FIRM
Frankfurt Institute for
Risk Management and Regulation

Positionspapier – September 2024

Inhalt

POSITIONSPAPIER DES ROUND TABLE ARTIFICIAL INTELLIGENCE

Künstliche Intelligenz: Herausforderung und Chance für das Modellrisikomanagement von Banken[i]

Von Dr. Sebastian Fritz-Morgenthal, Philipp Adamidi, Dr. Jochen Papenbrock

Einleitung

Die Nutzung von Künstlicher Intelligenz (KI) in Banken verspricht eine Vielzahl von Vorteilen, von der Effizienzsteigerung bis hin zur Verbesserung der Entscheidungsfindung. Gleichzeitig bringt der verstärkte Einsatz von KI-Modellen neue Herausforderungen mit sich, insbesondere im Bereich des Modellrisikomanagements. Bisher verfolgte Strategien bedürfen einer fundamentalen Überarbeitung. In diesem Artikel werden wir die regulatorischen Anforderungen und Regularien der Europäischen Bankenaufsichtsbehörde (EBA), der Europäischen Zentralbank (EZB), der Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) sowie des EU AI Act detailliert beleuchten. Nachfolgend werden wir darstellen, wie sich der erweiterte Einsatz von KI auf das Modellrisiko auswirkt und welche Herausforderungen aber auch insbesondere Chancen dies für das Modellrisikomanagement von Banken darstellt.

Bevor wir in das Thema einsteigen, möchten wir den „**Sense of Urgency**“ für das Thema KI/AI-Risikomanagement betonen. Bei vielen technologischen Trends (z.B. WWW/Internet, Cloud, Crypto, PSD / Payments etc.) war neben der „Early-Adopter“-Strategie auch ein abwartend-beobachtender Ansatz durchaus erfolgversprechend. Man wartet, bis sich eine stabile technologische Lösung etabliert, und springt dann auf. Bei der künstlichen Intelligenz würden wir von einer solchen Strategie dringend abraten. Tatsächlich befinden wir uns mitten in einem technologischen Paradigmenwechsel.

Technologischer Paradigmenwechsel

Von der einfachen Maschine, die Vorgaben ihrer Anwender 1:1 und mit hoher Effizienz umsetzt, hin zu einer Welt, in der Bewertungs- und Entscheidungsalgorithmen selbstlernend, dynamisch, statistisch statt deterministisch umsetzt, sich mehr oder weniger aktiv mit anderen Systemen vernetzt und dabei möglicherweise Produkte und Services kreiert, die wir jetzt noch gar nicht kennen.

Umso wichtiger erscheint es uns, dass Banken ihr Wissen und Verständnis der Grundlagen, des Entwicklungspotentials und der Risiken dieses technologischen Wandels aufbauen und beständig weiterentwickeln. Mit unserem Artikel wollen wir dazu einen Beitrag leisten.

Regulatorische Anforderungen

ANFORDERUNGEN DER EBA, EZB UND BAFIN[1]

Die regulatorischen Anforderungen der EBA, EZB und BaFin legen strenge Richtlinien für den Einsatz von Modellen in Banken fest. Diese Anforderungen konzentrieren sich auf mehrere Kernbereiche:

1. **Transparenz und Nachvollziehbarkeit:** Banken müssen sicherstellen, dass ihre KI-Modelle transparent und nachvollziehbar sind. Dies bedeutet, dass die Entscheidungsprozesse der Modelle klar dokumentiert und erklärbar sein müssen.

2. **Regelmäßige Validierung und Überprüfung:** KI-Modelle müssen regelmäßig validiert und auf ihre Leistung überprüft werden. Dies umfasst Tests zur Modellgenauigkeit sowie die Bewertung der Modellannahmen und der verwendeten Daten.

3. **Governance und Kontrolle:** Es muss ein umfassender Governance-Rahmen vorhanden sein, der die Entwicklung, Implementierung und Überwachung von KI-Modellen regelt. Dazu gehört auch die Unabhängigkeit der Modellvalidierungsteams.

4. **Datenqualität:** Die Qualität der Daten, die in KI-Modelle einfließen, ist von entscheidender Bedeutung. Banken müssen sicherstellen, dass ihre Daten vollständig, korrekt und relevant sind.

5. **Proportionalität:** Die Komplexität der KI-Modelle muss im Verhältnis zu ihrem Einsatzgebiet und den damit verbundenen Risiken stehen. Modelle, die in kritischen Bereichen wie der Kreditvergabe oder dem Risikomanagement eingesetzt werden, unterliegen strengeren Anforderungen.

EU AI ACT[2]

Der EU AI Act ist ein umfassender Regulierungsrahmen, der den Einsatz von KI in der Europäischen Union regelt. Er kategorisiert KI-Systeme nach ihrem Risiko und legt spezifische Anforderungen für die verschiedenen Risikokategorien fest:

1. **Verbotene KI-Systeme:** Modelle, die ein unvertretbares Risiko darstellen, sind verboten. Beispiele im Bankwesen könnten KI-Systeme sein, die dazu verwendet werden, Kundenverhalten auf manipulative Weise zu beeinflussen, etwa durch extreme Formen von personalisierten Angeboten, die auf unethische Weise gestaltet sind, um Kunden zu riskanten Finanzentscheidungen zu drängen.

2. **Hochrisiko-KI-Systeme:** Modelle, die in kritischen Bereichen wie dem Finanzwesen eingesetzt werden, unterliegen strengen Regulierungen. Ein Beispiel sind KI-Modelle zur Kreditvergabe, die das Risiko von Kreditausfällen vorhersagen. Diese Systeme müssen umfassende Risiko- und Sicherheitsbewertungen durchlaufen, eine hohe Transparenz aufweisen und kontinuierlich überwacht werden, um sicherzustellen, dass sie keine diskriminierenden Entscheidungen treffen.

3. **Geringes Risiko und Minimalrisiko:** Modelle, die ein geringes oder minimales Risiko darstellen, unterliegen weniger strengen Anforderungen. Ein Beispiel wäre ein KI-gestütztes Chatbot-System, das einfache Kundenanfragen bearbeitet. Solche Systeme müssen grundlegende Transparenzanforderungen erfüllen, wie z. B. die Offenlegung, dass es sich um eine KI handelt.

Banken müssen sicherstellen, dass ihre KI-Modelle den Anforderungen des EU AI Acts entsprechen, insbesondere hinsichtlich der Klassifizierung und den damit verbundenen regulatorischen Maßnahmen. Dies beinhaltet insbesondere eine kontinuierliche Überwachung des Modellrisikos mit möglicher Reklassifizierung und entsprechenden Anpassungen. Dabei sollte auch unterschieden werden zwischen ML Modellen, die rein auf internen Daten trainiert sind, und GenAI Foundation Modellen, deren Trainingsdatenbasis im Zweifel nicht vollumfänglich bekannt ist. Insofern muss sich die Evaluierung an der Komplexität der Modellklasse orientieren. Bei logistischer Regression sind die Ergebnisse deterministisch, für ML-Verfahren bietet sich z. B. Shapley Values an (siehe auch unser FIRM-Papier[3]), bei Gen AI liegt ein Schwerpunkt auf den sogenannten „Halluzinationen“, die man vermutlich besser als einfache Performanceschwankungen bezeichnet.

Herausforderungen durch den Einsatz von KI-Modellen

Der verstärkte Einsatz von KI in Banken führt zu einer signifikanten Erweiterung des Modelluniversums, was mehrere spezifische Herausforderungen mit sich bringt:

1 Weitere Informationen finden Sie auf den offiziellen Websites der EBA (https://www.eba.europa.eu/regulation-and-policy/internal-governance/guidelines-on-internal-governance), EZB (https://www.bankingsupervision.europa.eu/press/publications/supervisoryreview/html/index.en.html) und BaFin (https://www.bafin.de/DE/Aufsicht/BankenFinanzdienstleister/Mindestanforderungen/mindestanforderungen_node.html).

2 Weitere Informationen finden Sie auf der offiziellen EU AI Act-Website (https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence).

3 https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2022.779799/full

1. **Komplexität der Modelle:** KI-Modelle, insbesondere solche, die auf maschinellem Lernen basieren, sind oft komplex und schwer nachvollziehbar. Dies erschwert die Validierung und die Interpretation der Ergebnisse.

2. **Erweiterung des Anwendungsbereichs:** Traditionell konzentrierte sich das Modellrisikomanagement auf Modelle zur Risikomessung. Mit dem Einsatz von KI in neuen Bereichen wie der Kundenberatung, der Beantwortung und Abwicklung von Kundenanfragen sowie im Mahnwesen erweitern sich die Anforderungen erheblich. Diese neuen Anwendungsbereiche bringen zusätzliche Risiken mit sich, da die Modelle nun direkt die Interaktion mit Kunden betreffen und deren Erfahrungen und Zufriedenheit erheblich beeinflussen können.

3. **Erklärbarkeit und Transparenz:** Regulatoren und interne Stakeholder verlangen eine hohe Transparenz und Erklärbarkeit von Modellen. Dies steht oft im Widerspruch zur inhärenten Komplexität vieler KI-Algorithmen.

4. **Kontinuierliche Überwachung und Anpassung:** KI-Modelle müssen regelmäßig überwacht und an veränderte Rahmenbedingungen angepasst werden, z. B. Identifizierung von Bias, eine Veränderung der Daten oder des Kontexts. In jedem Fall muss man testen, ob die Modell-Funktionalität noch gegeben ist. Dies erfordert umfangreiche Ressourcen und Fachwissen.

5. **Verantwortung und Ethik:** Der Einsatz von KI-Modellen, insbesondere in der Kundeninteraktion, erfordert eine klare ethische Verantwortung. Banken müssen sicherstellen, dass ihre KI-Modelle keine diskriminierenden Entscheidungen treffen und fair gegenüber allen Kunden sind. Darüber hinaus gilt selbstverständlich die DSGVO / GDPR.

Der Einsatz von KI-Modellen im Modellrisikomanagement

Trotz der genannten Herausforderungen bietet der Einsatz von KI im Modellrisikomanagement erhebliche Chancen:

1. **Effizienzsteigerung:** KI kann den Prozess der Modellentwicklung und -validierung erheblich beschleunigen und automatisieren.

2. **Bessere Entscheidungsfindung:** Durch die Analyse großer Datenmengen können KI-Modelle präzisere Vorhersagen und fundiertere Entscheidungen ermöglichen.

3. **Frühzeitige Risikoerkennung:** KI kann helfen, potenzielle Risiken frühzeitig zu identifizieren und geeignete Gegenmaßnahmen zu ergreifen.

4. **Innovationspotenzial:** Der Einsatz von KI fördert die Innovation innerhalb von Banken und kann zur Entwicklung neuer Produkte und Dienstleistungen führen. KI kann zum systematischen Testen verwenden werden. Allerdings sollte tunlichst vermieden werden, eine Blackbox mit einer anderen zu validieren. Das verlangt umfangreiche Vorbereitung.

5. **Rolle des Modell-Risikomanagements:** Ein aktives Modell-Risikomanagement kann dabei helfen, die Chancen des KI-Einsatzes optimal zu nutzen, indem es eine robuste Rahmenstruktur für die Entwicklung, Überwachung und Anpassung von KI-Modellen schafft. Durch die Implementierung strenger Kontrollmechanismen und regelmäßiger Überprüfungen können Risiken minimiert und die Zuverlässigkeit der Modelle gewährleistet werden. Darüber hinaus kann ein proaktives Risikomanagement dabei helfen, regulatorische Anforderungen zu erfüllen und das Vertrauen von Stakeholdern in die KI-Systeme zu stärken.

6. **KI validiert KI:** Eine besondere Herausforderung liegt im Einsatz der KI zur Modellvealidierung durch KI. Es besteht das Risiko der Selbstähnlichkeit. Auch wenn man ChatGPT mit Gemini validiert, ist das nicht unbedingt unabhängig, wenn diese größtenteils mit den gleichen Datensätzen trainiert worden sind. Auch die Verwendung von synthetisch generierten Daten bieten Chancen aber eben auch Risiken. Insofern gehören zu einer unabhängigen Validierung auch zwingend separat ermittelte / gesammelte Daten zu verwenden.

Rolle und Chancen des Modell-Risikomanagements

Für das Modell-Risikomanagement als klassische Second Line of Defense Funktion stellen sich mit dem zunehmenden Einsatz von KI-Technologien **zwei wesentlich-erweiternde Herausforderungen**:

1. Die Validierung von KI-Modellen ist grundsätzlich aufwändiger als von klassischen statistisch-deterministischen Modellen. Neben Alpha- und Beta-Fehler, Rang-und Gini koeffizienten, Quantilen und Ausreißern, Out-of-Sample-Tests und allgemeine Modellstabilität spielen Erklärbarkeit, Nachvollziehbarkeit und Fairness von

Modellen eine wesentliche Rolle. Der FIRM-AI-Roundtable hat zu dem Thema 2021 bereits einen Übersichtsartikel veröffentlicht[4]. Das verlangt zusätzliche Testverfahren und Routinen sowie bei häufig eine deutlich erhöhte Validierungsfrequenz.

2. Während bisher zu validierende Modelle ihren Fokus auf die Bewertung von Kapitalmarktprodukten, von Ausfallwahrscheinlichkeiten und Erlösquoten im Kreditgeschäft, Parametern der Verlustverteilungen im Operationellen Risikomanagement sowie den Berechnungen von RWA- und Kapitalquoten legten, breitet sich die KI praktisch überall im Bankgeschäft aus. Bots im Kundenverkehr, Erstellung und/oder Zusammenfassung von Finanz- und Investmentberichten, Investitionsempfehlungen, Abarbeitung von Beschwerden und vieles mehr.

Schon die **erste Herausforderung** lässt sich nicht einfach durch einen „More of the same"-Ansatz mit einer Vergrößerung der Modell-Risiko-Management-Abteilung lösen. Die auch bei der klassischen Modellvalidierung müssen Daten, Governance und Use Test berücksichtigt werden. Die Funktion verlangt nach zusätzlichen Qualifikationen und Fähigkeiten:

- KI-Methoden müssen in einem ersten Schritt (im Einklang mit dem EU-AI-Act) bzgl. Ihrer Risikorelevanz klassifiziert werden:

  a. Minimales Risiko – z. B. Optimierung interner Prozesse, keine Kundenklassifikation, keine Kundeninteraktion

  b. Moderates Risiko – z. B. Kundenklassifikation zur Segmentierung in Betreuung und Vertrieb

  c. Hohes Risiko – Risikobewertung von Kunden, Produkten, Transaktionen, bilanz- und GuV-relevanten Positionen

- Während KI-Methoden mit minimalen Risiko nicht zwingend validiert werden müssen, sollten solche mit moderatem Risiko validiert werden. Hoch-Risiko-Methoden werden üblicher im Rahmen regulatorischer Vorgaben zwingend reguliert

- In einem zweiten Schritt verlangt das nicht-deterministische, zufällige Element in der Validierung besondere Aufmerksamkeit. Wie stabil sind Modellergebnisse?

- Der dritte Schritt fokussiert auf die Nachvollziehbarkeit und Transparenz von Modellen. Sind die Modellvorhersagen im Großen und Ganzen nachvollziehbar? Neben der Nachvollziehbarkeit spielt insbesondere auch Fairness und Diskriminierungsfreiheit des Modells eine besondere Rolle.

- Erst der vierte Schritt entspricht der klassischen Modell-Validierung von Qualität und Stabilität der Vorhersagegüte.

Die **zweite Herausforderung** lässt sich nur durch einen Veränderungsprozess im gesamten Unternehmen bewältigen. Aller Voraussicht nach werden Modelle die Arbeitsprozesse von jedem/jeder Mitarbeiter*in verändern. Das verlangt einerseits, dass jede*r ein grundsätzliches Verständnis von Modell-basierten Ergebnissen (im Sinne von Prognosen, Bewertungen, Handlungsempfehlungen) entwickelt und wie diese zu interpretieren sind. Ganz praktisch läuft dies auf **Weiterqualifikationsmaßnahmen** für jede*n Mitarbeiter*in hinaus.

Modellrisikomanager*innen haben funktionsbedingt dieses Wissen und können deshalb beim „**Upskilling**" aller Kolleg*nnen eine aktive Rolle spielen. Andererseits müssen Unternehmen ein grundlegendes Verständnis entwickeln, dass idealerweise alle KI-Modell-Risiken aktiv gemanagt werden müssen. Die Modell-Risiko-Managementfunktion spielt hierbei ebenfalls eine zentrale Rolle im Sinne einer zweiten Verteidigungslinie, aber sie kann dies nicht allein leisten. Vertriebs-, Produkt- und Service-Bereiche müssen für die von ihnen eingesetzten KI-Methoden und -Modelle aktives Modell-Risikomanagement betreiben, um nachhaltigen Geschäfts-Erfolg zu gewährleisten. Wie können Modelle dynamisch angepasst werden, ohne das Vertrauen der Kunden zu gefährden? Wie können fälschlich diskriminierende Modell-Eigenschaften schnell identifiziert und Modelle entsprechend korrigiert werden? Wie lässt sich Model Drift rechtzeitig erkennen und eindämmen. Die **aktive Zusammenarbeit** zwischen Vertrieb, Produkt, Service und Modellrisiko ist somit ein **kritischer Erfolgsfaktor** für den Einsatz von KI-Methoden.

Ein weiterer kritischer Erfolgsfaktor ist der Einsatz von Technologie zur Prüfung und Validierung von KI-Methoden und -Instrumenten. „**KI prüft KI**" ist keine ferne Vision, sondern in vielen Bereichen gelebte Realität. Modellrisikomanager*innen können und sollten diese Chancen aktiv nutzen, deren Herausforderungen aber auch aktiv managen.

4 Financial Risk Management and Explainable, Trustworthy, Responsible AI, https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2022.779799/full

Der optimale Risk Return bei Modellrisiken sieht sicher anders aus als bei klassischen Kredit- oder Marktpreisrisiken. Er kann jedoch im Sinne einer Pareto-Effizienz nicht Null sein. Banken, Regulatoren, Kunden und Anbieter müssen sich auf Minimalstandards einigen. **Technische und regulatorische Zertifizierungen** können und werden hierbei eine entscheidende Rolle spielen, um die Komplexität der Modelle aktiv und nachhaltig nutzen und weiterentwickeln zu können.

Das Modellrisikomanagement der KI-basierten Applikationen kann durch durch Leitplanken-Technologien ('Guardrails') verbessert werden. NeMo Guardrails zum Beispiel ist ein Open-Source-Toolkit zum einfachen Hinzufügen von programmierbaren Guardrails zu Konversationsanwendungen basierend auf Large Language Modellen (LLM). Guardrails sind spezifische Möglichkeiten, die Ausgabe eines großen Sprachmodells zu steuern, wie z.B. nicht über Politik zu sprechen, auf eine bestimmte Art und Weise auf bestimmte Benutzeranfragen zu reagieren, einem vordefinierten Dialogpfad zu folgen, einen bestimmten Sprachstil zu verwenden, strukturierte Daten zu extrahieren und mehr. Entwickler können auf einfache Weise programmierbare Guardrails zwischen dem Anwendungscode und dem LLM hinzufügen.

Die Vorteile sind:

- Aufbau vertrauenswürdiger, sicherer und geschützter LLM-basierter Anwendungen: Sie können Leitplanken definieren, um Konversationen zu leiten und zu schützen; Sie können das Verhalten Ihrer LLM-basierten Anwendung in Bezug auf bestimmte Themen festlegen und verhindern, dass sie sich an Diskussionen über unerwünschte Themen beteiligt.

- Sichere Verbindung von LLM-Modellen und anderen Diensten

- Kontrollierbarer Dialog: Sie können das LLM so steuern, dass er vordefinierten Konversationspfaden folgt, so dass die Interaktion nach bewährten Konversationsverfahren gestaltet und Standardbetriebsverfahren (z. B. Authentifizierung, Support) umgesetzt werden können.

NeMo Guardrails bietet auch mehrere Mechanismen zum Schutz einer LLM-basierten Chat-Anwendung gegen gängige LLM-Schwachstellen, wie Jailbreaks und Prompt Injections.

Fazit

Der Einsatz von Künstlicher Intelligenz im Modellrisikomanagement von Banken stellt eine bedeutende Herausforderung dar, bietet jedoch auch immense Chancen. Die Einhaltung der regulatorischen Anforderungen der EBA, EZB, BaFin und des EU AI Acts ist dabei unerlässlich. Banken müssen sicherstellen, dass ihre KI-Modelle robust, nachvollziehbar und kontrollierbar sind. Gleichzeitig sollten sie die Potenziale von KI nutzen, um ihre Entscheidungsprozesse zu verbessern und Wettbewerbsvorteile zu erzielen. Ein ausgewogenes und gleichzeitig aktiv agierendes Modellrisikomanagement wird dabei der Schlüssel zum Erfolg sein.

DIE AUTOREN:

Dr. Sebastian Fritz-Morgenthal
Advisense

Philipp Adamidis
QuantPI

Dr. Jochen Papenbrock
NVIDIA

i Dieser Artikel wurde mit Unterstützung durch Chat GPT 4o erstellt