Eidgenössische Finanzmarktaufsicht FINMA
Autorité fédérale de surveillance des marchés financiers FINMA
Autorità federale di vigilanza sui mercati finanziari FINMA
Swiss Financial Market Supervisory Authority FINMA

FINMA-Aufsichtsmitteilung 08/2024

Governance und Risikomanagement beim Einsatz Künstlicher Intelligenz

18. Dezember 2024

Laupenstrasse 27, 3003 Bern, Tel. +41 (0)31 327 91 00, www.finma.ch

Inhaltsverzeichnis

1 Einleitung

Der Einsatz von Künstlicher Intelligenz (KI) im Finanzmarkt nimmt zu.[1] Für die Beaufsichtigten ist dies mit Chancen, aber auch mit Risiken verbunden. Mit dieser Aufsichtsmitteilung macht die FINMA auf entsprechende Risiken und das Erfordernis einer angemessenen Erfassung, Begrenzung und Kontrolle dieser Risiken aufmerksam.

Bislang existiert in der Schweiz keine KI-spezifische Gesetzgebung. Im Finanzmarktrecht erfassen die technologieneutralen, prinzipienbasierten aufsichtsrechtlichen Anforderungen an eine wirksame Governance und ein wirksames Risikomanagement die Risiken aus dem Einsatz von KI. Wie auch im internationalen Umfeld gefordert, erwartet die FINMA, dass sich Beaufsichtigte, die KI einsetzen, aktiv mit den Auswirkungen dieser Nutzung auf ihr Risikoprofil auseinandersetzen und ihre Governance, ihr Risikomanagement und ihre Kontrollsysteme entsprechend ausrichten. Dabei sind neben Grösse, Komplexität, Struktur und Risikoprofil der Beaufsichtigten insbesondere die Wesentlichkeit der genutzten KI-Anwendungen sowie die Eintrittswahrscheinlichkeit der aus der Nutzung dieser Anwendungen resultierenden Risiken zu berücksichtigen. [2]

2 Erkenntnisse aus der Aufsicht

Die Risiken aus dem Einsatz von KI liegen hauptsächlich im Bereich der operationellen Risiken[3], insbesondere Modellrisiken (z.B. mangelnde Robustheit, Korrektheit, Bias und Erklärbarkeit) sowie IT- und Cyber-Risiken. Weiter resultieren sie aus einer steigenden Abhängigkeit von Drittparteien wie etwa Anbietern von Hardware-Lösungen, Modellen oder Cloud-Dienstleistungen in einem zunehmend konzentrierten Markt.[4] Schliesslich bestehen Rechts- und Reputationsrisiken sowie Herausforderungen bei der Zu-

[1] Zur Verbreitung von KI im Finanzmarkt siehe FSB, The Financial Stability Implications of Artificial Intelligence, 14.11.2024 (nachfolgend: FSB), S. 3 ff.

[2] Mögliche (nicht abschliessende) Faktoren, die die Wesentlichkeit einer Anwendung beeinflussen, sind: Bedeutung für die Einhaltung der Finanzmarktgesetzgebung, finanzielle Auswirkungen auf das Unternehmen, Rechts- und Reputationsrisiken, Relevanz des Produktes für das Unternehmen, Anzahl betroffener Kundinnen und Kunden resp. Anlegerinnen und Anleger, Typen von Kundinnen und Kunden resp. Anlegerinnen und Anlegern (retail/institutionell), Wichtigkeit des Produktes für Kundinnen und Kunden resp. Anlegerinnen und Anleger, Konsequenzen bei Fehlern oder Ausfall. Mögliche (nicht abschliessende) Faktoren, die die Eintrittswahrscheinlichkeit der mit den Risiken verbundenen Ereignisse beeinflussen, sind: Komplexität (z.B. Erklärbarkeit, Vorhersehbarkeit), Art und Menge der verwendeten Daten (z.B. unstrukturierte Daten, Integrität, Zweckmässigkeit, Personendaten), ungeeignete Entwicklungs- oder Überwachungsprozesse, Grad der Autonomie und Prozesseinbindung, Dynamik (z.B. kurze Kalibrierungszyklen), Vernetzung mehrerer Modelle, Potenzial für Angriffe oder Ausfälle (z.B. erhöht wegen Outsourcing).

[3] Vgl. Art. 89 ERV: Mit operationellen Risiken wird die Gefahr von Verlusten bezeichnet, die in Folge der Unangemessenheit oder des Versagens von internen Verfahren, Menschen oder Systemen oder in Folge von externen Ereignissen eintreten.

[4] Vgl. auch FSB, S. 16 ff.

ordnung von Verantwortlichkeiten aufgrund des autonomen und schwer erklärbaren Handelns dieser Systeme und verstreuter Zuständigkeiten für KI-Anwendungen bei den Beaufsichtigten.

Nachfolgend sind beispielhaft Massnahmen zur Adressierung spezifisch aus KI-Anwendungen resultierender Risiken aufgeführt, die die FINMA im Rahmen der laufenden Aufsicht, namentlich bei Aufsichtsgesprächen und in ersten spezifischen Vor-Ort-Kontrollen, beobachtet hat. Dies soll die Beaufsichtigten dabei unterstützen, Risiken aus internen und externen KI-Anwendungen zu erkennen, zu beurteilen, zu steuern und zu überwachen.

2.1 Governance

Die FINMA beobachtete, dass sich die Beaufsichtigten in erster Linie auf Datenschutzrisiken, weniger aber auf Modellrisiken wie mangelnde Robustheit und Korrektheit, Bias, mangelnde Stabilität und Erklärbarkeit fokussieren. Zudem erfolgt die Entwicklung von KI-Anwendungen oftmals dezentral, sodass es herausfordernd ist, konsistente Standards umzusetzen, Verantwortlichkeiten klar und an Mitarbeitende mit entsprechenden Fähigkeiten und Erfahrungen zuzuweisen und alle relevanten Risiken zu adressieren. Bei extern eingekauften Anwendungen und Dienstleistungen hatten die Beaufsichtigten teilweise Schwierigkeiten, zu eruieren, ob KI enthalten ist, welche Daten und Methoden verwendet werden und ob eine ausreichende Due Diligence existiert.

Die FINMA beurteilte, ob bei Beaufsichtigten mit vielen oder wesentlichen Anwendungen eine KI-Governance vorhanden ist, die u.a. ein zentral geführtes Inventar einschliesslich einer Risikoklassifizierung und daraus folgender Massnahmen, die Festlegung von Zuständigkeiten und Verantwortlichkeiten bei Entwicklung, Implementierung, Überwachung und Nutzung von KI, Vorgaben zu Modell-Tests und unterstützenden Systemkontrollen, Dokumentationsstandards sowie breite Schulungsmassnahmen umfasst. Im Fall von Outsourcing beurteilte sie, ob die Beaufsichtigten zusätzliche Tests, Kontrollen und Vertragsklauseln, die Verantwortlichkeiten und Haftungsfragen regeln, implementiert und sichergestellt hatten, dass die betrauten Dritten über die nötigen Fähigkeiten und Erfahrungen verfügen.

2.2 Inventar und Risikoklassifizierung

Die FINMA beobachtete, dass die Beaufsichtigten KI teilweise eng definierten, um sich auf vermeintlich grössere oder neue Risiken zu fokussieren. Für viele Beaufsichtigte stellte es eine Herausforderung dar, die Vollständigkeit von Inventaren sicherzustellen, da häufig KI-Entwicklung und -Anwendung im Unternehmen breit gestreut und seit dem Aufkommen generativer KI Anwendungen für jedermann zugänglich sind. Auch hatten nicht alle Beaufsichtigten konsistente Kriterien festgelegt, um Anwendungen zu identifizieren, die aufgrund ihrer Wesentlichkeit sowie ihrer spezifischen Risiken und deren

Eintrittswahrscheinlichkeit besonderer Aufmerksamkeit im Risikomanagement bedürfen.[5]

Die FINMA beurteilte, ob bei den Beaufsichtigten eine ausreichend breite Definition von KI vorlag,[6] da auch klassische Anwendungen ähnliche Risiken aufweisen können und gleiche Risiken gleich zu adressieren sind.[7] Sodann beurteilte sie Vorhandensein und Vollständigkeit von KI-Inventaren sowie die Risiko-Klassifizierung von KI-Anwendungen.

2.3 Datenqualität

Die FINMA beobachtete, dass die Beaufsichtigten teilweise keine Vorgaben und Kontrollen definiert haben, um die Datenqualität bei KI-Anwendungen sicherzustellen.

KI-Anwendungen lernen oftmals automatisiert und ohne menschlichen Eingriff aus Daten. Die Datenqualität ist daher häufig wichtiger als die Auswahl des konkreten Modells. Gleichzeitig können Daten falsch, inkonsistent, unvollständig, nicht repräsentativ oder veraltet und daher von schlechter Qualität sein. Historische Daten können einen Bias enthalten, der sich in zukünftigen Prognosen weiterträgt, oder sie können aufgrund eines veränderten Umfelds für die Prognose nicht mehr repräsentativ sein. Bei eingekauften Lösungen haben die Beaufsichtigten häufig keinen Einfluss auf bzw. kennen die zugrundeliegenden Daten nicht. Dies kann dazu führen, dass diese für die Beaufsichtigten oder das konkrete Anliegen nicht passend sind und die Gefahr der unbewussten Verwendung von gezielt manipulierten Daten steigt. Seit dem verstärkten Einsatz von KI werden ausserdem mehr unstrukturierte Daten wie Texte und Bilder ausgewertet, für die eine Beurteilung der Qualität erschwert sein kann.

Die FINMA beurteilte, ob die Beaufsichtigten in ihren internen Weisungen und Richtlinien Vorgaben definiert haben, um sicherzustellen, dass Daten vollständig, korrekt und integer sind und die Verfügbarkeit von und der Zugang zu Daten gesichert ist.

[5] Tendenziell sind Risiken erhöht, wenn KI zur Einhaltung von Aufsichtsrecht oder zur Ausführung kritischer Funktionen eingesetzt wird oder wenn die Kundschaft oder Mitarbeitende von ihren Ergebnissen stark betroffen sind. Die Kriterien zur Klassifizierung sollten von den Beaufsichtigten festgelegt werden.

[6] Vgl. den Definitionsansatz der OECD: OECD, Explanatory Memorandum on the Updated OECD Definition of an AI System, OECD Artificial Intelligence Papers, March 2024 (No. 8).

[7] KI ist nicht per se eine Hochrisiko-Anwendung. Das mit ihr verbundene Risiko hängt von der Komplexität, Adaptivität und Autonomie der jeweiligen Anwendung, ihrem Anwendungsbereich und ihrer Integration in Prozesse ab.

2.4 Tests und laufende Überwachung

Die FINMA beobachtete bei den Beaufsichtigten teilweise Schwächen bei der Auswahl von Performance-Indikatoren, Tests und laufender Überwachung.

Die FINMA beurteilte, ob die Beaufsichtigten Tests zur Sicherstellung der Datenqualität und Funktionsfähigkeit der KI-Anwendungen vorsehen, die eine Prüfung auf Genauigkeit, Robustheit und Stabilität sowie ggfs. Bias beinhalten.[8] Sie beurteilte, ob Fachpersonen des jeweiligen Anwendungsbereiches hierzu Fragestellungen und vordefinierte Erwartungen lieferten und ob vorab festgelegte Performance-Indikatoren gesetzt wurden, um zu beurteilen, wie gut eine KI-Anwendung die gesteckten Ziele erreicht.[9] Bei regelmässig durchzuführenden Kontrollen beurteilte die FINMA beispielsweise, ob die Beaufsichtigten Schwellenwerte oder andere Validierungsmethoden definiert hatten, um Korrektheit und fortlaufende Qualität der Outputs zu gewährleisten.[10] Zudem beurteilte sie, ob die Beaufsichtigten Veränderungen in Input-Daten überwachen, um sicherzustellen, dass Modelle auch bei verändertem Umfeld anwendbar bleiben (Erkennung und Behandlung von Datendrift). Zur Überwachung gehört auch die Analyse von Fällen, in denen die Ausgabe von den Anwendern ignoriert oder verändert wurde, da solche manuellen Korrekturen Rückschluss auf Schwachstellen geben können. Schliesslich beurteilte die FINMA, ob die Beaufsichtigten Überlegungen zur Erkennung und Behandlung von Ausnahmen vorab anstellen.

2.5 Dokumentation

Die FINMA beobachtete, dass die Beaufsichtigten teilweise über keine zentralen Vorgaben zur Dokumentation verfügen und die vorhandene Dokumentation teilweise nicht ausreichend detailliert und empfängerorientiert ist.

Bei wesentlichen Anwendungen beurteilte die FINMA, ob die Beaufsichtigten in der Dokumentation den Zweck der Anwendung, Datenauswahl und -aufbereitung, Modellauswahl, Performance-Masse, Annahmen, Limitierungen,

[8] Es existiert eine Vielzahl an Tests, um die Leistung und die Ergebnisse einer Anwendung zu beurteilen. Hierzu gehören unter anderem Tests, bei denen die Anwendenden das korrekte Ergebnis kennen und prüfen, ob die Anwendung es liefert (z.B. Backtesting, Out-of-Sample Testing), konstruierte Tests, um zu verstehen, wie sich die Anwendung in bestimmten Grenzfällen verhält (z.B. Sensitivitätsanalysen oder Stress-Testing), Tests mit falschen Input Daten (z.B. Adversarial Testing), oder auch Tests gegen zusätzliche, ggfs. einfachere Benchmark-Modelle. Ausserdem können mit Tests potenzielle Grenzen der Anwendung bewertet und Ergebnisse auf "Wiederholbarkeit" geprüft werden.

[9] Je wesentlicher und komplexer die Anwendung und je weniger über die Funktionsweise des Systems oder die zugrundeliegenden Daten bekannt ist, umso wichtiger ist es, vor dem produktiven Einsatz, bei Veränderungen und – insbesondere aufgrund der Adaptivität heutiger Anwendungen – laufend zu beurteilen, ob die Anwendung entsprechend ihrem Zweck funktioniert. Zudem sind Überlegungen zu Fallback-Mechanismen wichtig, um vorbereitet zu sein, wenn die KI sich in eine unerwünschte Richtung entwickelt und nicht mehr die ursprünglich definierten Ziele erfüllt.

[10] Hierzu können z.B. Stichproben, Backtesting, vordefinierte Testfälle oder Benchmarking beitragen.

Testing und Kontrollen sowie Fallback-Lösungen adressieren. Bei der Datenauswahl betrachtete die FINMA, ob die Beaufsichtigten Datenquellen und Prüfungen der Datenqualität inklusive Integrität, Korrektheit, Zweckmässigkeit, Relevanz, Bias und Stabilität darlegten. Ausserdem betrachtete sie, wie die Beaufsichtigten Robustheit und Zuverlässigkeit sowie Nachvollziehbarkeit der Anwendung sicherstellen und ob sie eine angemessene Einstufung in eine Risikokategorie sowie die dazugehörige Begründung und Prüfung vornehmen.

2.6 Erklärbarkeit

Die FINMA beobachtete, dass Ergebnisse häufig nicht nachvollzogen, erklärt oder reproduziert und somit nicht kritisch beurteilt werden können.

Wenn Entscheidungen gegenüber Anlegerinnen und Anlegern, der Kundschaft, den Mitarbeitenden, der Aufsicht oder der Prüfgesellschaft begründet werden mussten, beurteilte die FINMA die Erklärbarkeit der Anwendungen vertieft. Hierzu gehört u.a. die Treiber der Anwendungen oder das Verhalten unter verschiedenen Bedingungen zu verstehen, um die Plausibilität und Robustheit der Ergebnisse beurteilen zu können.

2.7 Unabhängige Überprüfung

Die FINMA beobachtete nicht in allen Fällen eine klare Abgrenzung zwischen der Entwicklung von KI-Anwendungen und der unabhängigen Überprüfung.

Zudem beobachtete sie, dass nur wenige Beaufsichtigte eine unabhängige Prüfung des gesamten Modellentwicklungsprozesses durch dafür qualifiziertes Personal durchführen, um Modellrisiken konsistent zu identifizieren und zu reduzieren.

Bei wesentlichen Anwendungen beurteilte die FINMA, ob die unabhängige Prüfung die Abgabe einer objektiven, versierten und unvoreingenommenen Meinung über die Angemessenheit und Zuverlässigkeit eines Verfahrens für einen bestimmten Anwendungsfall umfasst und ob die Ergebnisse der unabhängigen Überprüfung bei der Entwicklung der Anwendung berücksichtigt wurden.

3 Ausblick

Das Risikoverständnis beim Einsatz von KI bei Beaufsichtigten ist in Entwicklung. Gestützt auf ihre Erfahrungen aus der Aufsicht und in Anlehnung

an relevante internationale Entwicklungen wird auch die FINMA ihre Erwartungen an eine angemessene Governance und ein angemessenes Risikomanagement der Beaufsichtigten im Zusammenhang mit KI weiterentwickeln und sofern nötig im Markt transparent machen. Dabei strebt die FINMA wie bei anderen relevanten Risikotreibern einen technologieneutralen, proportionalen sowie sektorübergreifend einheitlichen Ansatz an und berücksichtigt wesentliche Unterschiede zwischen den Sektoren sowie internationale Standards.