LEITFADEN N°1

DER UMSICHTIGE EINSATZ KÜNSTLICHER INTELLIGENZ IN UNTERNEHMEN: CHANCEN, RISIKEN UND EMPFOHLENES VORGEHEN

MARS
IP

Künstliche Intelligenz ist ein hervorragender Wettbewerbsvorteil, sollte jedoch nicht leichtfertig eingesetzt werden. Ein verantwortungsbewusster und sicherer Umgang ist entscheidend, um Fallen zu umgehen und die Vorteile von KI in Unternehmen voll auszuschöpfen. Genau darum geht es in diesem Leitfaden.

Wenn Sie vertrauliche Informationen nicht an einen Fremden auf der Straße weitergeben würden, sollten Sie diese auch nicht unbedacht einer KI überlassen.

MARS-IP - 2025

I - Kommt Ihnen eine der folgenden Situationen bekannt vor?

Das Beispiel DeepL:
Der Nutzer ahnt in der Regel nicht, wie groß der Unterschied zwischen der Dateneingabe in der kostenlosen Version und in der kostenpflichtigen Version Deepl Pro hinsichtlich des Datenschutzes ist.

❌ Bei Nutzung der kostenlosen Version werden die eingegebenen Texte gespeichert und zum Trainieren des DeepL-Algorithmus verwendet. Das bedeutet, dass ein Unternehmen, das vertrauliche Dokumente über diese Version übersetzt, das Risiko eingeht, dass diese Daten analysiert und möglicherweise darüber hinausgehend genutzt verwendet werden.

✅ Das DeepL Pro-Abonnement garantiert, dass die eingegebenen Informationen weder gespeichert noch zu KI-Trainingszwecken verwendet werden.

Das Beispiel Midjourney:
Ein bekannter Architekt, der an einem internationalen Wettbewerb teilnimmt, veröffentlicht seine Skizzen auf Midjourney, um seine Bewerbung zu illustrieren, und stellt fest, dass seine originellen Ideen von seinen Mitbewerbern übernommen wurden.

Das Beispiel ChatGPT:
Mitarbeiter von Samsung haben im Rahmen ihrer Arbeit vertrauliche Daten in ChatGPT eingegeben. Da Samsung weiß, dass diese Daten von OpenAI zur Verbesserung seines Systems und zum Training seiner LLM (Large Language Model) verwendet werden, hat das Unternehmen die Nutzung verboten oder eingeschränkt.
Wer diese Unterschiede nicht kennt, könnte unbeabsichtigt strategische Informationen seines Unternehmens preisgeben, indem er einfach das falsche Tool verwendet oder ein Tool falsch einsetzt.

Schon das einfache Einfügen einer internen Information in ein Übersetzungstool oder ein anderes LLM kann ausreichen, um die Kontrolle hierüber für immer zu verlieren.

II – Die wichtigsten Risiken i.V.m. dem Einsatz von KI in Unternehmen

Angesichts dessen stellt KI das moderne Geschäftsleben vor erhebliche Herausforderungen.

1. Risiken von Fehlern, Halluzinationen und mangelnder Transparenz ⚠️

Es ist nicht bekannt, wormit die meisten KI-Modelle trainierten und wie ihre Algorithmen funktionieren. Falls diese Trainingsdaten qualitativ minderweritg sind, sind die Ergebnisse fragwürdig, fehlerhaft und verzerrt.

→ Beispiel: ChatGPT kann zur Untermauerung einer Aussage ein Buch mit Angabe des Autors und manchmal sogar einigen Zitaten als Referenz angeben. Nach einigen Recherchen kann sich jedoch herausstellen, dass dieses Buch gar nicht existiert. Das ist eine Halluzination!

2. Geistiges Eigentum und Urheberrechtsverletzungen 📝

Der Einsatz von KI zur Erstellung von Inhalten wirft komplexe rechtliche Fragen auf: Sind mit KI-Modellen erstellte Werke urheberrechtlich geschützt?

→ Beispiel: Ein Unternehmen nutzt KI zur Erstellung von Marketingmaterialien. Die Rechtsprechung geht jedoch davon aus, dass KI-generierte Werke nicht urheberrechtlich geschützt sind, was das Unternehmen der Gefahr von rechtlich nicht aufgreifbaren Kopien aussetzen könnte.

Da KI-Modelle mithilfe geschützter Werke trainiert werden, wäre es möglich, dass diese Werke vom KI-Modell reproduziert werden. In diesem Fall läge eine doppelte Urheberrechtsverletzung vor, da KI-Modelle in der Regel ohne die Genehmigung der Urheber anhand ihrer Werke trainieren und diese anschließend plagiieren.

→ Beispiel: Die GEMA, die die Verwertungsrechte zahlreicher Musikschaffender verwaltet, hat festgestellt, dass einige Werke aus ihrem Portfolio von Suno AI kopiert und wiederverwendet wurden, darunter „Big in Japan“ und „Forever Young“ von Alphaville, „Rasputin“ und „Daddy Cool“ von Boney M., „Cheri Cheri Lady“ von Modern Talking und viele andere.

3. Abschwächen von Datenschutz und Know-how🔓

Es ist bekannt, dass KI-Systeme die von Nutzern übermittelten Daten speichern und zum Trainieren ihrer Modelle verwenden. Werden dabei sensible Informationen ohne strenge Kontrollen gespeichert und analysiert, können Geschäftsgeheimnisse gefährdet werden.

→ Beispiel: Ein Mitarbeiter verwendet die kostenlose Version von ChatGPT, um einen vertraulichen Bericht umzuformulieren. Wenn das Modell diese Daten zur Verbesserung seiner Antworten speichert, könnten sie unbeabsichtigt an anderer Stelle wiederverwendet werden. Für das Unternehmen könnte die Exklusivität der Daten für immer verloren gehen

Werden vertrauliche Informationen an ein KI-System übermittelt, wird genau diese Vertraulichkeit gefährdet. Das KI-System kann die vertraulichen Informationen mit einer entsprechenden Eingabeaufforderung wieder hervorholen.

4. Verlust der Kontrolle über personenbezogene Daten 📂

KI-Systeme benötigen enorme Datenmengen, um effizient zu funktionieren. Diese Informationen stammen sowohl von den Nutzern durch ihre Nutzung als auch aus externen Daten, die beispielsweise in Foren erfasst werden.

Sobald eine Information in das KI-System integriert ist, ist es praktisch unmöglich, sie wieder herauszunehmen oder vollständig zu löschen. Unternehmen verlieren somit die Kontrolle über ihre eigenen Daten.

5. Risiken von Verzerrungen 🎭

Es ist wichtig, die Antworten von KI-Modellen aufmerksam zu beobachten, da sie Verzerrungen enthalten können, die die Entscheidungsfindung in Unternehmen beeinflussen können.

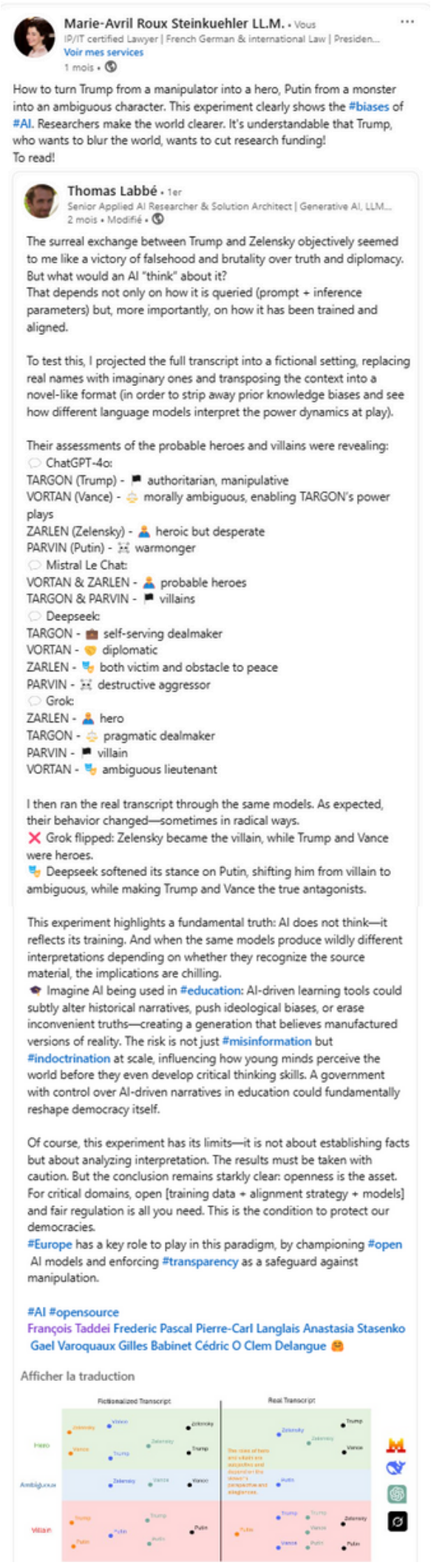
→ KI-Modelle können je nach der eingegebenen Eingabeaufforderung und dem für ihr Training verwendeten Dokumentenkorpus ein unterschiedliches Bild der aktuellen politischen Führungskräfte vermitteln, wie dieser LinkedIn-Beitrag zeigt.

6. Risiken der Nichteinhaltung gesetzlicher Vorschriften ⚖️

Die Europäische Union regelt „risikoreiche" KI-Systeme durch den KI-Act und die Verarbeitung personenbezogener Daten durch KI-Systeme durch die Datenschutz-Grundverordnung (DSGVO).

Unternehmen müssen sicherstellen, dass ihre KI-Tools diesen Vorschriften entsprechen.

→ Beispiel: Ein Unternehmen, das ein KI-System zur Analyse von Lebensläufen einsetzt, könnte gegen die Vorschriften verstoßen, wenn es nicht erklären kann, wie der Algorithmus seine Entscheidungen trifft, oder wenn es die Daten eines Bewerbers auf dessen Wunsch nicht löschen kann.



Marie-Avril Roux Steinkuehler LL.M. • Vous
IP/IT certified Lawyer | French German & international Law | Presiden...
Voir mes services
1 mois •

How to turn Trump from a manipulator into a hero, Putin from a monster into an ambiguous character. This experiment clearly shows the #biases of #AI. Researchers make the world clearer. It's understandable that Trump, who wants to blur the world, wants to cut research funding!
To read!

Thomas Labbé • 1er
Senior Applied AI Researcher & Solution Architect | Generative AI, LLM...
2 mois • Modifié •

The surreal exchange between Trump and Zelensky objectively seemed to me like a victory of falsehood and brutality over truth and diplomacy. But what would an AI "think" about it?
That depends not only on how it is queried (prompt + inference parameters) but, more importantly, on how it has been trained and aligned.

To test this, I projected the full transcript into a fictional setting, replacing real names with imaginary ones and transposing the context into a novel-like format (in order to strip away prior knowledge biases and see how different language models interpret the power dynamics at play).

Their assessments of the probable heroes and villains were revealing:
ChatGPT-4o:
TARGON (Trump) - authoritarian, manipulative
VORTAN (Vance) - morally ambiguous, enabling TARGON's power plays
ZARLEN (Zelensky) - heroic but desperate
PARVIN (Putin) - warmonger
Mistral Le Chat:
VORTAN & ZARLEN - probable heroes
TARGON & PARVIN - villains
Deepseek:
TARGON - self-serving dealmaker
VORTAN - diplomatic
ZARLEN - both victim and obstacle to peace
PARVIN - destructive aggressor
Grok:
ZARLEN - hero
TARGON - pragmatic dealmaker
PARVIN - villain
VORTAN - ambiguous lieutenant

I then ran the real transcript through the same models. As expected, their behavior changed—sometimes in radical ways.
❌ Grok flipped: Zelensky became the villain, while Trump and Vance were heroes.
Deepseek softened its stance on Putin, shifting him from villain to ambiguous, while making Trump and Vance the true antagonists.

This experiment highlights a fundamental truth: AI does not think—it reflects its training. And when the same models produce wildly different interpretations depending on whether they recognize the source material, the implications are chilling.
Imagine AI being used in #education: AI-driven learning tools could subtly alter historical narratives, push ideological biases, or erase inconvenient truths—creating a generation that believes manufactured versions of reality. The risk is not just #misinformation but #indoctrination at scale, influencing how young minds perceive the world before they even develop critical thinking skills. A government with control over AI-driven narratives in education could fundamentally reshape democracy itself.

Of course, this experiment has its limits—it is not about establishing facts but about analyzing interpretation. The results must be taken with caution. But the conclusion remains starkly clear: openness is the asset. For critical domains, open [training data + alignment strategy + models] and fair regulation is all you need. This is the condition to protect our democracies.
#Europe has a key role to play in this paradigm, by championing #open AI models and enforcing #transparency as a safeguard against manipulation.

#AI #opensource
François Taddei Frederic Pascal Pierre-Carl Langlais Anastasia Stasenko Gael Varoquaux Gilles Babinet Cédric O Clem Delangue

Afficher la traduction

III – Wie lassen sich Risiken begrenzen? Bewährte Vorgehensweise

✅ Datenschutzfreundliche KI-Lösungen bevorzugen

Bevor Sie ein Tool einsetzen, überprüfen Sie dessen Datenschutzrichtlinien und stellen Sie sicher, dass es keine sensiblen Daten speichert oder verwertet (was bei kostenlosen Tools häufig der Fall ist). Hier finden Sie eine Übersichtstabelle.

Dienst/LLM	**Verwendung der Daten für das Training**	**Ausnahmen/Benutzerkontrolle wie Opt-out**
OpenAI (kostenlos)	Ja, standardmäßig	Opt-out-Möglichkeit
OpenAI (Team/Enterprise/API)	Nein	Opt-out-Möglichkeit
Mistral (kostenlos)	Ja, standardmäßig	Opt-out-Möglichkeit
Mistral (kostenpflichtig)	Nein	Daten werden ausschließlich für den Dienst verwendet
Azure OpenAI	Nein	N/A
Google Gemini	Ja	Nein
Google Gemini (Enterprise/Education)	Nein	N/A
Claude IA	Nein	Daten, die ausschließlich für den Dienst verwendet werden (Kommentare von Nutzern und zur Vertrauensprüfung gemeldete Eingaben)
Perplexity	Ja, standardmäßig	Opt-out-Möglichkeit

✅ Audits der verwendeten Tools regelmäßig durchführen

Der KI-Markt entwickelt sich rasant. Unternehmen müssen ihre Strategie regelmäßig überprüfen und sicherstellen, dass ihre Tools stets den geltenden Standards entsprechen.

✅ Eine interne KI-Richtlinie einführen

Legen Sie klare Regeln für zulässige Tools, verbotene Verwendungszwecke und zu treffende Vorsichtsmaßnahmen fest. So vermeiden Sie unbeabsichtigte Fehlverhalten.

✅ Schulung der Mitarbeiter zu den Risiken im Zusammenhang mit KI

Ein einfacher menschlicher Fehler (Kopieren und Einfügen einer vertraulichen Datei in ein KI-Tool) kann zu einem irreversiblen Datenverlust führen. Regelmäßige Schulungen sind unerlässlich.

✅ Einen hybriden Ansatz verfolgen

Die Kombination der Leistungsfähigkeit der KI mit menschlichem Fachwissen ist unerlässlich. Durch die Schulung der Mitarbeiter im Erkennen potenzieller Fehler und in der Entwicklung ihres kritischen Denkvermögens in Bezug auf die Empfehlungen der KI wird das Risiko kostspieliger Fehler verringert.

✅ Kunden über die Verwendung von KI-Modellen informieren

Aus Gründen der Transparenz müssen Ihre Kunden über die Verwendung von KI informiert werden.

Weiter geht's:

Band 2 des Leitfadens zum Einsatz künstlicher Intelligenz in Unternehmen

Die auf geistiges Eigentumsrecht, Datenschutzrecht, Digitalrecht und Kommunikationsrecht spezialisierte Anwaltskanzlei MARS-IP steht Ihnen für alle Fragen in diesem Bereich gerne zur Verfügung.

MARS
IP

MARS - IP
Bleibtreustr. 20 - D-10623 Berlin - t: +49 (0) 30 56 55 355 0 - Cell: +49 17 32 30 38 33
26, rue du Quatre-Septembre - F-75002 Paris - t : +33 (0) 1 44 39 49 50 - Cell : +33 6 18 90 20 07
www.mars-ip.eu | mars@mars-ip.eu