

# A Title

a subtitle

Adrian Joost<sup>1,\*</sup>

<sup>1</sup>*Fachhochschule Graubünden*

*\*E-Mail Adressen: adrian.joost@stud.fhgr.ch*

8. Januar 2026

## Zusammenfassung

Abstract

## 1 Introduction

Die Erfolge in der Suizidprävention sind seit Jahren stagnierend (404 error). Ein Suizid ist nicht nur ein Verlust einer einzelnen Person, sondern hat auch weitreichende schwere Einflüsse auf die Hinterbliebenen. Daher ist es wichtig, die Forschung in diesem Bereich voran zu treiben, um Erkenntnisse zu gewinnen, wie eine geplante Selbsttötung erkannt und verhindert werden kann.

Es gibt einige Versuche, mit verschiedenen Lösungsstrategien einen Algorithmus zu entwickeln, der vorhersagen kann, welche Menschen Suizidgefährdet sind (Words of Suicide, Predicting the Risk of Suicide). Allerdings zeigen diese Studien oft, dass die Genauigkeit eher schwach sind. Hinzu kommt, dass Suizid ein selten auftretendes Verhalten ist (404 error). Das bedeutet, dass wenn 1 aus 100 Personen suizidgefährdet ist, und ein Modell mit 99% Genauigkeit Suizide vorhersagen kann, so würde das Modell zwar diese eine aus 100 Personen erkennen, aber eben auch 9 weitere Personen, die nicht suizidgefährdet sind.

Da sich Suizid auch in der Sprache äussern kann (Words of Suicide), stellt sich die Frage, ob die neuen grossen Sprachmodelle (LLM) bei der Erkennung von Suizidgefährdeten hilfreich sein können. Anders als traditionelle Machinelearning Algorithmen wie Support Vector

Machines oder Lineare Regressionen sind LLMs auf die Verarbeitung und Interpretation von Sprache spezialisiert.

Da die LLMs in sehr vielen Bereichen Einzug erhalten haben, hat die Europäische Union den EU Artificial Intelligence Act verabschiedet. Artikel 1 dieses Acts soll den Einsatz von Artificial Intelligence (AI) Systemen regulieren mit dem Ziel, die Innovation durch AI zu fördern und gleichzeitig negativen Folgen von AI entgegenzuwirken (European Union, 2024). Den Einsatz eines AI Systems zur Erkennung von Suizidgefährdeten würde durch Annex III, Paragraph 5 a und d als ein <High Risk AI System> eingestuft. Das bedeutet, dass selbst wenn LLMs perfekt Suizidgefährdungen früherkennen können, dürfen sie nur eingesetzt werden, wenn sie den Auflagen entsprechen.

Zu diesen Auflagen gehört auch, dass das Modell interpretierbar sein muss. Das heisst, eine menschliche Person muss nachvollziehen können, warum das System bei einer Person anschlägt und bei einer anderen nicht. Die Erklärbarkeit von LLMs ist aber noch immer ein ungeklärtes Feld der Forschung (High Stakes, XAI). (ExplainShapley) hat ein System vorgeschlagen, dass mit Hilfe der Shapley Values zu erklären versucht, welche Wörter des Inputs zum Output beigetragen haben.

In dieser Arbeit wird

## **2 Research Question**

## **3 Methodology**

## **4 Results**

## **5 Diskussion**

## **Literatur**

European Union. (2024). *Regulation (eu) 2024/1689 of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act)*. Official Journal of the European Union. (OJ L, 2024)