



**ZUKUNFTSORIENTIERTE ANALYSE DER KI-PRÜFUNG IM KONTEXT DES EU AI ACT**

**TASSILO HEINEMANN (MASTERSTUDIUM INFORMATIK)**

Betreuer: Prof. Dr. Marcel Tilly, Prof. Dr. Andreas Krüger

**ROSENHEIMER INFORMATIKPREIS  
INF-MASTER**

KI ist die prägende Technologie dieser Dekade. Neben den beeindruckenden technischen Möglichkeiten ist eine ebenso entscheidende, aber häufig unterschätzte Frage, die der Governance und Rechenschaftspflicht: Wer trägt die Verantwortung für die Ergebnisse von KI und wie wird diese überprüft? Diese Frage will die EU mit dem EU AI Act, der seit dem 1. August 2024 schrittweise in Kraft tritt, ins Zentrum unternehmerischen Handelns rücken. Aus der Unternehmenssicht ergibt sich hier ein komplexes Geflecht aus Gesetzesparagrafen, Risiken und Compliance-Anforderungen. Für die Prüfungsbranche stellt sich die dringende Frage, wie sie der daraus entstehenden Nachfrage begegnen wird, und Vertrauen in eine der transformativsten Technologien unserer Zeit gewährleisten kann.

**Regulatorisches Gerüst**

Zwar stellt der regulatorische Druck sicherlich den größten Anreiz dar, doch die Notwendigkeit einer Prüfung von KI-Systemen entsteht nicht allein daraus. KI-Prüfung ist die Antwort auf die inhärenten Risiken, die mit dem zunehmend transformativen Einsatz von KI in kritischen Prozessen einhergehen: Von Datenschutzverletzungen und Manipulationen, bis zu physischer Sicherheit in Bereichen wie dem autonomen Fahren. Wie sich KI genau auf unsere Gesellschaft auswirken wird, lässt sich erst sicher bewerten, wenn die Technologie bereits weiter etabliert ist – dann wird allerdings die Steuerung nur noch sehr schwer möglich (das sog. Collinridge-Dilemma, nach Collinridge, 1980). Hier versucht die EU mit dem AI Act den Zielkonflikt durch die risikobasierte Regulierung aufzulösen: Es werden verbotene Praktiken – wie unterschwellige Manipulation – definiert, gewisse Systeme müssen Transparenzpflichten einhalten, im Zentrum stehen aber besonders Hochrisiko-KI-Systeme, denen strenge präventive Anforderungen an Risikomanagement, Dokumentation und Governance auferlegt werden.

Das Ergebnis der Auseinandersetzung mit dieser Regulatorik ist die Feststellung einer fundamentalen Unsicherheit: Der AI Act gibt zwar das Was klar vor, überlässt das Wie der Umsetzung allerdings größtenteils Normen (man denke an anerkannte ISO/IEC-Normen) – die derzeit noch im Entwicklungsstadium sind, was für Unternehmer und Prüfer zu Anforderungsunklarheiten führt. Auch sind Normen, ihrer Natur geschuldet, generisch gehalten und benötigen starke Anpassungen. Dazu kommt, dass es komplexe Zusammenspiele des AI Acts mit anderen Rechtsgebieten wie der DSGVO, dem Produkthaftungsrecht und dem Urheberrecht gibt. Eine isolierte Betrachtung des AI Acts wird einer holistischen Compliance-Strategie nicht gerecht.

**Methodisches Vorgehen**

Um die Lücke zwischen den abstrakten regulatorischen Vorgaben und den konkreten Bedürfnissen und Herausforderungen der Praxis zu schließen, wurde ein zweiteiliger Forschungsansatz verfolgt: Zunächst eine systematische Literaturanalyse, um das Feld der KI-Prüfung theoretisch zu erschließen. Bestehende Konzepte, Anforderungen und Frameworks wurden kartiert, um eine theoretische Basis zu schaffen, die auf etablierten Ansätzen der Auditing Theory von Ian Dennis (1991) und dem Prüfungsstandard IDW PS 861 aufbaut.

Diese Grundlage wurde anschließend durch qualitative Experteninterviews mit sechs Fachleuten aus Bereichen der KI-Prüfung, Rechtsberatung und KI-Anwendung, mit praxisrelevanten Einblicken angereichert. So konnten die in der Theorie identifizierten Konzepte an der Realität gespiegelt werden; unerkannte Herausforderungen aufgedeckt und die tatsächlichen Bedürfnisse von Prüfern und Unternehmern erfasst werden.

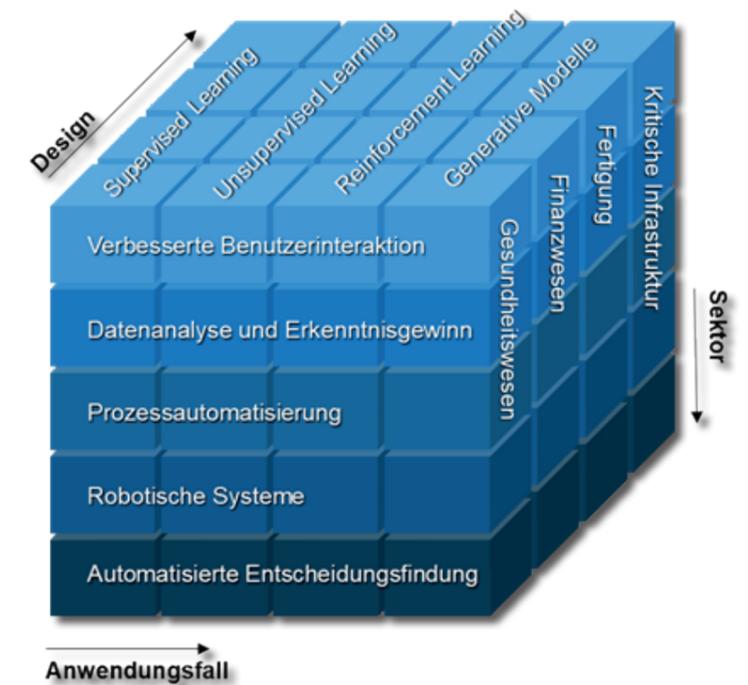
**Ergebnisse und Praktische Implikationen**

Aus dieser Synthese von Regulatorik, Theorie und Praxis wurden mehrere Ergebnisse abgeleitet. Die „KI-Prüfung“ kann nicht als standardisierte Checkliste verstanden werden, sondern ist ein hochgradig kontextabhängiger Prozess.

Es wurde ein dreidimensionales Modell der KI-Prüfung entwickelt, welches den Prüfungsraum entlang der Achsen Design (z. B. Supervised Learning, Generative Modelle), Anwendungsfall (z. B. Kreditbewertung, prädiktive Wartung) und Sektor (z. B. Finanzwesen, Gesundheitswesen) aufspannt. Jedes KI-System lässt sich als ein spezifischer Punkt in diesem Raum verorten, was illustriert, dass Prüfungsstrategien und -schwerpunkte individuell angepasst werden müssen. So wie es nicht „die“ KI gibt, kann es nicht „die“ KI-Prüfung geben.

Darauf aufbauend identifiziert die Arbeit die zentralen Herausforderungen, die sich in vier Kategorien einteilen lassen:

- Grundlegende Herausforderungen betreffen ungelöste Probleme wie die bereits erwähnte Anforderungsunklarheit durch fehlende Normen und die schwierige Operationalisierung ethischer Prinzipien.
- Operative Herausforderungen umfassen praxisnahe Hürden für Prüfungsgesellschaften, darunter der Mangel an Fachkräften, die komplexe Verteilung von Verantwortlichkeiten in teils globalen KI-Wertschöpfungsketten sowie die Skalierbarkeit von Prüfverfahren.
- Technische Herausforderungen beziehen sich auf die inhärenten Eigenschaften von KI, wie die Opazität von „Black-Box“-Modellen, die systemische Dynamik durch kontinuierliches Lernen und die Schwierigkeit, faktische Korrektheit zu gewährleisten.
- Betriebliche Herausforderungen entstehen im Zusammenspiel von Mensch und KI im Unternehmen, etwa bei der Gewährleistung einer angemessenen menschlichen Kontrolle, ohne in „Kontrollfaulheit“ zu verfallen.



Beispielhafte Dimensionen der KI-Prüfung im Würfel

Abschließend leitet die Arbeit aus der Analyse bestehender Frameworks (u. a. von NIST, Fraunhofer IAIS, ISACA) eine Auswahlhilfe für die Prüfungspraxis ab. Anstatt ein weiteres starres Modell vorzuschlagen, wird die Entwicklung eines modularen „Prüfungsbaukastens“ empfohlen. Dieser Ansatz kann es Prüfern ermöglichen, basierend auf einer fundierten Risikoanalyse die relevanten Prüfungsmodule für den spezifischen Anwendungsfall flexibel zu kombinieren.

Die Arbeit liefert damit eine theoretische Durchdringung des Themas, eine pragmatische Landkarte für die bevorstehenden Aufgaben der KI-Prüfung und unterstreicht die Notwendigkeit einer interdisziplinären Zusammenarbeit, um Vertrauen in die KI der Zukunft zu schaffen.