OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG
FAKULTÄT FÜR INFORMATIK

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

KI-Sicherheit im Diskurs domänenspezifischer Anwendungsfelder

Sandro Hartenstein
sandro.hartenstein@hwr-berlin.de

Herausforderungen Low-Code orientierter KI-Ansätze
Fraunhofer IESE Kaiserlautern
12.11.2024

Agenda

- Motivation
- Umfrage
- Diskussionsrunden – KI-Sicherheit (je ca 10 Minuten)
  - Runde 1: Sprachverarbeitung mittels LLMs in der Mediationsanalyse
  - Runde 2: Totholzerkennung via Drohnen
  - Runde 3: Vandalismus-Erkennung und visuelle Echtzeitklassifikationsalgorithmen
- Zusammenfassung und Abschluss

Motivation

Quelle: https://chatgpt.com/share/672dd785-d4bc-800d-a98f-eb983341071f

Motivation

Der zunehmende Einsatz von KI-Systemen in sicherheitskritischen und sensiblen Anwendungsbereichen stellt uns vor neue Herausforderungen.

Wir betrachten drei konkrete Anwendungsfälle:

- Die Analyse vertraulicher Mediationsgespräche mittels Large Language Models
- Die automatisierte Erkennung von Totholz durch Drohnensysteme im Forst
- Die KI-gestützte Videoanalyse zur Erkennung von Vandalismus

Motivation

Aktuelle Relevanz

Diese Anwendungsfälle repräsentieren unterschiedliche Facetten moderner KI-Systeme:

- Verarbeitung sensibler personenbezogener Daten
- Autonome Entscheidungsfindung in der realen Umgebung
- Automatisierte Überwachung und Analyse im öffentlichen Raum

Forschungsbedarf

Die Entwicklung geeigneter Sicherheitsstandards für diese Systeme erfordert

- Technische Robustheit und Zuverlässigkeit
- Rechtliche und datenschutzrelevante Konformität
- Ethische Vertretbarkeit und gesellschaftliche Akzeptanz

Motivation

Bedeutung des interdisziplinären Austauschs

Ihre unterschiedlichen Expertisen sind entscheidend für:

- Identifikation kritischer Sicherheitsanforderungen
- Entwicklung praxistauglicher Lösungsansätze
- Bewertung möglicher Zielkonflikte

Ziele der heutigen Diskussionsrunde sind

- Verknüpfung verschiedener Fachperspektiven
- Aufzeigen von Ansätzen für sichere KI-Systeme

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG
FAKULTÄT FÜR INFORMATIK

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Umfrage 5min

https://www.menti.com/al77parxiw1v

Diskussionsrunden

Der Teamsstream soll aufgezeichnet, später transkribiert und für Forschungszwecke ausgewertet werden.

Ziel der Analyse ist es, die Differenzen der anwendungsbezogenen Anforderungen an KI-Sicherheit aufzuzeigen.

Runde 1: Mediationsanalyse und LLMs

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Einordnung:

- KI-gestützte Analyse von Mediationsgesprächen

Technischer Kontext:

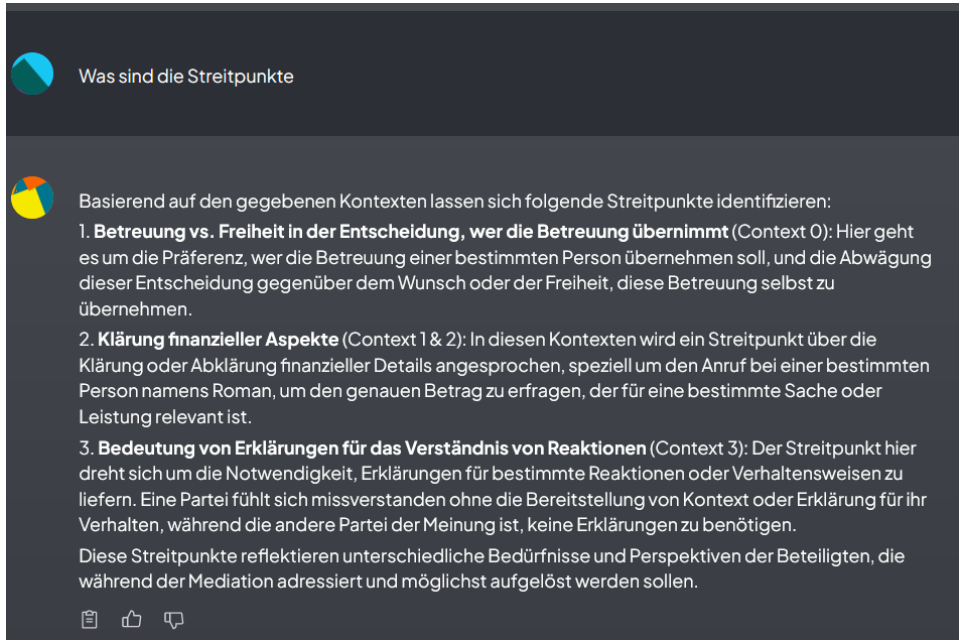
- Large Language Models analysieren Gesprächsprotokolle

Aktuelle Herausforderungen:

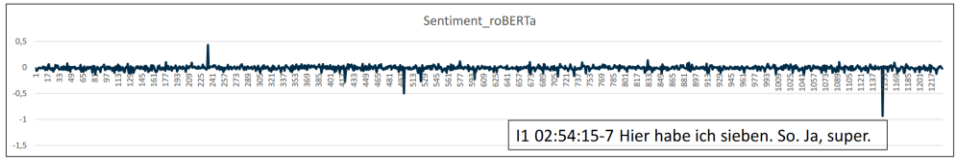
- Vertraulichkeit sensibler Gesprächsinhalte
- Zuverlässigkeit der KI-Analyse
- Balance zwischen Analysetiefe und Datenschutz

Diskussionsziel:

- Ansätze für vertrauenswürdige Analysen



Retrieval Augmented Generation Fall5



Sentiments im Zeitverlauf Fall5 S3

Runde 2: Totholzerkennung via Drohnen

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Einordnung:

- Automatisierte Waldinspektion mittels Drohnen

Technischer Ansatz:

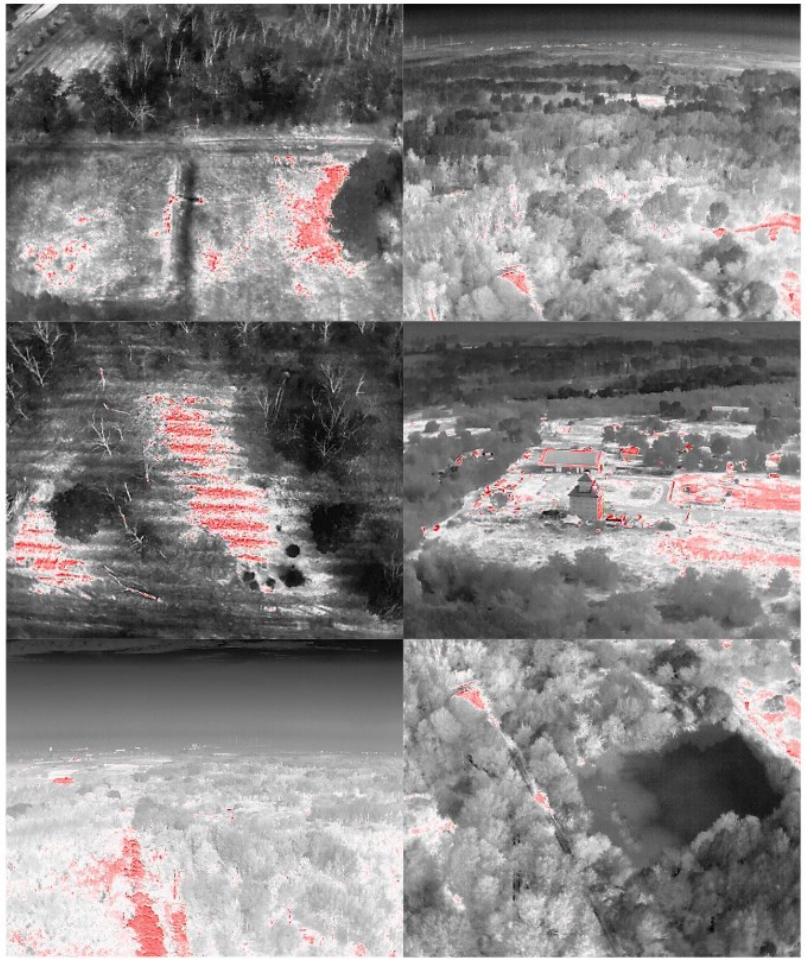
- Visuelle KI-Klassifikation von Luftbildern (Wärme- und RGB Bilder)

Zentrale Aspekte:

- Präzision der Erkennung
- Sicherheit im Flugbetrieb
- Umgang mit Umwelteinflüssen

Diskussionsziel:

- Ansätze für robuste visuelle KI-Systeme



Quelle:https://blog.hwr-berlin.de/schmietendorf/wp-content/uploads/2024/07/rodner-tahai-jul2024.pdf

Runde 3: Vandalismus-Erkennung und visuelle Klassifikationsalgorithmen

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Einordnung:

- KI-gestützte Videoüberwachung öffentlicher Räume

Technische Ansatz:

- Visuelle KI-Algorithmen zur Echtzeiterkennung von Objekten und Veränderungen

Zentrale Aspekte :

- Echtzeitanalyse von Videodaten
- Balance zwischen Sicherheit und Privatsphäre
- Vermeidung von Fehlalarmen

Diskussionsziel:

- Ansätze für sichere Überwachung mit KI

Quelle: https://universe.roboflow.com/nimra-sardar-n5hc6/vandalism-detection-2/images/qzbGuR9MO6RIV85OHovj

Zusammenfassung und Abschluss

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG
FAKULTÄT FÜR INFORMATIK

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Vielen Dank!

Sandro Hartenstein
sandro.hartenstein@hwr-berlin.de

Herausforderungen Low-Code orientierter KI-Ansätze
Fraunhofer IESE Kaiserlautern
21.11.2023