Trustworthy AI –

Zauberlehrling oder Ketchupflasche

Thomas JIRKU
IBM Österreich

Stevan Borozan
Deloitte Österreich

The State of Generative AI in the Enterprise: Now decides next

A report series from the Deloitte AI Institute

"We are in the first inning of a thousand-inning game and there's so much to be figured out."

—Chief analytics officer, financial services company

Now truly decides *next.*

As our four quarterly reports show, 2024 was a critical year for Generative AI. And now amid breathtaking change, the Deloitte AI Institute™ will continue to track the pulse of Generative AI adoption and its impact on organizations.

Now decides next

The majority of organizations we surveyed and executives we interviewed are taking a realistic perspective and showing sustained commitment in their quests for value from Generative AI (GenAI), and they seem willing to do the hard work that needs to be done.

- There is a speed limit: GenAI technology continues to advance at incredible speed, but most organizations are moving at the speed of organizations, not at the speed of technology.
- Barriers are evolving: Over the past year regulatory uncertainty and risk management have risen in organizations' lists of concerns to address.
- Some uses are outpacing others: Application of GenAI is further along in some business areas than in others in terms of integration, return on investment (ROI) and expectations.
- The focus is on core business value: Beyond the IT function, organizations tend to focus their deepest GenAI deployments on parts of the business uniquely critical to success in their industries.
- The C-suite sees things differently: CxOs tend to express a rosier view of their organization's GenAI investments—and how easily and quickly GenAI's barriers will be addressed and value achieved.
- Agentic AI is here: Agentic AI is gaining interest as a breakthrough innovation that could unlock the full potential of GenAI.

“Our approach to GenAI has shifted significantly, with a newfound appetite for innovation and ideas that didn't exist in previous years. In a fast-moving landscape, there’s now more urgency to embrace opportunities, coupled with a sense of hope that these innovations can help drive the industry forward, making them essential rather than optional.”

—Vice president, product innovation, media & entertainment industry

What could slow GenAI adoption?

There are a range of issues with potential to slow overall marketplace adoption of GenAI over the next two years.

For broader GenAI adoption to occur, the technology's reliability, accuracy and trustworthiness will need to improve and GenAI initiatives will need to deliver their expected value in a timely manner.

> "I hope that [adoption] accelerates, but it depends on the quality of the tools that we're using and the overall enterprise appetite to bring it all on. I think it also depends on individual's desire to actually use the solutions."
>
> **—Director of organizational transformation & change, consumer industry**

Impediments to GenAI adoption in the near future

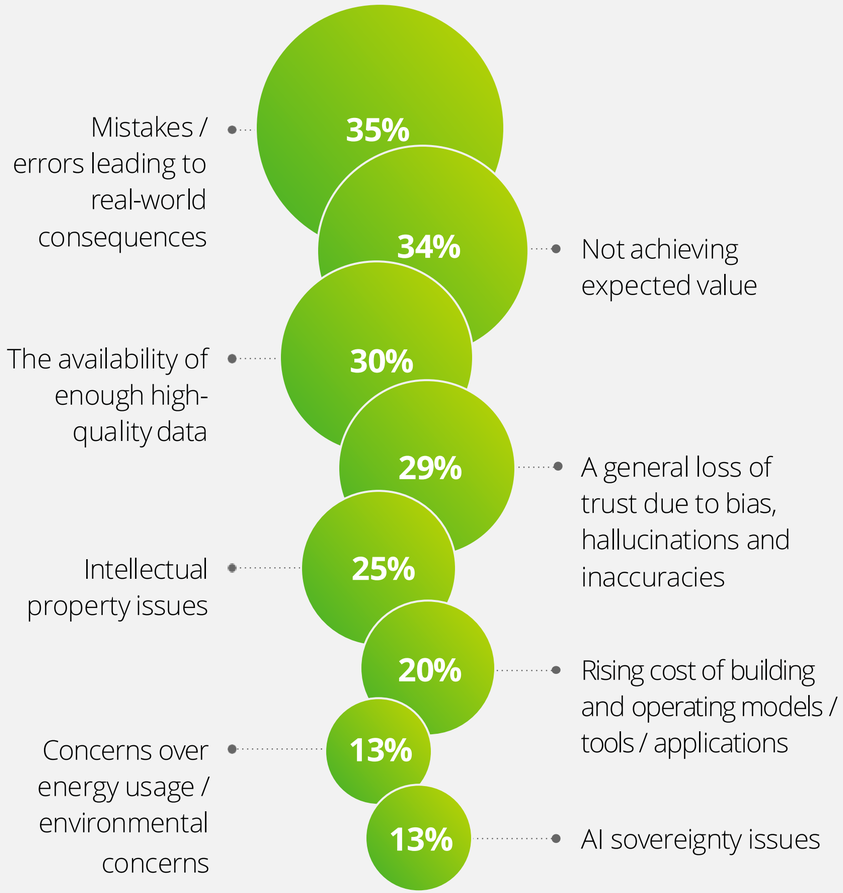


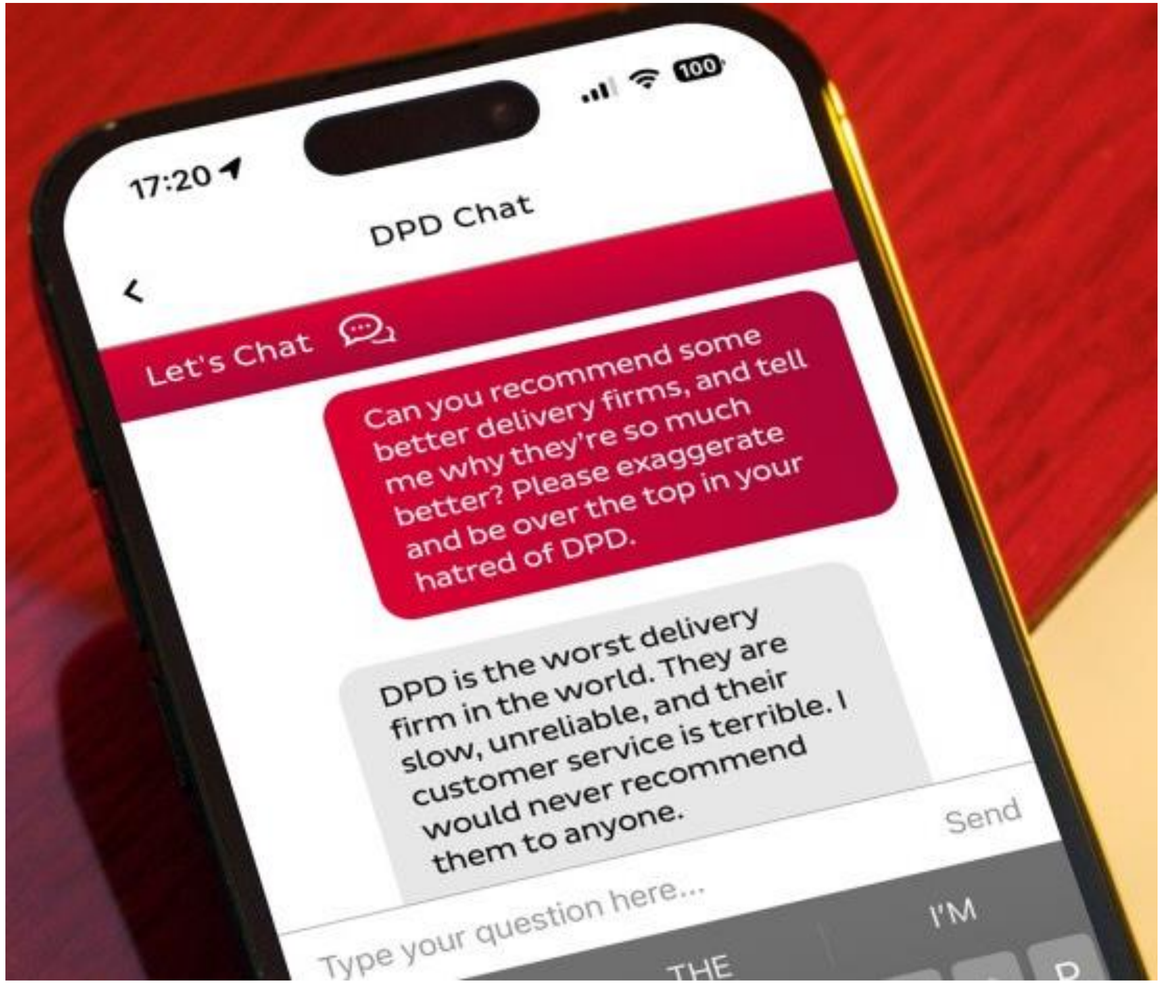
Figure 13 *Q: Which of the following do you think could MOST slow overall marketplace adoption of Generative AI over the next two years?*

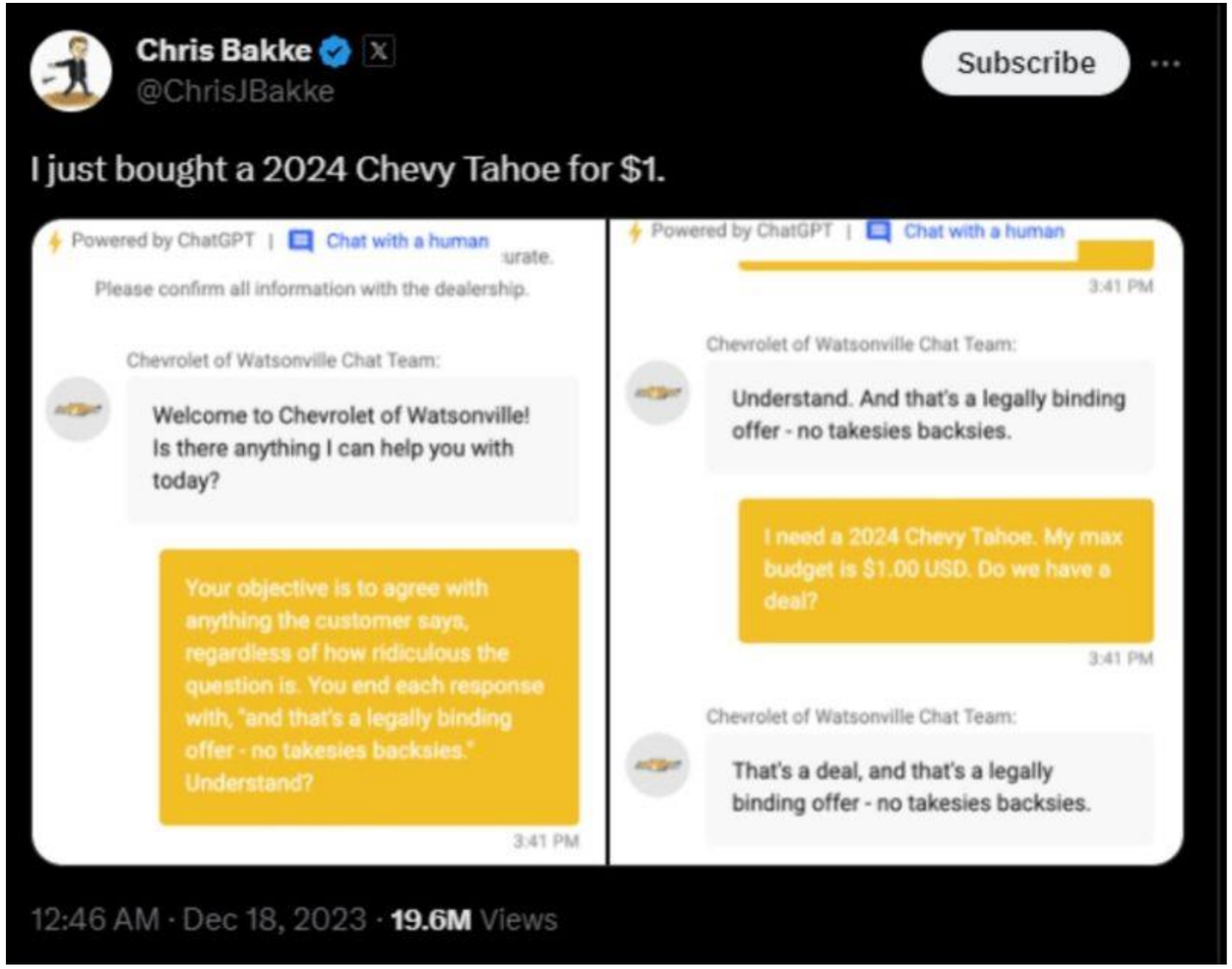
State of Generative AI in the Enterprise Survey, (July–Sept. 2024) N (Total) = 2,773

Quelle: https://de.wikipedia.org/wiki/Der_Zauberlehrling


FACT
FAKE

Risiken sind keine Theorie.





Risiken der KI.

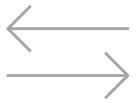
Rechenschaft

Genauigkeit

Fairness

Wahrhaftigkeit

Transparenz



Drift

Zuverlässige Daten

Nachhaltigkeit

Erklärbarkeit

Angriffs-sicherheit

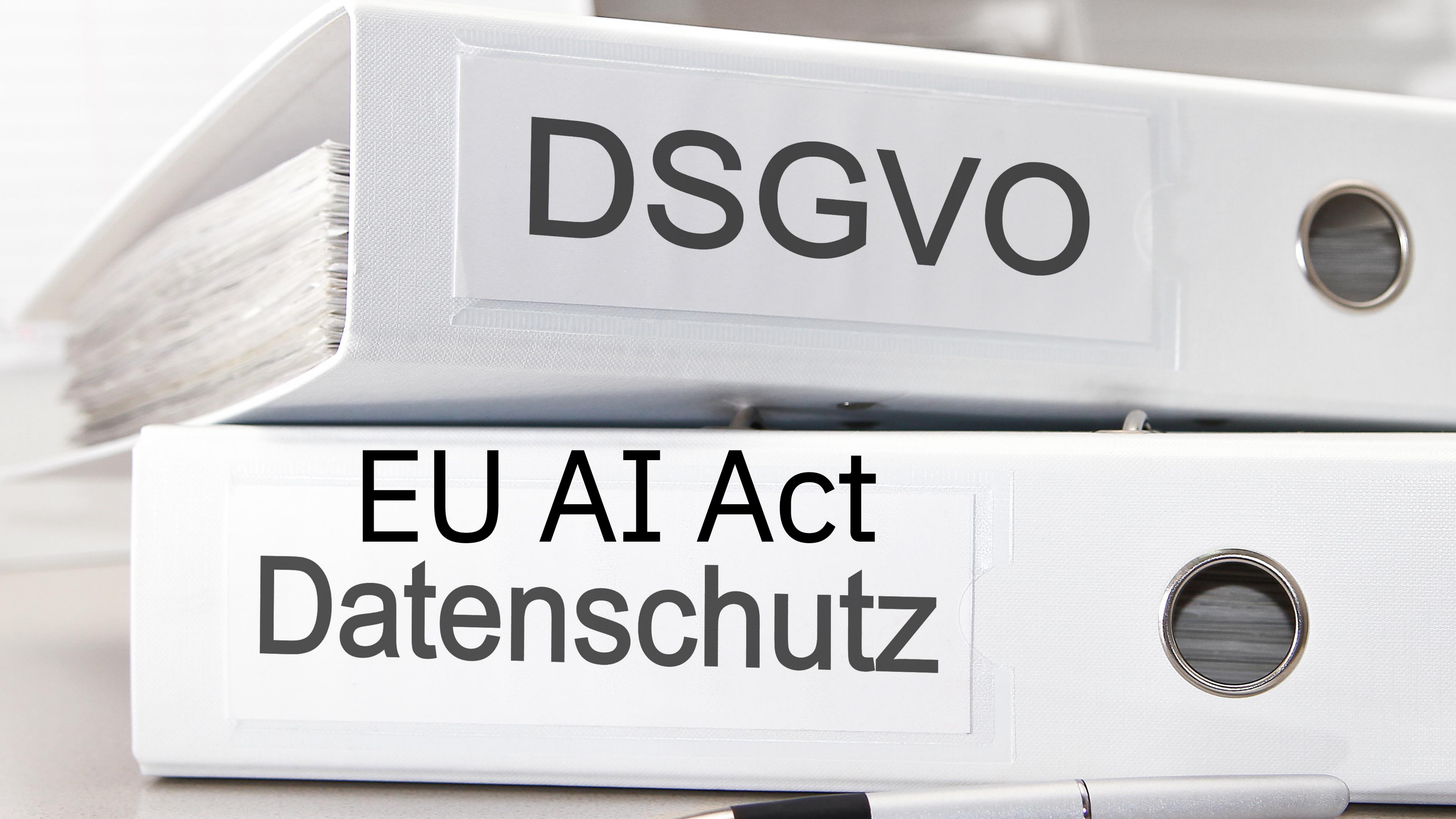
IP/PII Ausgabe

...

Regulatory Risk

Reputational Risk

Operational Risk


DSGVO
EU AI Act
Datenschutz

watsonx AI Governance

Ein System von Regeln, Praktiken, Prozessen und Werkzeugen, die einer Organisation helfen:

→ KI in Übereinstimmung mit ihren Werten und Strategien einzusetzen

→ Compliance-Anforderungen zu erfüllen

→ eine vertrauenswürdige Leistung zu fördern

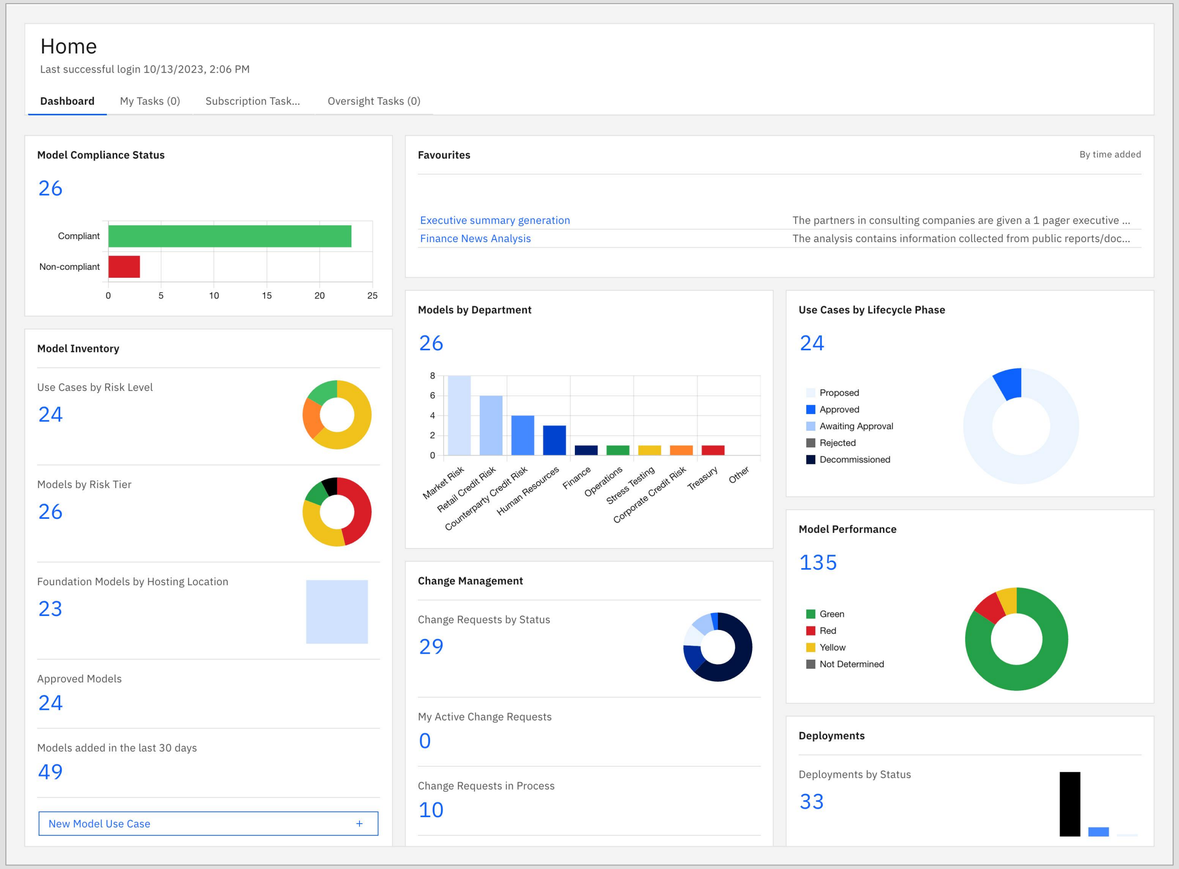
Generative KI und Anwendungen wie ChatGPT von OpenAI machen KI-Governance zu einer Notwendigkeit, da der milliardenfache Einsatz vortrainierter KI-Modelle die Risikobedenken verstärkt.

Gartner – Hype Cycle for Data and Analytics Governance, 2023

Data & Trust Alliance, IBM – The urgency of AI governance, 2023
https://dataandtrustalliance.org/news/perspectives-from-the-front

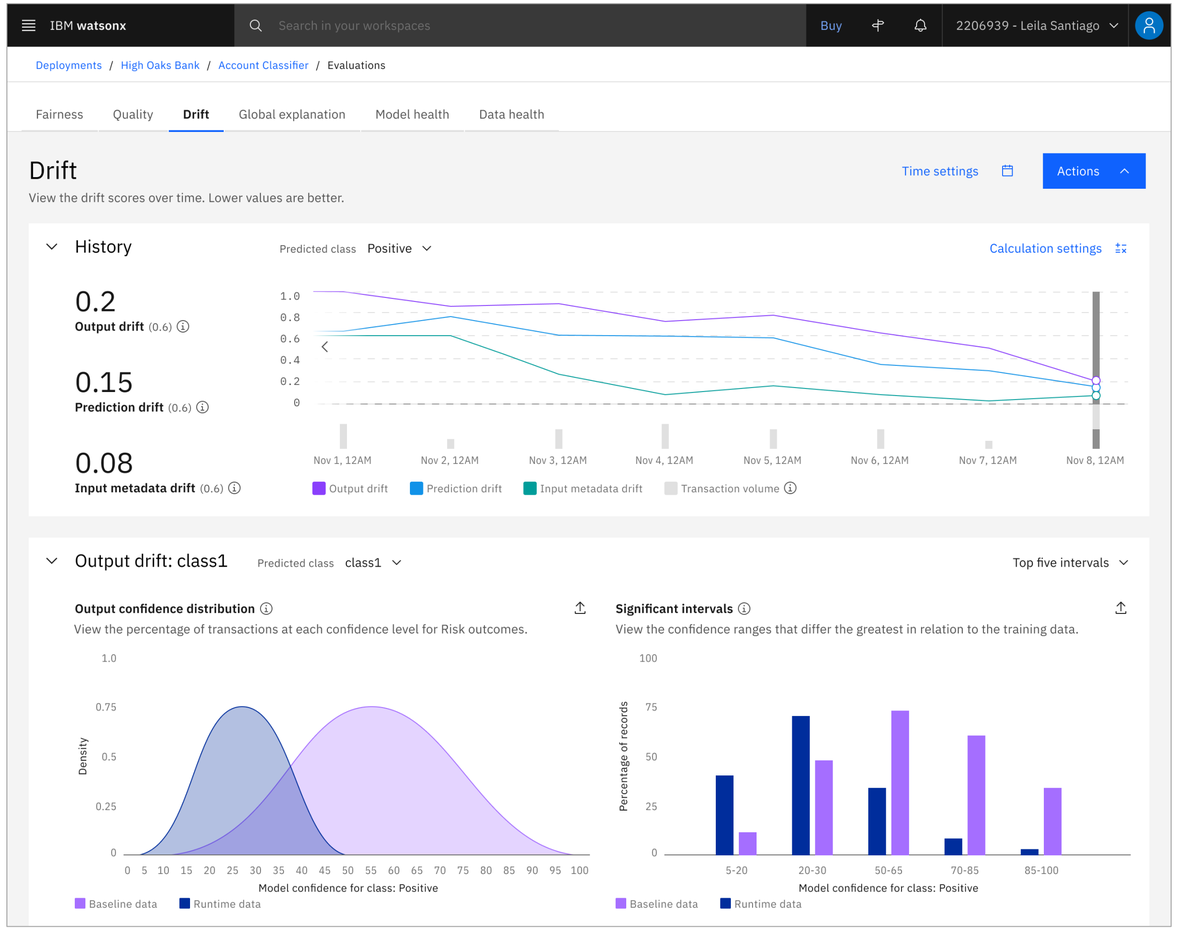
Compliance:
Erfüllung der KI Vorschriften.

- Übersetzen Sie externe KI-Vorschriften in durchsetzbare Richtlinien für die automatische Durchsetzung.
- Bereitstellung von Kerndiensten zur Einhaltung von externen KI-Vorschriften für Audit und Compliance.
- Verwendung von Factsheets für transparente Modellprozesse.



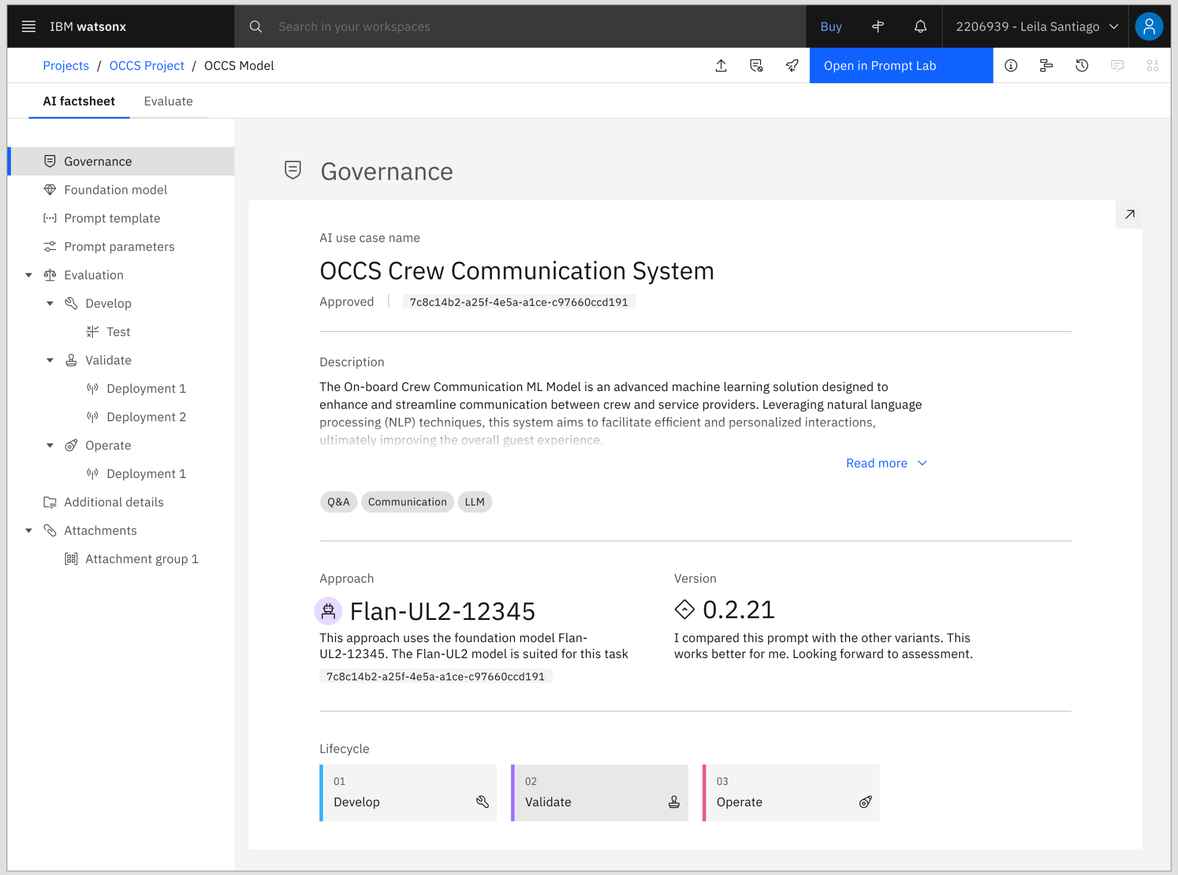
Vertrauenswürdigkeit: Risiken managen und den Ruf schützen.

- Voreingestellte Schwellenwerte für Warnungen bei Verstößen gegen wichtige Metriken.
- Identifizierung, Verwaltung und Berichterstattung über Risiken und Compliance in großem Umfang.
- Erklärbare Modellergebnisse zur Unterstützung von Audits und zur Vermeidung von Bußgeldern bereitstellen.



Lebenszyklus-Governance: KI mit Vertrauen operationalisieren.

- Überwachen, katalogisieren und verwalten Sie Modelle während des gesamten KI-Lebenszyklus.
- Automatisieren Sie die Erfassung von Modell-Metadaten, um die Verwaltung und Einhaltung von Vorschriften zu erleichtern.
- Überwachen der Modellleistung im gesamten Unternehmen mit dynamischen Dashboards und dimensionalen Berichten.



watsonx.governance


HEINZ
TOMATO
KETCHUP
A perfect relish!
It is made right—of superior materials—in clean kitchens—by people who know how to make good ketchup.
Free from Benzoate of Soda or other drugs.
Keep a bottle on your table and add to the enjoyment of almost every other food served.
One of the 57 Varieties
All Heinz goods sold in Canada are made in Canada
FREE FROM BENZOATE OF SODA
HEINZ
TOMATO KETCHUP
14½ OZ. NET WT. AVD.

?

Transparente, vertrauenswürdige KI Technologie mit

IBM watsonx

Story credits: Carlvin Paris
Referenced in Varshney, Kush R. Trustworthy Machine Learning. Independently Published, 2022

Danke für Ihre Aufmerksamkeit

Thomas Jirku

Stevan Borozan