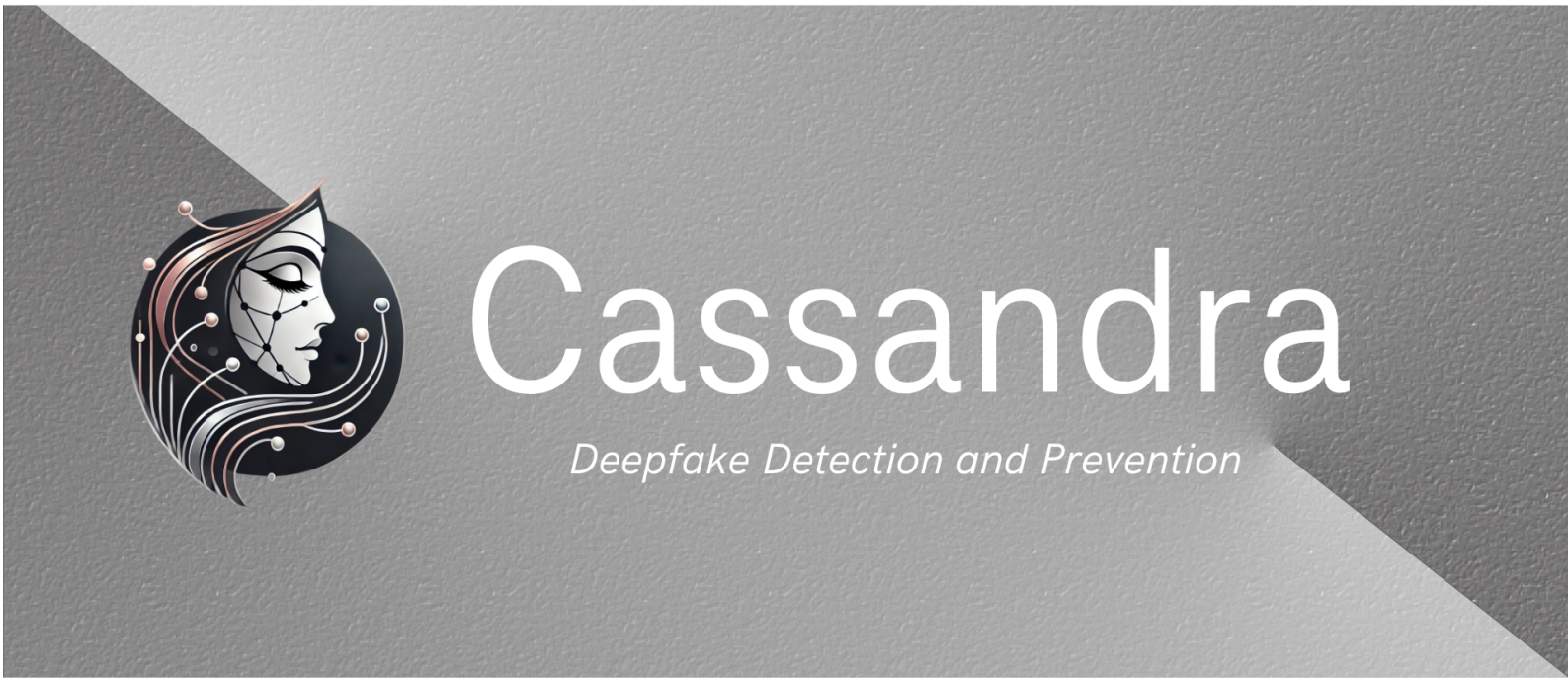

Einreichungskonzept:

Cassandra for Deepfake Detection and Prevention

1. Projektidee

Unser Projektvorschlag basiert auf einem innovativen Ansatz zur **Früherkennun**g von Deepfake-Methoden. Im Gegensatz zu bestehenden reaktiven Lösungen, die darauf abzielen, Deepfakes zu identifizieren, nachdem sie erstellt oder verbreitet wurden, setzen wir auf die **Prävention**. Mithilfe eines KI-gestützten Systems, das kontinuierlich Kontextinformationen, Trends und Narrative in digitalen Inhalten analysiert, möchten wir Deepfake-Techniken vorab erkennen und mögliche Manipulationen verhindern.

Die Herausforderung, die wir annehmen, besteht darin, Deepfake-Angriffe bereits in der Planungs- oder Entstehungsphase zu antizipieren.
Unser System wird nicht nur auf technische Merkmale von Medieninhalten achten, sondern auch soziokulturelle Trends, neue Technologien und Kommunikationsmuster in den digitalen Medien verfolgen, um aufkommende Bedrohungen frühzeitig zu identifizieren.

2. Lösungsansatz

Unsere Lösung orientiert sich am **Cassandra-Ansatz** von Professor **Jürgen Wertheimer** (*literarische Texte als Prognoseinstrumente für die Gewalt-Prävention*).

Diese Methode nutzt qualitative Daten – im Fall von Deepfakes wären das Texte, Fotos Videos und andere Medienformate –, um **Muster zu erkennen**, die auf manipulative Absichten hinweisen.

Ein KI-gestütztes System analysiert fortlaufend digitale Inhalte (*z. B. in sozialen Netzwerken, Nachrichtenportalen oder Foren*) und sucht nach:

- aufkommenden technischen Trends, die zur Erzeugung von Deepfakes genutzt werden könnten.

- Veränderungen in der visuellen und sprachlichen Darstellung von Informationen.

- Mobilisierungsnarrativen, Desinformationskampagnen oder manipulativen Erzählmustern.

- Diskussionen und technologischen Innovationen in Foren oder wissenschaftlichen Publikationen, die neue Deepfake-Tools oder -Techniken behandeln.

Das System erstellt eine umfassende Datenbank mit diesen Trendanalysen, die es ermöglicht, **Bedrohungen zu identifizieren, bevor sie sich voll entfalten**.

3. Technologie und Skalierbarkeit

Das Herzstück unseres Ansatzes ist eine **adaptive KI-Plattform**, die maschinelles Lernen und fortschrittliche Datenanalyse kombiniert.

Diese Plattform ist in der Lage, große Datenmengen in Echtzeit zu verarbeiten und Muster zu identifizieren, **die auf die Entwicklung und Nutzung neuer Deepfake-Techniken hinweisen**. Durch den Einsatz von Natural Language Processing (*NLP*) und Computer Vision wird die Plattform Trends in Text-, Audio- und Videoinhalten erkennen.

Die Lösung wird in mehreren Phasen entwickelt und getestet, wobei der Fokus auf der Integration in bestehende digitale Infrastrukturen liegt.

Ziel ist es, die Plattform so zu gestalten, dass sie sowohl von **Einzelpersonen** als auch von **großen Organisationen** genutzt werden kann.

Die Skalierbarkeit der Lösung wird es ermöglichen, das System flexibel an verschiedene Nutzeranforderungen anzupassen, sei es zur Analyse von Social-Media-Plattformen, Nachrichtenportalen oder Unternehmensinfrastrukturen.

4. Vorteile und Innovation

Der **präventive Charakter** unserer Lösung unterscheidet sich maßgeblich von den derzeitigen Deepfake-Detection-Ansätzen, die in der Regel reaktiv arbeiten.

Wir bieten:
- Früherkennung: Durch die Analyse von Trends können Bedrohungen identifiziert werden, bevor sie auftreten.
- Prävention: Der Schwerpunkt liegt auf der Entwicklung von Gegenmaßnahmen, bevor Deepfake-Angriffe ausgeführt werden.
- Kontinuierliche Anpassung: Unsere Plattform passt sich automatisch an neue Manipulationsmethoden an, da sie kontinuierlich lernt und ihre Analysefähigkeiten verbessert.
- Nahtlose Integration: Die Lösung kann in bestehende Systeme integriert werden, um eine umfassende Überwachung der gesamten digitalen Wertschöpfungskette zu gewährleisten.

5. Projektdurchführung *(TRL 3 bis TRL 7)*

In der ersten Phase *(TRL 3)* entwickeln wir einen Proof-of-Concept, der zeigt, wie unser System erste Anzeichen von Deepfake-Bedrohungen erkennt. In Phase Zwei *(TRL 6-7)* werden wir diese Lösung in realen Umgebungen testen und optimieren.

Dabei werden wir mindestens drei verschiedene Anwendungsfälle adressieren, z. B. Social Media, Nachrichtenportale und Videokonferenzsysteme.

6. Team und Expertise

Unser Team setzt sich aus Fachleuten der Bereiche KI, maschinelles Lernen und digitale Medienanalyse zusammen. **Constantin Sohn** wird die technische Leitung des Projekts übernehmen und sich um die Systementwicklung und -integration kümmern, während **Gunnar Sohn** die strategische Ausrichtung und die Marktintegration verantwortet. Unsere Expertise in digitalen Medien und KI-gestützter Analyse ermöglicht es uns, innovative Lösungen zu entwickeln, die sowohl technologisch als auch gesellschaftlich relevant sind.

7. Finanzierung und Zeitplan

Stufe 1 Proof-of-Concept und Initiale Tests *(TRL 3 – TRL 5)*
Finanzierung: Bis zu 350.000 €

KI-Spezialisten (*2 Personen, 7 Monate*) Entwicklung der Modelle für die Früherkennung in Bild, Text und Video, Testdatensätze erstellen und trainieren, kontinuierliche Modellverbesserung. *Kosten: 150.000 €*	**Datenwissenschaftler** *(1 Person, 7 Monate)* Datenverarbeitung, Feature-Engineering, Analyse von Trends und Anomalien in den Daten. *Kosten: 60.000 €*	**Entwickler für Backend/ Frontend** *(2 Personen, 7 Monate)* Aufbau der API-Schnittstellen, Integration der Modelle in eine skalierbare Infrastruktur, Implementierung von Testumgebungen. *Kosten: 90.000 €*

Ziel dieser Phase ist die Entwicklung eines **Proof-of-Concept** (PoC), der zeigt, wie Deepfake-

Bedrohungen frühzeitig erkannt werden können, bevor sie voll ausgereift sind.

1. Team und Personalkosten

Cloud-Computing-Ressourcen Für die Verarbeitung von großen Datenmengen, Modelltraining und Tests in Echtzeit (AWS, Google Cloud). *Kosten: 20.000 €*	**Software-Lizenzen** Tools für Datenanalyse, Bild- und Videoverarbeitung, KI-Modellierung (z.B. TensorFlow, OpenCV, PyTorch). *Kosten: 10.000 €*	**Projektmanagement und Administration** Kosten für administrative Aufgaben, Koordination der Teams, Kommunikation mit externen Partnern und Experten. *Kosten: 20.000 €*

2. Infrastruktur und Tools

Begründung für den Aufwand in Stufe 1

- **Entwicklung des Proof-of-Concepts**
  Diese Phase erfordert intensives Modelltraining und Testen, um ein funktionsfähiges System zu entwickeln, das in der Lage ist, erste Anzeichen von Deepfakes zu erkennen. Die größte Herausforderung besteht darin, effiziente und skalierbare Algorithmen zu erstellen.
- **Integration in bestehende Systeme**
  Das Team benötigt ausreichend Ressourcen, um eine erste Version des Prototyps zu entwickeln, der in bestehende digitale Infrastrukturen integriert werden kann.

Stufe 2 Erweiterung und Realtests *(TRL 6 – TRL 7)*
Finanzierung: Bis zu 425.000 €

Ziel dieser Phase ist es, den PoC weiterzuentwickeln und das System in realen Umgebungen zu testen und zu optimieren.
Diese Phase umfasst die Verifizierung in mindestens drei verschiedenen Anwendungsfällen (Social Media, Nachrichtenportale, Videokonferenzsysteme).

KI-Spezialisten *(3 Personen, 6 Monate)* Weiterentwicklung und Optimierung der Modelle, Integration neuer Erkenntnisse aus den Tests in realen Umgebungen. *Kosten: 180.000 €*	**Datenwissenschaftler** *(1 Person, 6 Monate)* Erweiterte Datenanalysen und Feature-Engineering basierend auf realen Anwendungsszenarien. *Kosten: 70.000 €*	**Entwickler für Backend/ Frontend** *(2 Personen, 6 Monate)* Anpassung des Prototyps an spezifische Plattformen (Social Media, Nachrichtenportale, etc.) und Optimierung der API. *Kosten: 110.000 €*

1. Team und Personalkosten

Cloud-Computing und Serverkosten Erweiterte Serverkapazitäten für Echtzeitverarbeitung, Tests in mehreren Szenarien. *Kosten: 25.000 €*	**Erweiterte Software-Lizenzen und Tools** Für spezifische Plattformintegrationen und zusätzliche Features. *Kosten: 10.000 €*	**Reisekosten und Workshops** Durchführung von Workshops mit Plattformpartnern, Realtests vor Ort und in Kooperationsumgebungen. *Kosten: 20.000 €*

2. Infrastruktur und Tools

Begründung für den Aufwand in Stufe 2

- **Erweiterung und Realtests**
  In dieser Phase ist es entscheidend, dass das System in realen Szenarien getestet wird, um seine Wirksamkeit unter verschiedenen Bedingungen zu beweisen.
- **Spezifische Anpassungen und Optimierungen**
  Die Integration in verschiedene Plattformen (Social Media, Nachrichtenportale, Videokonferenzsysteme) erfordert spezifische Anpassungen des Systems und umfangreiche Tests zur Sicherstellung der Interoperabilität.

Gesamtfinanzierung und Fazit

Die beiden Stufen erfordern eine klare Aufteilung der Ressourcen zwischen der initialen Entwicklung und den Realtests.

Jede Phase ist wichtig, um sicherzustellen, dass das System präventiv auf Deepfake-Bedrohungen reagieren kann und gleichzeitig flexibel und skalierbar bleibt.

8. Zusammenfassung

Unser Projekt zielt darauf ab, eine bahnbrechende Lösung zur Früherkennung und Prävention von Deepfakes zu entwickeln.

Durch den **präventiven Ansatz**, der auf die **Analyse von Trends und Kontextinformationen** setzt, können wir eine skalierbare und innovative Lösung schaffen, die in bestehenden digitalen Infrastrukturen integriert werden kann.

Wir sind überzeugt, dass unsere Lösung das Potenzial hat, einen bedeutenden Beitrag zum Schutz der digitalen Integrität im Zeitalter der Künstlichen Intelligenz zu leisten.

Einreichungsvoraussetzungen

- Der Cassandra-Ansatz zur Früherkennung basiert auf der kontinuierlichen Analyse von Kontexttrends und technologischem Fortschritt.

- Skalierbarkeit und Integration sind zentrale Elemente unserer Lösung, die sowohl für Einzelpersonen als auch für große Plattformen einsetzbar ist.

- Unsere Lösung wird von TRL 3 (Laborstadium) zu TRL 7 (Prototyp) weiterentwickelt und umfasst mindestens drei Anwendungsfälle: Social Media, Nachrichtenportale und Videokonferenzsysteme.