Anschließend wird das Modell auf neue Daten angewendet, um Ergebnisse in Form von z.B. Inhalten, Vorhersagen, Klassifizierungen oder Empfehlungen zu generieren. Ziel ist dabei menschliche Aktivitäten zu unterstützen oder letztlich auch bestimmte Entscheidungen zu treffen.

Unsere Definition von KI entspricht der Definition eines KI-Systems gemäß AI Act (siehe Kasten oben). Auf der Grundlage dieser Definition verstehen wir, dass Ad-hoc-Analysen mit ML-Techniken außerhalb des Geltungsbereichs des AI Acts liegen.

[1] Alle weiteren Referenzen basieren auf dem „General Approach“ des Rates der Europäischen Union vom 6. Dezember 2022 „Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Regeln für künstliche Intelligenz (Artificial Intelligence Act) und zur Änderung bestimmter Rechtsakte der Union“.

[2] Rat der Europäischen Union: Pressemitteilung vom 6. Dezember 2022.

[3] Stellungnahme Deutschlands zum Vorschlag des AI Acts, 25. November 2022; Stellungnahme Deutschlands zum Vorschlag des AI Acts, 8. November 2022.

COMMERZBANK

Definition gemäß AI Act: KI System

System der künstlichen Intelligenz

„System der künstlichen Intelligenz“ (KI-System) ein System, das so konzipiert ist, dass es mit Elementen der Autonomie arbeitet, und das auf der Grundlage maschineller und/oder vom Menschen erzeugter Daten und Eingaben durch maschinelles Lernen und/oder logik- und wissensgestützte Konzepte ableitet, wie eine Reihe von Zielen erreicht wird, und systemgenerierte Ergebnisse wie Inhalte (generative KI-Systeme), Vorhersagen, Empfehlungen oder Entscheidungen hervorbringt, die das Umfeld beeinflussen, mit dem die KI-Systeme interagieren; (Vorschlag des AI Act per 6. Dezember 2022, Artikel 3)

„Ein System, das ausschließlich von natürlichen Personen definierte Regeln anwendet, um automatisch Operationen auszuführen, sollte nicht als KI-System gelten.“ (Vorschlag AI Act vom 6. Dezember 2022, S.6 (6))

Detaillierte Definitionen von maschinellem Lernen und der logik- und wissensgestützten Konzepte findet sich auf S. 6-7, 6a und 6b.

E. Risikobasierter Ansatz

KI hat das Potenzial, die Fähigkeiten traditioneller Modelle bei weitem zu übertreffen. Es ist eine andere Liga von Modellen, u.a. weil nun große Datensätze schneller verarbeitet und Entscheidungen automatisiert getroffen werden können. KI-Innovationen führen daher zu Verschiebungen bei Risiko- und Return-Trade-Offs. Erfolgreiche KI-Implementierung muss immer die Kosten der Risikomitigation berücksichtigen. Dazu gehören Transparenz gegenüber den Nutzern und eine konsistente Performance der Systeme (z. B. Genauigkeit, Robustheit sowie Cybersicherheit).

Die genauen Abwägungen bei der Auswahl der Modelle und deren Umsetzung sind situationsspezifisch:

- Eine hohe Datenqualität ist gleichermaßen Voraussetzung für eine robuste Entwicklung und Implementierung von traditionellen und ML-Modellen. Die Erhöhung der Menge und des Spektrums der verfügbaren Daten verbessert im Allgemeinen die Modellergebnisse.
- In einigen Fällen können einfachere datengetriebene Ansätze oder „klassische“ statistische Methoden eine Vorhersagekraft liefern, die den fortgeschritteneren KI-Modellen gleicht, jedoch ohne einige der damit verbundenen Risiken. Dennoch erfordert die Beschreibung komplexer, nicht-linearer Beziehungen in der Regel die Nutzung von ML-Methoden.
- Durch die Symbiose von Mensch und Maschine bei Labelling, Feedbackschleifen und Entscheidungsfindung können die Effizienz und Effektivität eines Prozesses erhöht und gleichzeitig die Risiken gemindert werden.

Die Commerzbank verfolgt einen risikobasierten Ansatz. Systeme, die KI enthalten, werden in KI-Risikoklassen eingeteilt, abhängig von einer Vielzahl von Faktoren wie Modellkomplexität, Wirkung und geschäftlicher Kritikalität.

Im Bankwesen spielt die Modellierung seit Jahrzehnten eine wesentliche Rolle. Sie ist das Herzstück zahlreicher Geschäftsprozesse und unterstützender Aktivitäten. Viele Modellierungsaktivitäten sind für den Erfolg unseres Unternehmens von entscheidender Bedeutung und haben keinen Einfluss auf die Grundrechte betroffener Personen. Darüber hinaus variieren ML-Methoden erheblich in ihrer Komplexität und können statisch sein (bis zum Deployment neuer Versionen) oder auf kontinuierlichem Lernen basieren. Daher deckt der risikobasierte

Ansatz der Commerzbank die Komplexität des Modells, dessen Wirkung und Kritikalität ab. Dies führt zu den folgenden Risikokategorien:

Klasse 1: Verbotene Praktiken[4]

Klasse 2: Hochrisiko-KI-Systeme (gemäß AI Act Annex III **oder** intern klassifizierter Bankkritikalität)

Klasse 3: KI-Systeme mit Transparenzverpflichtungen

Klasse 4: Low-Risk KI-Systeme

Klasse 5: Ad-Hoc-Analyse

Klasse 6: Keine KI

Der Vorschlag des AI Acts sieht ebenfalls einen risikobasierten Ansatz vor. Dessen Methodik klassifiziert Hochrisiko-KI-Systeme (gemäß Artikel 6 und Anhang III), die erhebliche Risiken für die Gesundheit und Sicherheit oder die Grundrechte von Personen darstellen. Um diese Risiken angemessen abzubilden und mitigieren, müssen für Hochrisiko-KI-Systeme (gemäß AI Act) eine Reihe horizontaler verbindlicher Anforderungen für vertrauenswürdige KI erfüllt werden (Titel III; Kapitel 2). Außerdem müssen sie eine Konformitätsbewertung (Artikel 43) vollziehen, bevor sie in der EU auf den Markt gebracht werden können. Darüber hinaus sind sie in einer EU-Datenbank für Hochrisiko-KI-Systeme (Artikel 51) zu registrieren.

Für Hochrisiko-KI-Systeme, die nach Anhang III des AI Act gelistet sind, bestehen unter anderem folgende Anforderungen:

- Risikomanagementsystem (Artikel 9),
- Daten und Daten-Governance (Artikel 10),
- Technische Dokumentation (Artikel 11),
- Aufzeichnungspflichten (Artikel 12),
- Transparenz und Bereitstellung von Informationen für die Nutzer (Artikel 13),
- Menschliche Aufsicht (Artikel 14),
- Genauigkeit, Robustheit und Cybersicherheit (Artikel 15).

Der Risikoklassifizierungsprozess des AI Acts ist branchenübergreifend. Daher muss die Methodik zur Klassifizierung von Hochrisiko-KI-Systemen nach dem AI Act einfach und leicht verständlich sein. Das Gesetz sieht eine Klassifizierung in Form einer Liste von risikoreichen Anwendungen vor und führt somit zu einer digitalen „Ja oder Nein“-Entscheidung.

Kreditwürdigkeit

Die folgende Bankfunktion ist gemäß Absatz 5 (b) des Anhangs III des Vorschlags des Europäischen Rates als ein Hochrisiko-KI-System gekennzeichnet:

„KI-Systeme, die bestimmungsgemäß für die Kreditwürdigkeitsprüfung oder Kreditpunktebewertung natürlicher Personen verwendet werden sollen, mit Ausnahme von KI-Systemen, die von Anbietern, die Kleinstunternehmen oder kleine Unternehmen im Sinne der Begriffsbestimmung im Anhang der Empfehlung 2003/361/EG der Kommission sind, für den Eigengebrauch in Betrieb genommen werden.“

Während die Absicht eindeutig ist, nämlich die Fähigkeit der Menschen zu gewährleisten, Zugang zu Krediten zu erhalten, lässt die Formulierung insbesondere im englischen Text („evaluate the creditworthiness“) Raum für Interpretation.

Während der gesamten Dauer einer Kundenbeziehung kann es dazu kommen, dass wir möglicherweise Informationen verarbeiten, die mit der Bonität einer Person zusammenhängen (sogenannte Lifecycle-Anwendungen, z. B. Tools zur Informationsbereitstellung oder

[4] Verboten ist der Einsatz von Künstlicher Intelligenz für manipulative, ausbeuterische und soziale Kontrollpraktiken. Weitere Definitionen und Details finden sich in Artikel 5 des AI Acts.

Klassifizierung). Diese Fälle sind oft kleine, innovative Wege zur Verbesserung der Effizienz und Effektivität (z.B. bei verbesserten Reaktionszeiten und Kosten), wirken sich aber nicht auf die Grundrechte aus. Natürlich müssen diese Fälle entsprechend ihren potenziellen Risiken validiert werden. Würde man sie allerdings mit dem damit verbundenen Zeit- und Kostenaufwand wie ein Hochrisiko-KI-System behandeln, hätte das erhebliche Auswirkungen auf die Effizienz, Innovation und Digitalisierung im Bankensektor.

Es ist nicht die Absicht dieses White Papers, die Standards zu senken, sondern für eine realistische, effektive und angemessene Umsetzung zu werben. Eines der Ziele des AI Acts besteht darin, die Fähigkeit der Kunden, einen Kredit zu erhalten, zu wahren. Wir halten es daher für selbstverständlich, dass es hier ausschließlich um Situationen geht, in denen es sich um eine „Ja oder Nein"-Kreditentscheidung handelt. Daher kommen wir zu dem Schluss, dass Lifecycle-Anwendungen genau das sind, was mit der Klausel zu „völlig unwesentlichen" KI-Komponenten erreicht werden soll (Artikel 6).

Machine Learning Governance

Das Risikomanagement der Commerzbank folgt dem Prinzip der „Three Lines of Defense":

- Jede Einheit (Segmente und Funktionen) bildet die 1st Line of Defense gemäß ihrer operativen Verantwortung und ist direkt für die Identifizierung und das Management von Risiken in ihrem eigenen Managementbereich verantwortlich. Dabei müssen die angegebenen Risikostandards und -richtlinien eingehalten werden.
- Die 2nd Line of Defense legt für jede Art von Risiko Standards für geeignete Risikomanagementverfahren fest, überwacht und gewährleistet die Anwendung dieser Standards und analysiert und bewertet diese Risiken.
- Die 3rd Line of Defense wird von der internen Revision ausgeführt.

Diese Struktur gilt auch für das Management von Risiken, die sich aus der Implementierung, dem Einsatz und der Nutzung von KI ergeben. Die meisten Risiken im Zusammenhang mit dem Einsatz von KI-Modellen und -Systemen in Finanzdienstleistungen sind nicht neu und sind aus dem erfolgreichen Umgang mit traditionellen Modellen in der Vergangenheit bekannt. Wir bauen daher auf vorhandenes Wissen und bestehende Strukturen innerhalb der Bank auf. Die Modellvalidierung als 2nd Line of Defense wird bei Bedarf durch verschiedene Funktionen wie Cyber Risk, Compliance und Operational Risk usw. unterstützt.

F. Zertifikate

Zertifikate und CE-Kennzeichnungen sollen das Vertrauen in die angebotenen Lösungen stärken und damit Innovationen und Investitionen insbesondere für kleine und mittlere Unternehmen fördern. Dieses Konzept ermöglicht es Unternehmen, zertifizierte Drittanbietersoftware zu kaufen, welches die Vertrauenswürdigkeit des KI-Systems und die Erfüllung der Anforderungen des AI Acts sicherstellt. Zwar ist die Absicht klar, Innovation zu unterstützen, doch das Konzept hat Grenzen:

- Zunächst stellt sich die Frage, was *genau* zertifiziert wird. Eine Zertifizierung kann immer nur eine Momentaufnahme oder vergangenheitsorientiert sein. Sie wird in Zukunft nicht über einen bestimmten Zeithorizont hinaus Stand halten. Eine Zertifizierung kann zwei unterschiedliche Ansätze verfolgen:
  - Test of Design: Prüfung, ob Kontrollen ordnungsgemäß gestaltet sind und ein definiertes Risiko mindern können.
  - Test of Effectiveness: Prüfung, ob Kontrollen über einen bestimmten Zeitraum durchgeführt wurden und wie erforderlich effektiv funktioniert haben.
- Es kann im Laufe der Zeit zu Änderungen der Daten oder in verschiedenen Umgebungen des Einsatzes zu unterschiedlichen Ergebnissen

kommen. Diese und andere Umstände könnten die Erklärungskraft und Aktualität der Zertifizierung eines KI-Systems grundlegend verändern. Daher ist es sinnvoll, den gesamten Modell-Lebenszyklus-Prozess inklusive Entwicklung, Evaluierung, Deployment, Monitoring und dem Deployment neuer Modelle zu zertifizieren. Doch ein zu komplexes Zertifikat kann von Menschen nicht leicht verstanden werden. Nur ein verständliches Zertifikat ist wirksam und kann wie beabsichtigt Vertrauen erzeugen.

- Die Zertifizierung von Drittanbietersoftware belegt immer nur, wie gut das Modell in der entworfenen Umgebung funktioniert. In anderen Umgebungen und mit unterschiedlichen Daten funktioniert das Modell möglicherweise weniger effizient oder nicht wie beabsichtigt. Folglich könnte die Zertifizierung als „Carte Blanche" missverstanden werden, um dieses Modell unabhängig von der individuellen Situation zu nutzen.
- Stand heute können KI-Inspektionskataloge nicht als endgültige Versionen verstanden werden. Zertifikate können daher nicht einfach im Laufe der Zeit verglichen werden.
- Außerdem muss im Bankenkontext sichergestellt sein, dass Drittanbieter die gleichen hohen Standards erfüllen wie Banken selbst. Verantwortung kann niemals ausgelagert werden. Es ist daher fraglich, inwieweit sich Banken auf ein Zertifikat verlassen können oder diese Art von Bewertungen selbst durchführen müssen. Folglich muss festgelegt werden, wie Banken Zertifikate nutzen können. Dies ist insbesondere für die effiziente Nutzung von „General Purpose AI" wesentlich.

Zusammenfassend lässt sich sagen, dass Zertifikate möglicherweise nicht den beabsichtigten Nutzen bringen. Vielmehr könnten sie aber schnell kostspielig werden und redundante Belastungen für Banken verursachen. Es ist hier wichtig zu beachten, dass die Modelle der Bonitätsbewertung bereits von den zuständigen Behörden geprüft werden, was selbst als ein Zertifikat von hoher Qualität angesehen werden sollte. Eine harmonisierte Standardisierung sollte berücksichtigen, dass die bestehende Aufsicht die Zertifizierungsanforderungen vollständig erfüllt.

G. Vertrauenswürdige und verantwortungsvolle KI

Im Vergleich zu der Situation vor zehn Jahren ermöglichen die heutige Rechenleistung und die umfangreichen Datenquellen, dass ML sein Anwendungsspektrum erweitert. Dies schafft die Möglichkeit, tief in die sensiblen und privaten Lebensbereiche von Menschen einzudringen. Daher müssen auch ethische Überlegungen berücksichtigt werden. Environmental Social Governance (ESG) ernst zu nehmen bedeutet auch, dass eine besondere Aufmerksamkeit auf die Vertrauenswürdigkeit und Verantwortung der eigenen KI-Aktivitäten ein absolutes Muss ist.

Es gibt viele Aspekte, die bei der Definition vertrauenswürdiger und verantwortungsvoller KI zu berücksichtigen sind. Die meisten dieser Konzepte sind nicht neu, sondern bereits Standardverfahren bei der Einführung von Software in den produktiven IT-Betrieb oder der Verarbeitung von Daten im Allgemeinen (vgl. BAIT, DSGVO etc.). Hier wollen wir uns auf die drei Aspekte mit der lebhaftesten öffentlichen Debatte und dem größten Interpretationsspielraum konzentrieren: Transparenz, Erklärbarkeit und Fairness.

Transparenz

Transparenz bedeutet, klar, offen und ehrlich darüber zu kommunizieren, wie und warum die Daten einer Person verwendet werden (vgl. Artikel 13). Darüber hinaus sind die Informationspflichten für die automatisierte Verarbeitung in den Artikeln 13 und 14 DSGVO festgelegt. „Die betroffene Person hat das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt" (Art.

22 Abs. 1 DSGVO). Eine ähnliche Absicht findet sich im AI Act, die besagt, dass natürliche Personen darüber informiert werden müssen, dass sie mit einem KI-System interagieren (Artikel 52).

Erklärbarkeit / Interpretierbarkeit

Die Erklärbarkeit von ML-Modellen befasst sich mit Aspekten, die von der Interpretierbarkeit von Input-Output-Beziehungen bis hin zu dem präzisen „Inner Working" von Modellen in mathematischer Hinsicht reichen. In ihrer Anwendung geht es im Wesentlichen darum, den Prozess von der Daten- und Modellauswahl bis hin zur Modellvalidierung und -überwachung zu rechtfertigen. So wird die Kontrolle über den beabsichtigten Verwendungszweck sichergestellt.

Der richtige Umfang und die richtige Form der Erklärung können nicht ohne Angabe von Adressaten und Kontext ermittelt werden. Man könnte sich vorstellen, alle Quellcodes von KI-Systemen zu veröffentlichen, aber kaum ein Verbraucher wird über die Fähigkeit und die Trainingsdaten verfügen, um diese nachzuverfolgen. Stattdessen müssen die breite Öffentlichkeit und insbesondere Verbraucher eine ausreichende Menge an verständlichen Informationen erhalten, damit sie getroffene Entscheidungen überprüfen können. Zum anderen muss das Material beispielsweise für externe Prüfungen (Audits) vollständig sein und über bestimmte Zeiträume aufbewahrt werden.

Valide Bedenken sprechen gegen die Offenlegung detaillierter Informationen über KI-Systeme. Insbesondere im Zusammenhang mit der Betrugsbekämpfung könnte die Offenlegung von Details über die Instrumente zur Ermittlung betrügerischer Aktivitäten dazu beitragen, diese zu umgehen. Das ist natürlich kein wünschenswertes Resultat. Darüber hinaus wurde Know-how oft intern entwickelt und mit hohen Investitionen in Zeit und Geld. Die Veröffentlichung dieser Informationen kann daher zu einem erheblichen Verlust an geistigem Eigentum und vertraulichen Geschäftslogiken sowie zur Entstehung von Nachteilen gegenüber Wettbewerbern führen.

Fairness

KI-Ethik und „Bias" gehören zu den am häufigsten diskutierten Themen rund um vertrauenswürdige KI in der Öffentlichkeit, doch die Verwendung dieser Terminologie variiert.

„Bias" (Voreingenommenheit) von Algorithmen, beschreibt eine Situation, in der ein in einem Algorithmus kodiertes ungerechtfertigtes und diskriminierendes Urteil erfolgt. Da ML-Modelle Muster aus vergangenen Daten lernen, „lernen" sie diese Vorurteile, die in der Quelle vorherrschen mit. Dies birgt das Risiko, dass die bestehenden Vorurteile durch Automatisierung und Algorithmennutzung sogar noch verstärkt werden.

Diskriminierung bedeutet in diesem Zusammenhang, dass jemand „aufgrund einer sozial hervorstechenden Eigenschaft erniedrigt oder als moralisch minderwertiger behandelt wird"[5]. Wenn wir von Fairness sprechen, meinen wir Prinzipien der Gerechtigkeit. Das prominenteste Prinzip, die Gleichheit, fordert, dass Diskriminierung verhindert, überwacht und gemildert wird.

Diskriminierung kann vielfältig sein und unabhängig von den Absichten der Designer an verschiedenen Stellen in der Modellierung in den Algorithmus kommen, z. B.[6]:

- Bei der Auswahl der Aufgabe selbst oder ihrer Formulierung
- Systematische Unterschiede zwischen einer Bevölkerungsstichprobe und der gesamten Bevölkerung aufgrund von Datenauswahl oder ungenauer Datenerfassung

[5] Vredenburgh, K. (2022), 'Fairness' (S.3), in: Justin B. Bullock and Others (eds), The Oxford Handbook of AI Governance (online edn, Oxford Academic, 14. Februar 2022), https://doi.org/10.1093/oxfordhb/9780197579329.013.8, abgerufen am 10. Januar 2023.

[6] Siehe Vredenburgh, Kate, 'Fairness' (S. 3-5) vgl. [7,] und Zweig, K. (2019) „ein Algorithmus hat kein Taktgefühl" S.208-220, München: Wilhelm Heyne Verlag.

COMMERZBANK

- Historische, menschliche Urteile wie Vorurteile oder etablierte Konventionen
- Bias könnte durch kontinuierliches Lernen eingeführt werden, bei dem die ursprünglichen Trainingsdaten nicht voreingenommen waren, aber neu erfasste Trainingsdaten es sind
- Manuelle Schwellenwerte, die zur Umwandlung von Vorhersagen in Entscheidungen verwendet werden, können zu unterschiedlichen Behandlungen führen
- Die Anwendungsumgebung bei der Modellbereitstellung kann zu Bias führen

Dies zeigt, dass Fairness in allen Phasen des Modelllebenszyklus mit entsprechender interner Governance sorgfältig überwacht werden muss. Dabei ist auch wichtig zu beachten, dass es unterschiedliche, aber gleichzeitig relevante Fairness-Maßnahmen geben kann, bei denen es unmöglich ist, beide gleichzeitig umzusetzen. In diesen Fällen ist eine detaillierte und kontextorientierte Betrachtung erforderlich.

Darüber hinaus ist es wichtig zu beachten, dass selbst wenn ein potenziell unterscheidendes Merkmal nicht aufgezeichnet oder später aus dem Datensatz gelöscht wird, es mathematisch immer noch möglich ist, dass das Modellergebnis von diesem Merkmal abhängt. Dies ist der Fall, wenn die unterscheidende Funktion mit der Ausgabe korreliert ist. Daher könnte das Modell immer noch dieses Merkmal (z. B. Amazon Hiring Tool[7]) unterscheiden. Allein durch das Löschen des diskriminierenden Merkmals besteht also keine Chance, Diskriminierung zu entdecken und aufzulösen, solange eine Korrelation zu anderen Merkmalen besteht.

Es ist weiterhin wichtig darauf hinzuweisen, dass robuste und vertrauenswürdige KI-Systeme auf der anderen Seite auch dazu beitragen können, unbewusste Vorurteile menschlicher Entscheidungsträger zu überwinden, da sie wiederholbare und nachvollziehbare Ergebnisse liefern.

Differenzierung ist eine der ältesten Aufgaben der Banken: Für Risikozwecke ist eine Differenzierung um statistisch signifikante Merkmale wie Einkommen notwendig, um die Fähigkeit des Kunden seine Schulden zurückzuzahlen, abzuschätzen. Die Banken nehmen Einlagen ein und verwenden sie bis zu einem gewissen Grad, um Kredite zu gewähren. Basierend auf einem großen Portfolio können einzelne tatsächliche Kreditausfälle kompensiert werden. Diese Form der Differenzierung schützt die einzelnen Kunden, die langfristige Rentabilität der Bank und damit die Stabilität des Finanzmarktes im Allgemeinen.

Die Erstellung von AI- und ML-Modellen ist ein iterativer Prozess. Natürlich sind hier Transparenz und Fairness eine wichtige Voraussetzung. Dennoch können viele entsprechende Validierungsschritte nur während oder nach dem Modellierungsprozess durchgeführt werden. Von Beginn an eine endgültige Fairness-Validierung zu fordern, ohne Innovation von Anfang an zu blockieren, ist daher weder im Bereich des Möglichen, noch ist es sinnvoll.

H. Beziehungen zu anderen Gesetzen und Gerichtsbarkeiten

Aus regulatorischer Sicht müssen Entwicklung, Training, Bewertung und der Einsatz von KI-Systemen neben den Vorgaben durch den anstehenden AI Act verschiedene Anforderungen erfüllen. Dazu gehören:

- Richtlinie 2013/36/EU über den Zugang zu den Tätigkeiten von Kreditinstituten und die aufsichtsrechtliche Überwachung von Kreditinstituten und Wertpapierfirmen
- EU-Datenschutzgrundverordnung (DSGVO)

[7] IIF-Bias und ethische Implikationen im maschinellen Lernen (S. 10).

- European Data Act
- European Data Governance Act
- Bankaufsichtliche Anforderungen an die IT (BAIT)
- Mindestvoraussetzungen für das Risikomanagement (MaRisk[8])
- EBA Guidelines on outsourcing arrangements [9]
- Digital Operational Resilience Act (DORA)
- EBA Guidelines on ICT and security risk management
- Richtlinie über Märkte für Finanzinstrumente insbesondere für Algorithmic Trading
- Mindestens indirekt durch viele weitere Regulierungen wie Gesetze zu Verbraucherschutz, Gleichberechtigung, Antidiskriminierung usw.

Andere Gerichtsbarkeiten arbeiten aktuell bereits daran, KI-bezogene Vorschriften mit unterschiedlichen Blickwinkeln und Prioritäten zu entwickeln. Darüber hinaus argumentieren andere Jurisdiktionen, dass ein Großteil der KI-Themen bereits ausreichend durch bestehende Vorschriften abgedeckt sind und möglicherweise nur entsprechend detailliert werden müssen. Diese Entwicklungen bergen das Risiko einer Fragmentierung des Marktes und stellen multinationale Organisationen vor große Herausforderungen.

I. Schlussfolgerung

Wir schätzen die allgemeine Ausrichtung des AI Acts – einen Ansatz für ein Gütesiegel von KI, die in Europa hergestellt oder verwendet wird. Definitionen sind auf dem neuesten Stand der Technik und ein risikobasierter Ansatz ist geeignet. Da aber das Gesetz für alle Branchen gelten soll, werden einige Bankspezifika nicht hinreichend berücksichtigt. Dies ist besonders relevant bei der Betrachtung der feinen Grenze zwischen Lifecycle-Kundenmanagement und Bonitätsbewertung sowie der Nachteile, die zusätzliche Zertifizierungen in einem bereits extern auditierten Gebiet wie der Bonitätsbewertung haben. Hier ist es wichtig, die bereits laufende Aufsicht durch die zuständigen Behörden als gleichwertig mit Zertifizierungen zu betrachten. Darüber hinaus werden die Begriffe rund um KI-Ethik wie Transparenz, Erklärbarkeit und Fairness in der Öffentlichkeit oft missverstanden. Da der AI Act nicht ausreichend detailliert darlegt, wie dies aus praktischer Sicht umgesetzt werden soll, haben wir eine Vorstellung davon gegeben, was diese Begriffe bedeuten und wie dies im Bankenkontext eingebettet werden kann.

Es ist nicht die Absicht dieses White Papers, die Standards zu senken, sondern auf die bereits bestehende und effektive Regulierung in der Bankenbranche hinzuweisen und für eine realistische, effektive und angemessene Umsetzung der neuen Anforderungen zu werben. Aufgrund der Verflechtungen von Künstlicher Intelligenz mit verschiedenen anderen Regulierungen, wie oben beschrieben, fordern wir eine kohärente Harmonisierung der Vorgaben.

[8] Mindestvoraussetzungen für das Risikomanagement (MaRisk) in der Version 16.08.2021 und der Konsultation 06/2022 - Entwurf der MaRisk in der Version 26.09.2022.

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