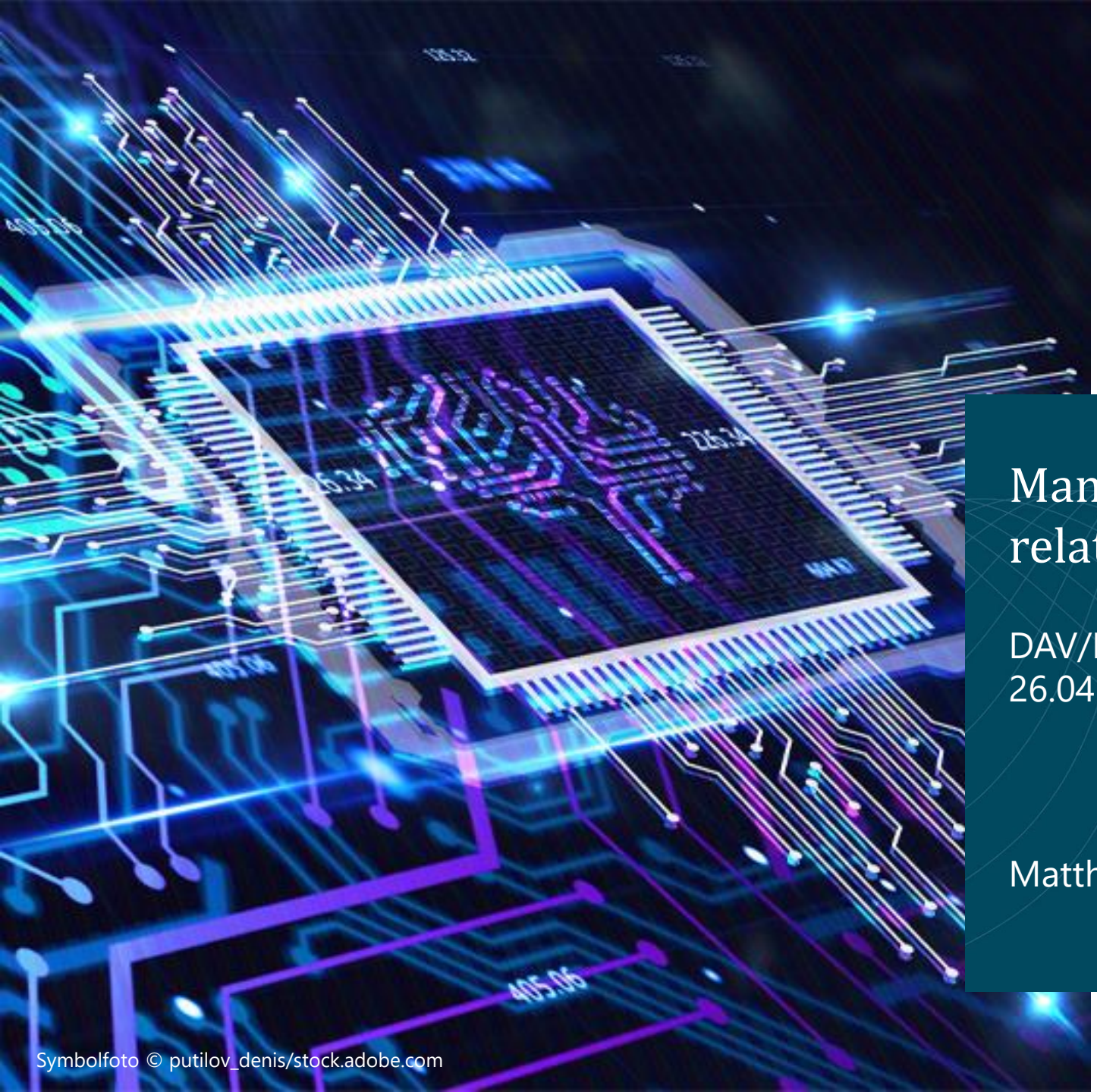
DAV/DGVFM
Jahrestagung
2024

*Dr. Dr. Matthias Fahrenwaldt, BaFin*

# Management von Modellrisiken aus KI/ML

DAV/DGVFM Jahrestagung, 26. April 2024



Symbolfoto © putilov_denis/stock.adobe.com

BaFin

Bundesanstalt für
Finanzdienstleistungsaufsicht

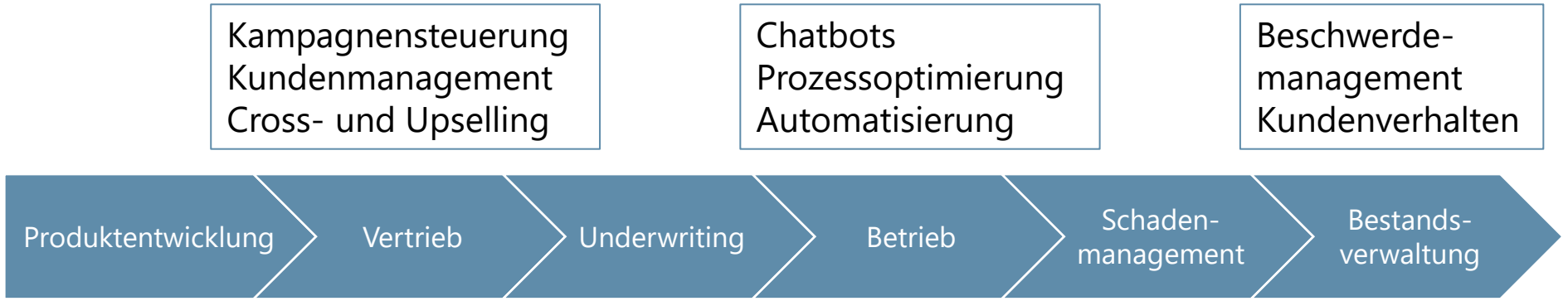
# Management of AI/ML-related Model Risks

DAV/DGVFM Jahrestagung,
26.04.2024

Matthias Fahrenwaldt

# Der Einsatz von KI/ML im Finanzsektor ist entlang der gesamten Wertschöpfungskette möglich – zumeist in unterstützender Funktion

## Beispiel Sachversicherung




**Noch kein signifikanter, aber zunehmender Einsatz von KI/ML in**

- Pricing
- Reservierung
- Risikomodellen

Zunehmender Einsatz generativer AI zur Analyse großer Mengen unstrukturierter Daten

# Alle Modelle sind falsch, aber manche sind nützlich – oder doch nicht?

**Jedes quantitative Modell führt zu Modellrisiken…**

- Beispiel Merton-Black-Scholes
- Ausgewählte Annahmen
  - Arbitragefreiheit
  - Keine Dividenden und Transaktionskosten
  - Konstanter risikoloser Zinssatz und Volatilität
  - Normalverteilung
- Keine dieser Annahmen ist in der Realität erfüllt, dennoch wird das Modell überall genutzt

**…und das Ignorieren der Modellrisiken kann ernsthafte Folgen haben**

Prominentes Beispiel ist der Untergang von LTCM im Jahre 1998[1]

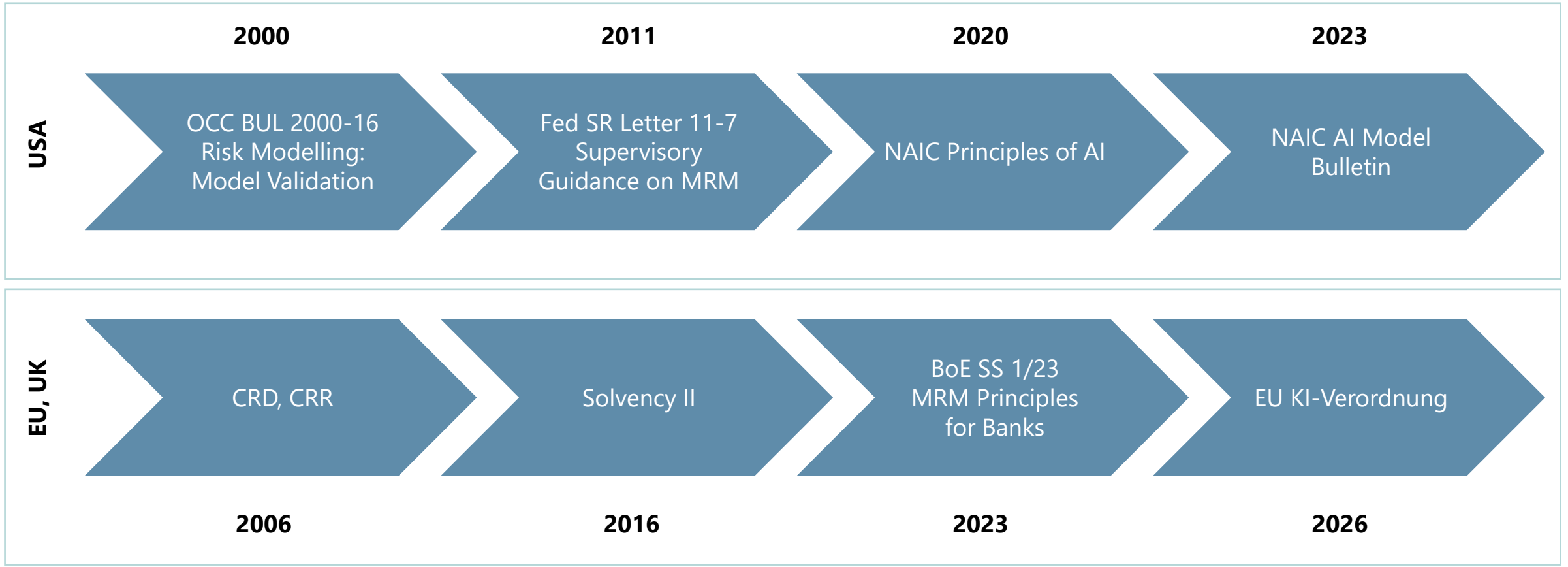
- Hintergrund
  - Hedgefonds gegründet 1993, Partner waren u. a. Myron Scholes, Robert Merton
  - Geschäftsmodell war Arbitragehandel zwischen Bond-Märkten
- Marktereignisse
  - Finanzkrise in Asien erreicht Europa
  - LTCM musste kurzfristig Positionen schließen, erzeugte Verluste, wurde gerettet
- Gründe für den Zusammenbruch
  - Entscheidungen auf Basis von Modellen, ohne diese zu hinterfragen
  - Hohe Returns hätten Investoren warnen sollen, die Reputation der Partner überstrahlte alles

Quelle: 1 Sweeting P. "Case studies." In: *Financial Enterprise Risk Management*. International Series on Actuarial Science. Cambridge University Press; 2011: 505-526.
Siehe auch Lowenstein R. *When Genius failed : The Rise and Fall of Long-Term Capital Management.* Random House; 2000.

# Modellrisiken sind schon länger im Fokus der Aufsicht

VEREINFACHT

**Meilensteine in der Regulierung des Modellrisikomanagement in den USA, UK und der EU**



# Sowohl CRD/CRR als auch Solvency II enthalten MRM-Vorgaben – jedoch nur für interne Modelle

VEREINFACHT

**CRD, CRR**

- Konzentriert auf interne Modelle und weiter detailliert u. a. in ECB Guide to Internal Models oder RTS on Assessment Methodology
- MRM-Anforderungen in mehreren Dimensionen
  - Governance
  - Validierung
  - Interne Revision
  - Modelländerungen
  - Auslagerung
  - Spezialvorgaben je Risikoart (Kredit-, Markt- und Kontrahentenrisiko)

**Solvency II**

- Konzentriert auf interne Modelle mit Schwerpunkt Validierung, weiter ausgeführt in Level-2-Texten
  - Unabhängige Validierung
  - Modelländerung
  - Dokumentation und Offenlegung
  - Explizite Erwähnung bestimmter Modellrisiken: Modelldesign, Datenqualität, Modellannahmen und Expertenschätzungen, Implementierung

Die BAIT/VAIT adressieren MRM-Themen aus IT-Sicht

- Softwareentwicklung
- Dokumentation
- Test
- Informationsrisikomanagement
- Datengovernance

# Das Federal Reserve Board hat einen Drei-Säulen-Ansatz für MRM-Vorgaben umgesetzt

VEREINFACHT

**MRM-Rahmenwerk**

**Entwicklung, Implementierung und Nutzung**

- Übergreifend
  - Dokumentation
  - Verständnis der Modellunsicherheit
  - Vorsichtiger Konservatismus
- Wissenschaftl. informiertes Modelldesign
- Daten: Qualität und Relevanz
- Test: Genauigkeit, Robustheit und Stabilität, Modellbeschränkungen
- IT : Daten-/Berichtsintegrität, Kontrollen
- Feedback der Nutzer

**Validierung**

- Mitarbeiter mit angemessenen Anreizen, Kompetenzen und Einfluss; Fähigkeiten und Expertise
- Erfasst alle Modellkomponenten
- Initiale und fortlaufende Validierung
- Kernelemente im Validierungsrahmenwerk
  - Konzeptionelle Evaluierung
  - Fortlaufendes Monitoring
  - Analyse des Modelloutputs
- Erfasst auch Modelle von Drittparteien

**Governance, Policies und Kontrollen**

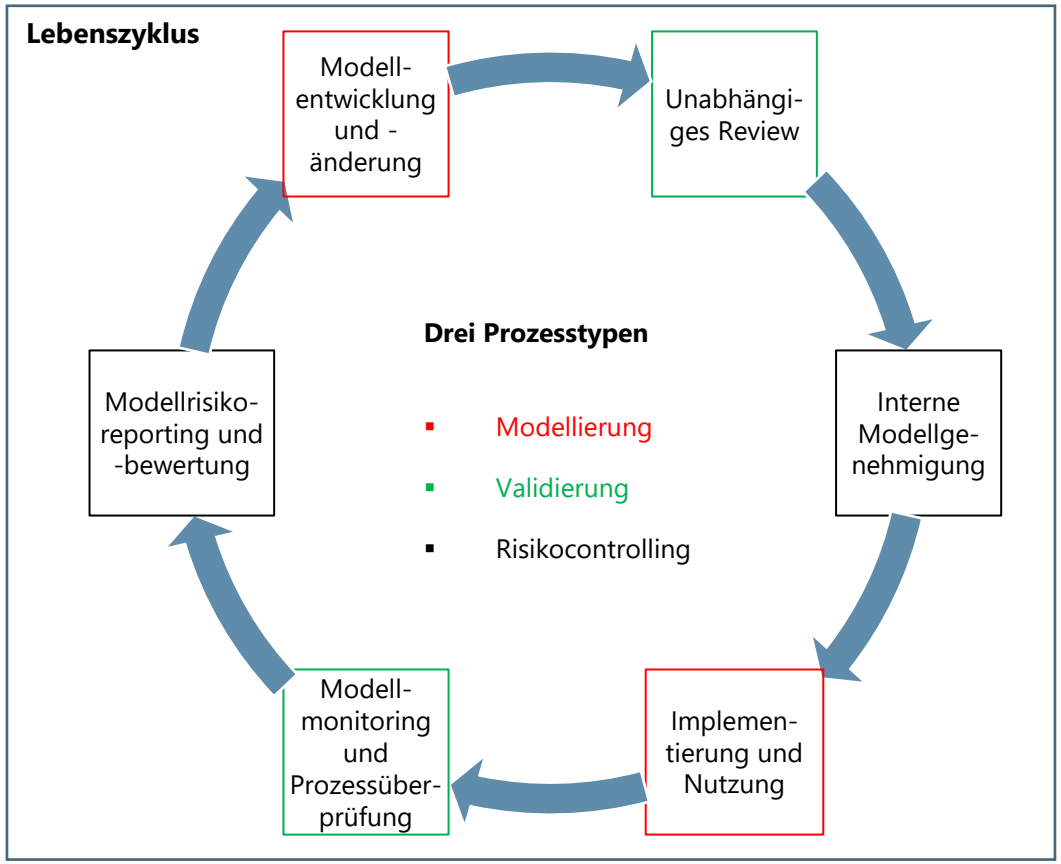
- Umfang/Granularität des Governance-rahmenwerkes entspricht Modellnutzung
- Kernthemen
  - Einbindung Vorstand und Senior Management
  - Policies und Prozesse
  - Interne Revision
  - Externe Ressourcen
  - Modellinventar
  - Dokumentation

**Was ist ein Modell**? „Quantitative method, system or approach that applies statistical, economic, financial, or mathematical theories, and assumptions to process input data into quantitative estimates."

Quelle: Federal Reserve Board SR Letter 11-7

# Die Bank of England strukturiert ihre MRM-Prinzipien entlang des Lebenszyklus eines Modells

VEREINFACHT




**1** Modellidentifizierung und Risikoklassifizierung
- Modelldefinition, Inventar und risikobasierte Klassifikation

**2** Governance
- Einbindung Vorstand, Benennung von MRM-Verantwortlichen

**3** Modellentwicklung, Implementierung und Nutzung
- Robuster Entwicklungsprozess, Test von Modellkomponenten, Bereinigung von Modellschwächen

**4** Unabhängige Validierung
- Fortlaufendes, unabhängiges und effektives Hinterfragen von Modellentwicklung und -nutzung

**5** Mitigierung von Modellrisiken
- Policies und Prozesse zur Mitigierung von Modellrisiken, unabhängiges Review von nachgelagerten Modellanpassungen

Quelle: Bank of England Supervisory Statement SS 1/23

# NAIC arbeitet an aufsichtlichen Erwartungen, einzelne US-Bundesstaaten entwickeln eigene Vorgaben

**NAIC Principles on AI**
(08/2020)

- Nicht bindende Guidance
- Themenfelder sind
  - Fairness/Ethik
  - Verantwortlichkeiten
  - Compliance
  - Transparenz
  - Sicherheit und Robustheit
- Ziel ist es, Innovation zu fördern und gleichzeitig den Verbraucher zu schützen

**NAIC Model Bulletin:**
**Use of AI Systems by Insurers**
(12/2023)

- Vorlage für Regulierung on US-Bundesstaaten
- Zur Zeit in Umsetzung in mehreren Staaten
- Beschreibt aufsichtliche Erwartungen
- Kernthemen sind
  - Governance
  - Risikomanagement/interne Kontrollen
  - Nutzung von KI-Systemen/Daten Dritter

**Definition**
„'AI System' is a machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, content (such as text, images, videos, or sounds), or other output influencing decisions made in real or virtual environments. AI Systems are designed to operate with varying levels of autonomy."

**Beispiel New York State**

- Entwurf eines Rundschreibens vom 17. Januar 2024
- Zur Nutzung von KI, v. a. im Underwriting und Pricing
- Schwerpunkt ist Fairness/ Diskriminierung
- Weitere Themen sind
  - Governance und Risikomanagement, u. a. Einbindung des Vorstands/Senior Management
  - Transparenz und Offenlegung
- Kommentarfrist bis März 2024

Quelle: NAIC Principles on Artificial Intelligence (14.8.2020); NAIC Model Bulletin Use of AI Systems by Insurers (2.12.2023); Ney York State Dept. of Financial Services, Proposed Insurance Circular Letter (17.1.2024).

# Die KI-Verordnung der EU kann als horizontale MRM-Regulierung verstanden werden

VEREINFACHT

**Verboten**

**Inakzeptable Risiken**

- z. B. Social Scoring, hohes Potenzial für Manipulation, ausbeutende Praktiken

**Erlaubt**

**Hohes Risiko**

- Annex III enthält Anwendungsfälle

**Finanzdienstleistungen**: Credit Scoring natürlicher Personen, Risikoeinschätzung und Pricing in Lebens-/Krankenversicherung

**Erlaubt**

**Niedriges Risiko**

- Interaktion mit Maschinen (z. B. KI Chatbot) in Kundenkontakt
- Deep Fakes

**Erlaubt**

**Minimales oder kein Risiko**

- z. B. KI in Videospielen
- Keine Regulierung
- Ggf. freiwilliger Verhaltenskodex

**Ausgewählte Vorschriften des Modellrisikomanagement**

- *Riskomanagementsystem*: Risikoidentifizierung, -bewertung, -mitigierung, Systemtests
- *Daten und Datengovernance*: Datengovernance für Trainings-, Validierungs- und Testdaten; z. B. Datenqualität, Aussagekraft, Bias
- *Technische Dokumentation*: muss ausreichend sein, um Compliance mit der Verordnung zu beurteilen
- *Aufzeichnung*: Aufzeichnungen müssen ein Monitoring erlauben, um zu sehen, dass das System wie gewollt funktioniert
- *Transparenz*: Anwender ("Bereitsteller") müssen Ergebnisse von Systemen Dritter interpretieren und angemessen anwenden können
- *Menschliche Aufsicht*: Output muss interpretierbar sein, Grenzen des Systems sollen erkennbar werden, Eingriff in den Betrieb muss möglich sein
- *Genauigkeit, Robustheit und Cyber Security*: u. a. Benchmarks und KPIs, Resilienz

# Die Nutzung von KI/ML führt zu einer Reihe aufsichtlicher Herausforderungen – Beispiele sind Skills, Governance und Validierung

**Aufsichtliche Herausforderung**

- Bestehende Regulierung muss entsprechend für KI/ML-Modelle interpretiert werden
- Der aufsichtliche Ansatz bleibt bestehen, jedoch werden manche Aspekte mehr betont

| | **Kernthemen** |
|---|---|
| **Skills und Fähigkeiten** | ▪ Expertise der Modellierer, Validierer<br>▪ Skills der Nutzer<br>▪ Fähigkeiten der internen Revision<br>▪ ML-Wissen im (Senior) Management |
| **Governance** | ▪ Kultur<br>▪ Rollen und Verantwortlichkeiten<br>▪ Policies<br>▪ Kontrollen in Prozessen<br>▪ Einbindung der internen Revision |
| **Validierung** | ▪ Unabhängigkeit<br>▪ Effektives Hinterfragen<br>▪ Behandlung von general purpose AI |

# Herausforderung 1: Die Nutzung von KI/ML im gesamten Unternehmen erfordert Skills in ausreichendem Maße

**BaFin-Analyse**

- ML-Experten sind vor allem in der Modellentwicklung und Validierung vorhanden, weniger in der internen Revision
- Geringeres ML-Wissen besteht in weniger modellbezogenen Organisationseinheiten, die jedoch ML nutzen
- Das Management und Senior Management verfügt häufig über eingeschränktere ML-Kenntnisse

**Themen zur Diskussion**

- Wie kann man Beschäftigte mit den erforderlichen Skills anwerben und im Unternehmen halten?
- Wie können Data Scientisten ausreichend versicherungsspezifisches Wissen aufbauen?
- Wie sieht effektives Upskilling aus?
- Welche ML-Kenntnisse benötigen Modellnutzer, Revisoren und Manager?

# Herausforderung 2: Aktuelle Governance-Rahmenwerke adressieren nicht alle Risiken aus ML

**BaFin-Analyse**

- Aktuelle Corporate Governance-Rahmenwerke werden häufig als ausreichend betrachtet
- Wenn diese aktualisiert werden, dann typischerweise an wenigen Stellen, um Risiken aus ML zu behandeln
- Anforderungen an die Erklärbarkeit von Modellen werden oftmals je nach Use Case definiert, "klassische" Anforderungen reichen nicht aus

**Themen zur Diskussion**

- Können bestehende MRM-Rahmenwerke für interne Modelle auch für ML-Anwendungen außerhalb der Säule 1 verwendet werden?
- Wie ist die Wechselwirkung zwischen Modellrisiken und IT-Risiken?
- Welche Skills sind für die Entwicklung/das Review der MRM-Governance erforderlich?
- Wie sollte die menschliche Kontrolle ausgestaltet sein – muss es lokale Ressourcen geben für wesentliche Modelle?

# Herausforderung 3: aktuelle Validierungsansätze sind möglicherweise ungeeignet für komplexe ML-Modelle

**BaFin-Analyse**

- Typischerweise gibt es keine Richtlinien für die Validierung von Modellen außerhalb von Säule 1
- In der Validierung von ML-Modellen verwenden manche Unternehmen mehrere andere ML-Ansätze
- Dies steht im Kontrast zu "klassischen" Modellen, wo Challengermodelle oftmals dieselbe Methodik verwenden wie produktive Modelle

**Diskussionsthemen**

- Braucht man ein übergreifendes Validierungs-Rahmenwerk?
- Was bedeutet Validierung für LLMs und general purpose AI?
- Braucht man Standards für die Erklärbarkeit?
- Wie können Fälle möglicher (in)direkter Diskrimierung identifiziert werden?
- Welche neuen Fähigkeiten sind für die Validierung komplexer Modelle erforderlich?