Wildauer Konferenz für Künstliche Intelligenz WiKKI - Wildau 2025

Beiträge zur Wildauer Konferenz für Künstliche Intelligenz 2025

https://doi.org/10.52825/th-wildau-ensp.v2i.2942

Published: 12 Sep. 2025

Erklärbarkeit und menschliche Aufsicht bei KI-Systemen in der Bildung

Regulatorische Rahmenbedingungen

Andre Bonne[1,*] und Uwe Schaffranke[2]

[1]Olymp.Services Bonne GmbH, Deutschland

[2]Landkreis Oder-Spree, Deutschland

*Korrespondenz: Andre Bonne, andre.bonne@olymp.services

Abstract. Der Einsatz Künstlicher Intelligenz (KI) in der beruflichen Bildung bietet erhebliches Potenzial, ist jedoch mit ethischen und regulatorischen Herausforderungen verbunden. Dieser Beitrag analysiert die ethischen und regulatorischen Implikationen von KI in der beruflichen Bildung und präsentiert Strategien zur Minimierung potenzieller Risiken. Anhand eines praxisorientierten Beispiels KFZ Elektromotor wird die Implementierung eines KI-Systems in einem beruflichen Bildungszentrum veranschaulicht. Es werden konkrete Lösungsansätze vorgestellt, um Data-Biases, Datenschutzverletzungen und andere ethische Bedenken zu adressieren. Durch die sorgfältige Selektion und Aufbereitung von Trainingsdaten sowie den Einsatz erklärbarer KI-Modelle wird die Entwicklung fairer, transparenter und zuverlässiger KI-Systeme in der Bildung gefördert. Der Beitrag betont die Notwendigkeit der Konformität mit ethischen Prinzipien bei der Entwicklung und Implementierung von KI-Systemen. Im Kontext des EU AI Acts wird insbesondere auf die Kategorie der hochrisikobehafteten KI-Systeme eingegangen, zu denen KI-Systeme in der beruflichen Bildung potenziell zählen können. Trotz der regulatorischen Herausforderungen werden Akteure ermutigt, KI zur Erstellung von Bildungsinhalten einzusetzen.

Die implementierte Lösung sieht eine menschliche Aufsicht mit entsprechenden IT-Systemen und Richtlinien vor, die als Middleware und Vertrauensinstanz fungiert.

Keywords: KI, Berufliche Bildung, EU AI Act, Ethische Aspekte, Didaktische Innovation, Bias in Trainingsdaten, Didaktische Lehrkonzept, KI-Sprachmodelle, Vertrauens Instanz, Menschliche KI-Aufsicht

1. Einleitung

Die Studie entwickelt ein KI-basiertes Werkzeug für die berufliche Bildung, fokussiert auf Trainingsdaten, Frameworks und KI-Modelle am Beispiel der KFZ-Elektroauto-Branche. Angesichts der Komplexität des Feldes werden didaktische Lehrkonzepte zur Förderung von Fach-, Sozial- und Handlungskompetenz analysiert. Im Folgenden sind die Lehrkonzepte, die wir für unseren Ansatz auf der Grundlage von Befragungen und inhaltlichen Schwerpunkten, ausgewählt haben:

1. **Dualität von Theorie und Praxis:** Verbindet theoretisches Lernen in der Berufsschule mit praktischen Erfahrungen im Betrieb.
2. **Handlungsorientierter Unterricht:** Fokussiert auf Lernen durch eigenes Handeln und Bearbeiten praxisnaher Aufgaben.
3. **Blended Learning:** Kombiniert traditionelle Präsenzphasen mit digitalen Lernangeboten.

In der beruflichen Weiterbildung im Bereich „KFZ-Elektroauto“ wurde aufgrund praktischer Erfahrungen das handlungsorientierte Lernen gewählt. Diese Arbeit konzentriert sich jedoch auf die Erstellung und Qualitätssicherung von Inhalten durch generative KI und geht daher nicht weiter auf die didaktische Lehrmethode ein.

1.1 Handlungsorientierter Blended-Learning-Ansatz

Handlungsorientierter Unterricht und Blended Learning werden kombiniert, um Lerninhalte effektiv zu erstellen. Die Generierung und Kontrolle der Inhalte stehen im Vordergrund. Die theoretische Grundlage wird online durch E-Learning vermittelt, gefolgt von praktischen Übungen. Die Kontrolle und Reflexion erfolgen durch Präsentationen, Feedback und Diskussionen in Online- und Präsenzphasen. Die Trainingsdaten entstehen aus der praktischen Anwendung und Reflexion, was eine praxisnahe Inhaltserstellung ermöglicht.

1.2 Handlungsorientiertes Blended-Learning-Projekt: Aufbau eines Elektromotors

1.2.1 Einleitung in die Theorie (Online-Phase)

Online-Kurse, Videos und interaktive Übungen vermitteln Grundlagenwissen zu Elektromotoren.

1.2.2 Kontrolle und Reflexion (Online- und Präsenz-Phase)

Lernende dokumentieren ihren Prozess und die Ergebnisse in Präsentationen. Feedback und Diskussionen erfolgen online und in Präsenz. Ein Qualitätsmanagementsystem überwacht und kontrolliert KI-generierte Inhalte. Automatisierte Bewertungen basieren auf Qualitätsprüfungen und Leitlinien. Die in den Phasen "Feedback" und "Reflexion" evaluierten Trainingsdaten werden für das Fine-Tuning verwendet.

1.3 Schritte zur Erstellung von hochwertige Trainingsdaten für die KI

1.3.1 Themenrecherche und Inhaltsgenerierung

Fachwissen wird genutzt, um Informationen zu Elektromotoren zu sammeln. Inhaltsstrukturierung: Leitlinien helfen, die Informationen zu strukturieren (z.B. Gliederungen, Stichpunkte, Texte).

1.3.2 Erstellung von Trainingsdaten

Multimedia-Inhalte (Präsentationen, Feedback, Reflexionen) liefern relevante Trainingsdaten für die KI. Quizfragen und Übungen, die von Fachleuten erstellt werden, dienen der Überprüfung des Lernerfolgs und können zur Validierung der KI-generierten Inhalte genutzt werden.

1.4 Schritte zur Erstellung von hochwertige Trainingsdaten für die KI

1.4.1 Themenrecherche und Inhaltsgenerierung

Fachwissen dient zur Informationssammlung über Elektromotoren. Leitlinien unterstützen die logische Strukturierung der Informationen (Gliederungen, Stichpunkte, Texte).

1.4.2 Erstellung von Trainingsdaten

Trainingsdaten für das KI-System werden aus Multimedia-Inhalten (Präsentationen, Feedback, Reflexionen) generiert. Experten erstellen Tests, die den Lernerfolg überprüfen und die KI-generierten Trainingsdaten validieren.

1.5 Integration des Qualitätsmanagementsystems (QMS)

1.5.1 Inhaltliche Kontrolle und Prüfung

KI-basierte Überprüfung der Inhalte auf Korrektheit, Relevanz und Vollständigkeit, inklusive Grammatik-, Rechtschreib-, Plagiats- und Faktenprüfung. Ein Qualitätsmanagementsystem mit klaren Richtlinien und Standards steuert die Inhaltserstellung und -prüfung unter Berücksichtigung ethischer Leitlinien.

1.5.2 Feedback-Schleifen und Anpassungen

Regelmäßige Feedbackschleifen mit Ausbildern, Experten und Lernenden. KI-Tools analysieren das Feedback und priorisieren es für das Qualitätsmanagement, welches die Einhaltung der Standards vor Veröffentlichung sicherstellt.

1.6 Beispiel für den Aufbau eines Elektromotors

KI-generierte Inhalte (Artikel, Bilder, Quizze) zu Elektrotechnik und Elektromotoren werden online bereitgestellt. Die Validität der Inhalte wird durch praxisbezogene, von Experten und im handlungsorientierten Lernen generierte Trainingsdaten sichergestellt. Die detaillierte Umsetzung wird in Methodik und Evaluation erläutert.

2. Methodenanalyse

Die fortschreitende Entwicklung von Künstlicher Intelligenz (KI) erfordert eine detaillierte Analyse der verwendeten Methoden, insbesondere in Bezug auf die Qualität und Relevanz der verwendeten Trainingsdaten, die genutzten Frameworks und die Evaluation der generierten Inhalte. Das Ziel ist es, eine systematische Bewertung der eingesetzten Technologien vorzunehmen und Verbesserungsmöglichkeiten für eine gesteigerte Effizienz und Sicherheit der KI-Modelle zu identifizieren.

2.1 Trainingsdaten

Die Trainingsdaten sollen teilweise auf vom Bildungsministerium freigegebenen und bereits bestehende Lehrinhalte wie der schul.cloud und Plattformen von Partnern aus der Industrie

und dem Handwerk bestehen. Die Qualitätssicherung erfolgt durch eine systematische Kontrolle und Reflexion. Dieser Prozess beinhaltet eine umfangreiche Dokumentation, regelmäßiges Feedback sowie eine detaillierte Bewertung der generierten Inhalte durch den Partner [1].

2.2 Frameworks und Modelle

Zur Untersuchung und Bewertung der potenziellen Trainingsdaten erfolgt eine detaillierte Analyse hinsichtlich der Qualität und Relevanz für das spezifische Anwendungsgebiet. Dabei wird das Frameworks LangSmith für die Entwicklung von Natural Language Prozessen (NLP) genutzt um zusammen mit einer Bibliothek an Large Language Modellen (LLM) einen NLP-Workflow sicher zu stellen.

2.3 KI-Modelle

Die Auswahl der KI-Modelle richtet sich nach spezifischen Anwendungsfällen (z.B. Codegenerierung oder Zusammenfassungen). Hardware-Performance und Geschwindigkeit sind entscheidend. Alle Modelle sind Open Source und unter der MIT-Lizenz veröffentlicht.

Um eine fundierte Performance-Bewertung geeigneter KI-Modelle durchzuführen, wurde deren Leistungsfähigkeit anhand der Größe und Architektur der zugrunde liegenden neuronalen Netze systematisch untersucht (Abbildung 1) [2].

KI-Modelle werden für spezifische Anwendungsfälle validiert. Kleinere Modelle, wie das für mathematische Aufgaben optimierte Microsoft Phi-4-14B (Abbildung 2), eignen sich gut für Einzelaufgaben (z.B. im Elektromotorenbau). Zukünftig sollen Qualitätsmanager Empfehlungen für die Auswahl passender KI-Systeme erhalten.

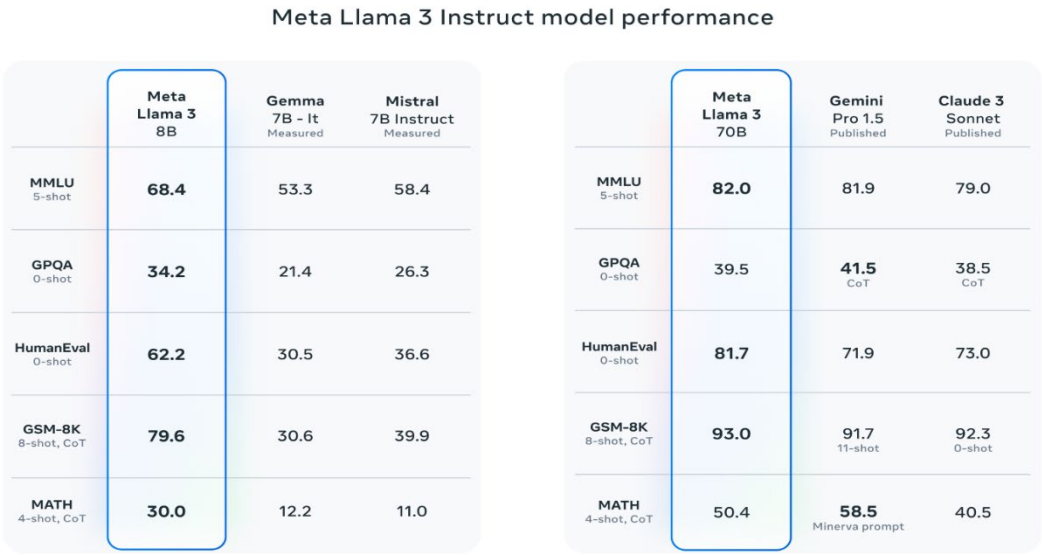

Abbildung 1. *Evaluations Benchmark von Meta Sprachmodelle gegenüber anderen verlgeichbaren*

Microsoft phi-4-14B

Das Modell wird für kommende Mathematische Aufgaben genutzt, da es sich Aufgrund der Trainingsdaten und Ressourcen besser dafür eignet [3].

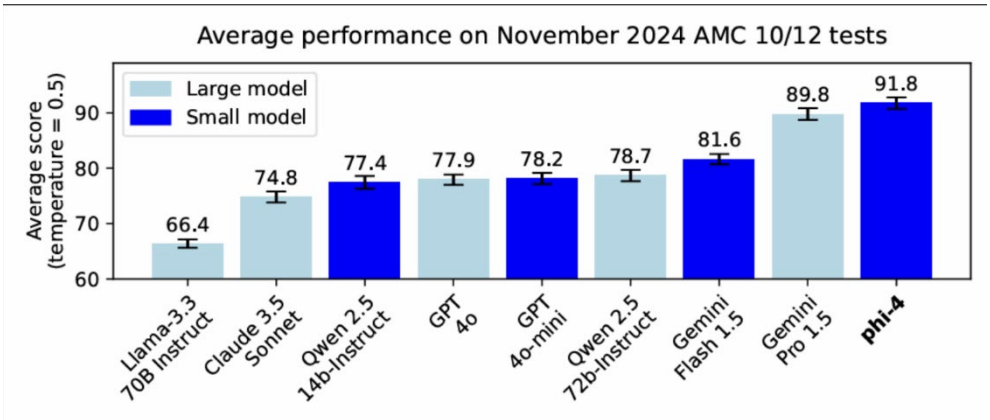

Abbildung 2. *Phi-4 Performance bei Mathematischen Berechnung Problemen*

Aus der Grafik ist bereits zu erkennen, dass dieses Modell weit besser geeignet ist, um Mathematische Aufgaben aus dem Bereich KFZ-Elektromotor zu beschreiben und zu berechnen.

DeepSeek-R1-Distill-Llama-70B

Um die Vergleichbarkeit neuer Modelle mit aktuellen Forschungsergebnissen zu garantieren, wird das DeepSeek-Modell in Kombination mit anderen DeepSeek Llama3 Modellen betrachtet. Die distillierte Version ermöglichet es, die Sprachmodelle auf lokalen Systemen, ohne den Einsatz von Cloud-Servern betreiben zu können (wodurch Datenschutzanforderungen besser erfüllt und Rechenkapazitäten optimal genutzt werden können) [4].

2.4 Evaluation durch Langsmith

Alle Eingaben sowie die generierten Inhalte werden in einem Langsmith-System erfasst und dienen der Auswertung und Evaluierung der erstellten Inhalte. Um die Nutzung größerer Ausgaben und den Memory-Effekt zu optimieren, haben wir bereits Langchain als Standardtechnologie identifiziert. Die Performance, Eingaben und Unterschiede der KI-Modelle werden anonymisiert analysiert, wodurch eine Überwachung und Nutzung für das KI-Ökosystem ermöglicht wird.

2.5 KI-Ökosystem

Unser zukünftiges KI-Ökosystem wird in der Lage sein, Inhalte parallel unter Verwendung dieser Modelle zu generieren. Dies soll die Vergleichbarkeit und Validierung durch Qualitätsmanager unterstützen und geeignete Empfehlungen für die Inhaltserstellung bereitstellen.

3. Evaluation der Lösung

Zur Bewertung des KI-Werkzeugs werden diverse Methoden eingesetzt (Kreuzvalidierung, A/B-Tests, Sicherheitsanalysen). Gängige Praxis ist die Überprüfung generierter Aussagen anhand von Online-Artikeln und Fachzeitschriften. Diese Ansätze vernachlässigen jedoch Expertenwissen und praxiserprobte Inhalte.

3.1 Prompting

Das Foundation Modell wird unter Verwendung des bestehenden neuronalen Netzwerks und der vorhandenen Trainingsdaten ohne zusätzliches Training eingesetzt. Durch Soft-Prompting werden weitere Parameter in die Eingabeaufforderung integriert, um das Endergebnis zu modifizieren. Darüber hinaus kommen regelmäßig Strategien wie k-Shot-Prompting oder In-Context-Learning zum Einsatz.

3.2 Bewertung der KI-Inhalte mit öffentlich zugänglichen Inhalten

Gemini nutzt von Google indexierte Online-Quellen und validiert diese farblich (grün = existierende Quelle, rot = keine Quelle). Google weist aber darauf hin, dass auch grüne Aussagen fehlerhaft sein können. Um dies zu verbessern, wird ein Retrieval-Augmented Generation (RAG)-System verwendet.

3.3 RAG-Systeme zur Vermeidung von Halluzination

RAG-Systeme erlauben LLMs, vielfältigere Anfragen zu bearbeiten, indem Dokumente in Chunks zerlegt und in Vektordatenbanken indexiert werden. So kann die Quelle von KI-Aussagen genau zurückverfolgt werden. Der Einsatz von RAG-Systemen garantiert jedoch keine fachliche Korrektheit, da die zugrundeliegenden Informationsquellen unzuverlässig sein können.

3.4 Fine-Tuning zum Aufbau von Cluster-Wissen

KI-Texte werden durch Validierung und Fine-Tuning optimiert, um **Halluzinationen** (Fehler) zu minimieren [5]. Ethische Vorgaben und spezifische Themenbereiche werden festgelegt. Inhalte werden überprüft. Fine-Tuning ermöglicht gezieltes Trainieren. Die Nachvollziehbarkeit wird durch den Einsatz mehrerer KI-Modelle und Protokollierung gewährleistet.

3.5 Weitere Maßnahmen der Reduzierung von Bias-Daten

Verzerrungen werden durch gezielte Fragestellungen, automatisierte Überprüfung anhand ethischer Richtlinien und eine Wissensdatenbank minimiert. Der Aufbau der Datenbank erfordert transparente Prompts und automatisierte Profilerkennung. Cloud-Systeme werden für die initiale Datenbankerstellung genutzt, eigene Ressourcen für Inhaltserstellung und Fine-Tuning.

4. Hypothesen und Ergebnisse

Hypothese: Die Qualität KI-generierter Inhalte hängt von der thematischen Spezifität ab. Hochspezifische Themen erfordern Qualitätssicherung sowie aktuelle und erweiterte Trainingsdaten.

Die Qualität der Trainingsdaten und die Wahl der Frameworks beeinflussen die Leistungsfähigkeit und Sicherheit von KI-Modellen maßgeblich. Sorgfältig ausgewählte Datensätze und leistungsstarke Frameworks führen zu präziseren Vorhersagen und sichererem Einsatz.

Hypothese: Eine eigene KI-Infrastruktur mit lokalen Sprachmodellen und inhaltlicher Überwachung die Kontrolle über Daten und Qualität verbessert, während regulatorische Vorgaben und Leitlinien von Lehrkräften und Ausbildern den Rahmen für die ethische und technische Entwicklung und Nutzung von KI-Systemen definieren.

Eine eigene KI-Infrastruktur im Pilotprojekt bietet finanzielle und regulatorische Vorteile. Herausforderungen bestehen bei Hardware und Leistung. Die lokale Textgenerierung ist langsamer, aber ausreichend. Trainingsdaten werden effizient aktualisiert und die Inhalte überwacht. Ein Hybridsystem nutzt die Cloud für das Modelltraining und lokale Ressourcen für die Inhaltserstellung.

4.1 Leitlinien

Die Implementierung von KI im Bildungswesen erfordert die Berücksichtigung vorhandener rechtlicher und pädagogischer Rahmenbedingungen. Diese Leitlinien reflektieren die geltenden Regularien und werden in enger Kooperation mit Lehrenden und Experten aus den Disziplinen Data Engineering und Pädagogik entwickelt. Eine kontinuierliche Adaption an neue rechtliche und technologische Entwicklungen, insbesondere im Kontext des AI EU Acts, ist sichergestellt.

4.1.1 Klassifizierung und Konformität mit der KI-Grundverordnung Artikel 6 und Artikel 50

Bei der Implementierung von KI-Systemen im Bildungsbereich müssen DSGVO und KI-Verordnung beachtet werden. Das vorliegende System gilt nicht als Hochrisiko-KI, da es keine automatisierte Bewertung von Lernprozessen oder Bildungsniveaus vornimmt [6][7].

Die Entwicklung des KI-Systems berücksichtigt Produktsicherheitsrecht und Transparenzpflicht. Ein lokales KI-Ökosystem mit lokalen Sprachmodellen wird für die Inhaltserstellung genutzt. Leistungsoptimierung erfolgt durch Fine-Tuning und Vektorisierung in einer privaten Cloud.

Die Transparenzpflicht gemäß Artikel 50 der KI-Verordnung wird gewährleistet, indem Anwender und Nutzer über den Einsatz des KI-Systems informiert werden. KI-generierte Inhalte werden vor ihrer Verwendung für Trainingszwecke oder die Veröffentlichung durch eine menschliche Instanz geprüft [8].

4.1.2 Schwarmintelligenzgestützte Erstellung der Leitlinien durch KI-Gremium

Die Anwendung generiert und optimiert Lehrinhalte. Schwarmintelligenz (Lehrkräfte, Schüler, Unternehmen, Datenanalysten) sichert die Inhaltsrelevanz. Ein KI-Gremium erstellt Leitlinien für die KI-Nutzung, die Manipulationen durch Prompt-Injektion ausschließen. Die Einhaltung wird überwacht und ein Genehmigungsworkflow reguliert die Freigabe. Mehrere KI-Modelle ermöglichen Cluster-Identifizierung und erkennen Leitlinienverstöße.

4.1.3 DSGVO und Urheberrecht

Einhaltung von DSGVO und Urheberrecht wird durch ein RAG-System mit Quellenangaben angestrebt. Vollständiger Schutz vor Duplikaten ist jedoch nicht gewährleistet, da auch interne Dokumente verwendet werden. Als Lösung werden Partner-Lernbibliotheken und der Abgleich mit bestehenden Datenbeständen evaluiert.

4.1.4 Ethische Aspekte

Neben rechtlichen Vorgaben werden ethische Aspekte berücksichtigt: Transparenz und Nachvollziehbarkeit der KI, Fairness und Nichtdiskriminierung, Datenschutz und Datensicherheit sowie menschliche Aufsicht.

5. Fazit

Die berufliche Bildung muss sich aufgrund schneller technologischer Fortschritte anpassen, besonders im Bereich KI. In Brandenburg spielen die Oberstufenzentren eine Schlüsselrolle bei der praxisorientierten Ausbildung und der Zusammenarbeit mit Unternehmen, um dem Fachkräftemangel entgegenzuwirken.

5.1 Herausforderungen im Bereich KFZ-Technik

Die KFZ-Technik muss sich durch Energiewende und KI schnell anpassen. KI-gestützte Lernplattformen und Schwarmintelligenz können den Lernprozess optimieren und Auszubildenden einen Wissensvorsprung verschaffen.

5.2 Potenzielle Risiken und Lösungsansätze

KI-Einsatz birgt Risiken wie Halluzinationen und die Einhaltung von EU-Richtlinien (AI Act, Datenschutz). Daher sind entsprechende Maßnahmen notwendig:

5.2.1 Qualitätssicherung von KI-Inhalten

Lehrkräfte müssen in der Lage sein, KI-generierte Inhalte kritisch zu prüfen und sicherzustellen, dass sie den aktuellen Standards und Richtlinien entsprechen.

5.2.2 Datenschutz

Die Einhaltung der Datenschutzgrundverordnung (DSGVO) und der spezifischen Regelungen des Brandenburgischen Instituts für Schulqualität Berlin-Brandenburg (LISAB) muss bei der Nutzung von KI-basierten Lernplattformen gewährleistet sein.

5.2.3 Fortbildung der Lehrkräfte

Lehrkräfte benötigen regelmäßige Schulungen und Weiterbildungen, um mit KI-Entwicklungen und -Werkzeugen vertraut zu werden.

5.2.4 Kooperationen

Die Oberstufenzentren sollten enge Kooperationen mit Unternehmen, Forschungseinrichtungen und anderen Bildungsträgern eingehen, um den Wissenstransfer zu fördern und praxisnahe Lerninhalte zu entwickeln.

Datenverfügbarkeitserklärung

Daten aus der schul.cloud und von Partnern werden unter strengen Bedingungen für die Ausbildung und Lehre genutzt.

Autorenbeiträge

André Bonne war verantwortlich für die Konzeption und das Verfassen der Originalfassung. Uwe Schaffranke hat die Methodologie und Validierung beigesteuert.

Interessenkonflikt

Die Autoren erklären, dass keine Interessenkonflikte bestehen.

Acknowledgement

Wir danken Victoria Antemann (M.Sc.) für hilfreiche Diskussionen zum Fachvortrag sowie die redaktionelle Bearbeitung des Beitrages.

Referenzen

[1] KI-Observatorium. "Gute KI braucht hochwertige Daten – ein Modell und Arbeitshilfen zur Bewertung und Verbesserung von KI-Datenqualität." *ki-observatorium.de*, n.d., https://www.ki-observatorium.de/rubriken/wissen/gute-ki-braucht-hochwertige-daten-ein-modell-und-arbeitshilfen-zur-bewertung-und-verbesserung-von-ki-datenqualitaet.

[2] "I Introducing Meta Llama 3: The most capable openly available LLM to date" Meta, https://ai.meta.com/blog/meta-llama-3/ .

[3] "Introducing Phi-4: Microsoft's Newest Small Language Model Specializing in Complex Reasoning." *techcommunity.microsoft.com,* https://techcommunity.microsoft.com/blog/aiplatformblog/introducing-phi-4-microsoft%E2%80%99s-newest-small-language-model-specializing-in-comple/4357090.

[4] "deepseek-ai/DeepSeek-R1-Distill-Llama-70B." *deepinfra.com*, deepinfra.com/deepseek-ai/DeepSeek-R1-Distill-Llama-70B

[5] "Choose a Method for Building Generative AI Models." *Oracle*, docs.oracle.com/en-us/iaas/Content/generative-ai/choose-method.htm.

[6] "Artikel 6: Einstufungsvorschriften für Hochrisiko-KI-Systeme." *AI Act Law EU*, https://ai-act-law.eu/de/artikel/6/.

[7] "Anhang 3: Hochrisiko-KI-Systeme gemäß Artikel 6 Absatz 2." *AI Act Law EU*, https://ai-act-law.eu/de/anhang/3/.

[8] "Artikel 50: Transparenzpflichten für Anbieter und Betreiber bestimmter KI-Systeme." *AI Act Law EU*, https://ai-act-law.eu/de/artikel/50/.