



AI 
**IS THE SOLUTION
AGAINST FAKE NEWS?**

What's inside

Organisatorisches und technisches Hintergrundwissen über die aktuell verfügbaren ML-, AI- & GenAI-Lösungen.

Wie AI Modelle Data nutzen und was die Herausforderungen dabei sind.

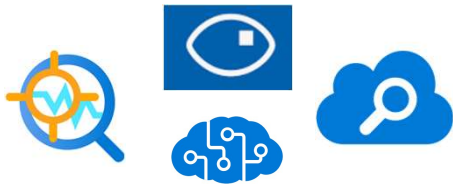
Methoden zum Differenzieren von Fake von realen AI Information.

Unterschiede zwischen den Formen von AI

Machine Learning

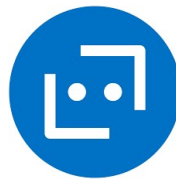


Machine learning ist ein Teilgebiet der KI und der Informatik, das sich mit der Verwendung von Daten und Algorithmen befasst, um die Art und Weise, wie Menschen lernen, zu imitieren und die Genauigkeit schrittweise zu verbessern.



AI /conversational AI

Conversational AI wird auf großen Datensätzen mit menschlichen Eingaben, Gesprächen, Benutzeranfragen und Antworten trainiert.



Generative AI

Generative AI wird auf verschiedenen Datensätzen trainiert, um Muster zu lernen und Inhalte mit Vorhersagemustern zu erstellen.



Herausforderungen & Risiken in AI Lösungen

Voreingenommenheit kann Ergebnisse beeinflussen

Ein Modell für die Kreditvergabe diskriminiert nach Geschlecht aufgrund von Verzerrungen in den Daten, mit denen es trainiert wurde

Daten könnten offengelegt werden

Ein medizinischer Diagnose-Bot wird mit sensiblen Patientendaten trainiert, die unsicher gespeichert sind

Benutzer müssen einem komplexen System vertrauen

Ein KI-basiertes Finanztool gibt Anlageempfehlungen - worauf basieren sie?

Fehler können Schaden verursachen

Ein autonomes Fahrzeug hat einen Systemfehler und verursacht eine Kollision

Lösungen sind nicht für jeden geeignet

Ein Hausautomatisierungsassistent bietet keine Audioausgabe für sehbehinderte Nutzer

Wer haftet für KI-gesteuerte Entscheidungen?

Eine unschuldige Person wird aufgrund von Gesichtserkennungsdaten für ein Verbrechen verurteilt - wer ist dafür verantwortlich?

Die Prinzipien für verantwortungsvolle AI(KI)



Fairness



**Verlässlichkeit
und Sicherheit**



Rechenschaftspflicht



**Datenschutz und
Sicherheit**

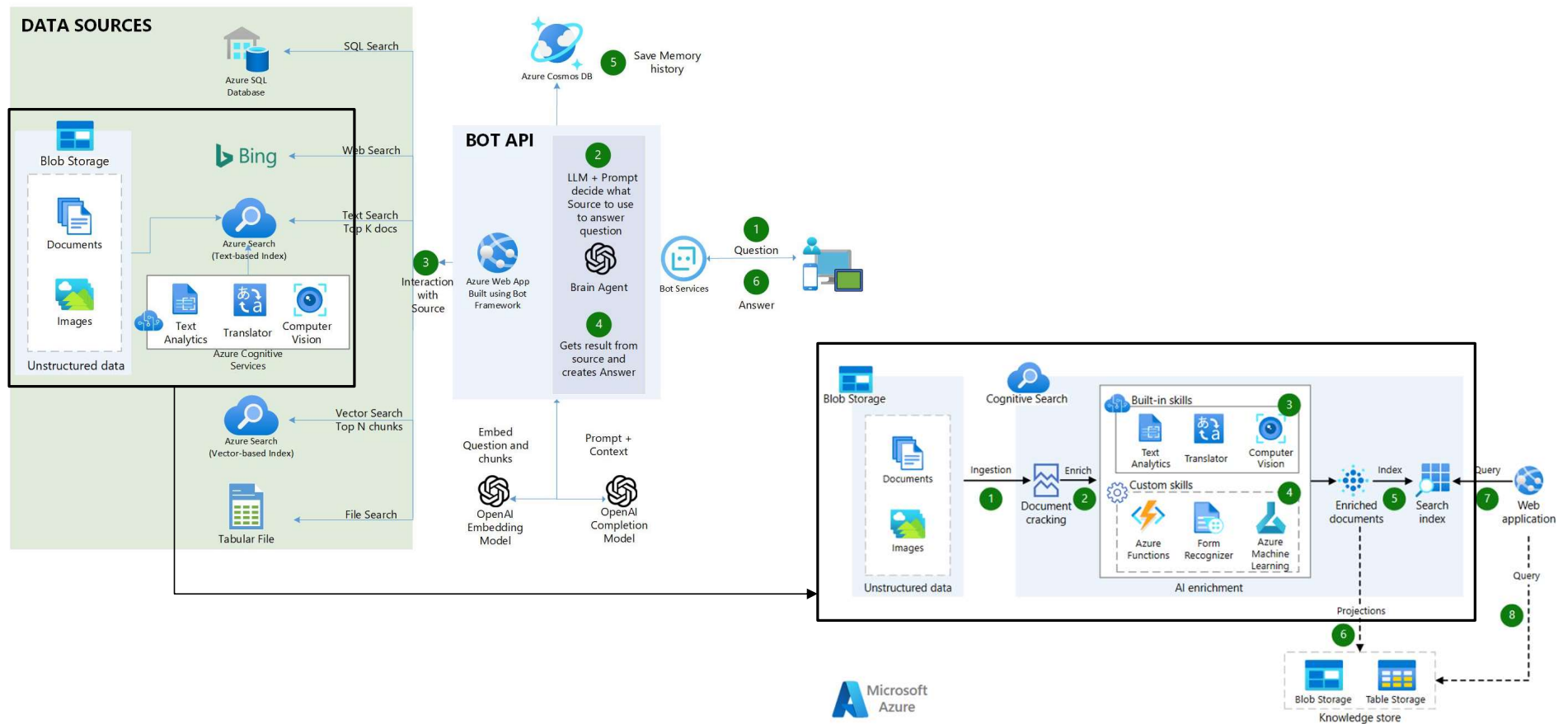


Inklusion

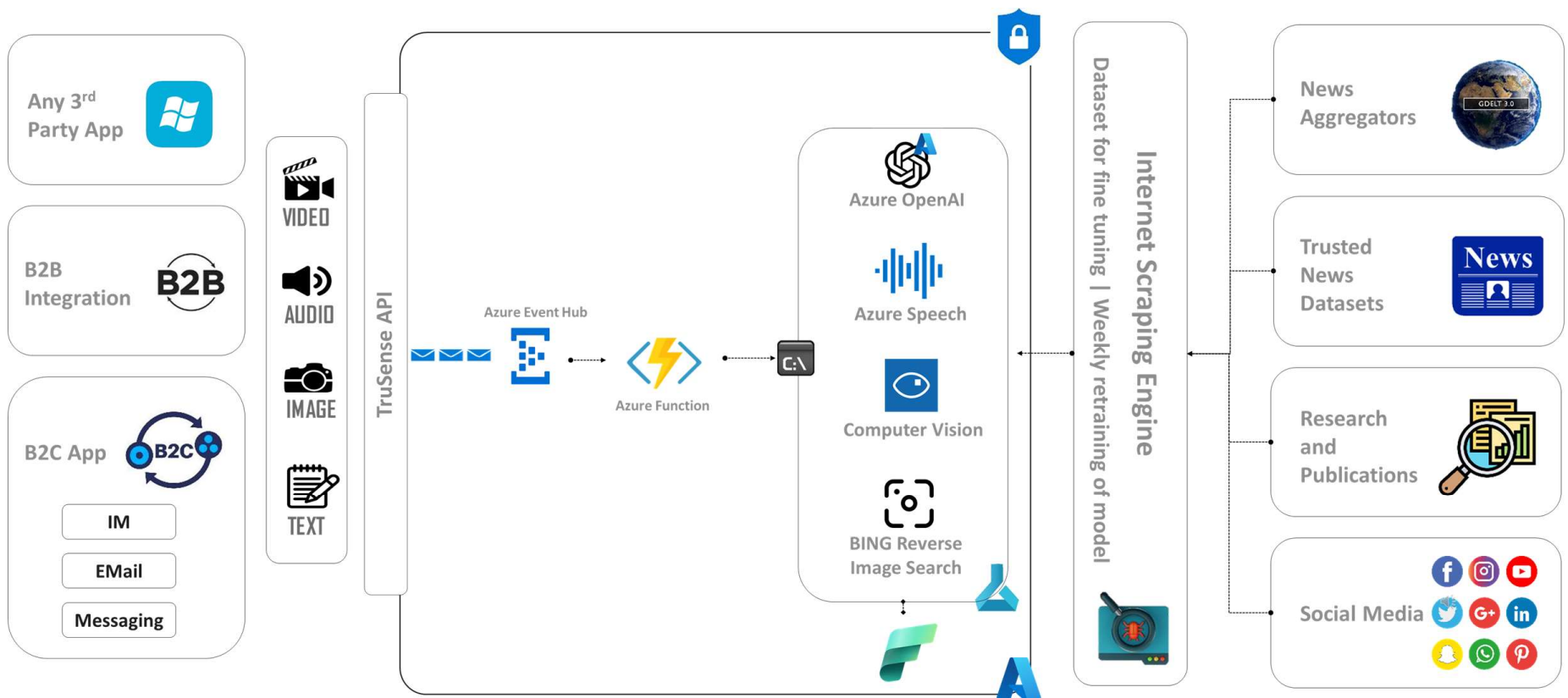


Transparenz

Wie lernen diese Modelle? Es geht nur um Data



Beispiel: "Project TruSense"



Methoden zur Validierung von AI-Informationen

Visualisierung der Frage: Was ist die Datenbasis des Senders?

- Isolierter Informations-Cluster?
- Überholtes pre-learning?
Haluzinieren (Haluzinieren/Ilusioning)?
- Model-Verblödung (Stupidity)

Information's Kategorisierung:

- Primär (Telemetrie Daten, Statistik Daten, direkter Kontakt)
- Sekundär (Etwas das über eine primär Ressource geschrieben ist, Presse Artikel)
- Tertiär (Nur eine Sammlung von Daten)

Der Weg vom Inside zur Konsequenz:

- 1 Stück Information ist als "Inside" zu betrachten
- 2-3 "Insides" mit verschiedenen Fokus Punkten oder Stati sind eine Information
- Aus mehreren Information zusammen kann ich eine Konsequenz zum Handeln ableiten (Raum, Kraft, Zeit - Prinzip)

Schlussfolgerung zu AI Lösungen

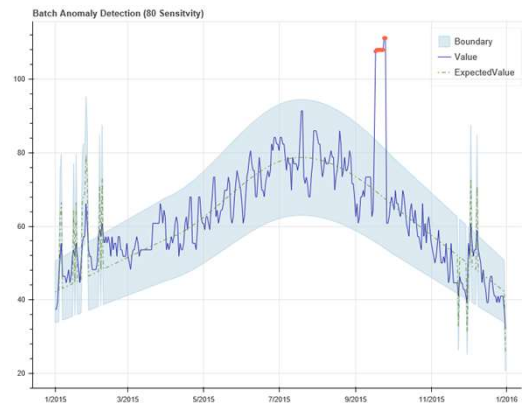
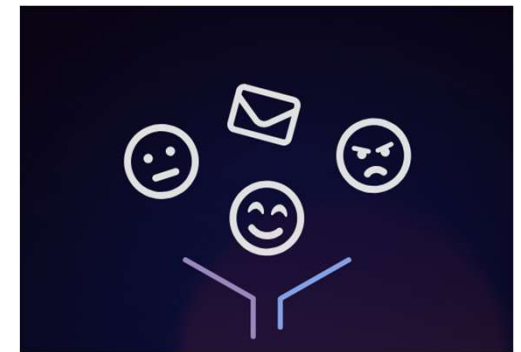
Kann die GenAI Fake News vermeiden?



Kann Gen AI genutzt werde um Fake News entgegenzuwirken?



Kann maschinelles Lernen dabei helfen, Fake News zu erkennen?



Was kommt als nächstes?

Mehr zweck gefundene vortrainierte Modelle

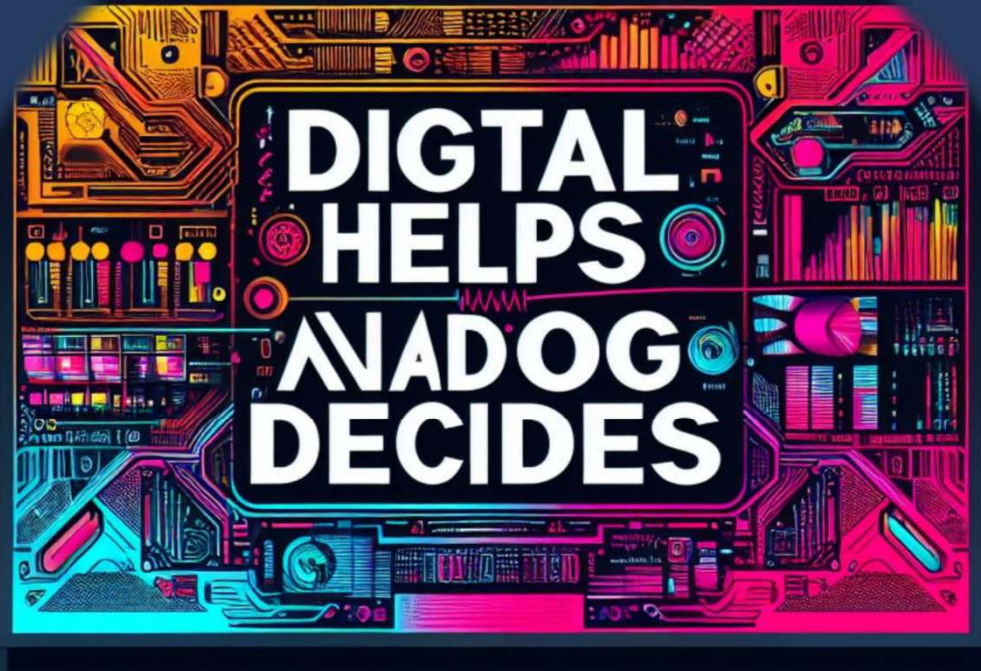
- GPT3.5 turbo-instruct
- Wisper

AI Lösungen werden als Basic Technology wahrgenommen

- Grundausbildung über Daten und das Verhalten damit
- Die "Principles of responsible AI" müssen in Gesetzen und Richtlinien verankert werden
- Schaffung internationaler Rahmenbedingungen für die Entwicklung, Umsetzung und den Betrieb von verantwortungsvoller KI

Die Entwickler Community wird?

- Sich wild auf dieses Thema stürzen und versuchen "Cracy Shit" zu erstellen
- Integrierte Lösungen nach der Q&A Phase versuchen zu erstellen um Leben zu verbessern



Thanks
You!