

Vertrag zur Abschlussarbeit

Vertrag zur Bachelorarbeit von **Kristoffer Schaaf** bei **Prof. Dr. Stefan Sarstedt**

Vorläufiger Titel:

„Entwicklung einer Software zur Erkennung von Fake News auf Bild.de“

1. Problemstellung

Beschreibung und Einordnung des Problems

Fake News sind ein allgegenwärtiges Problem. Als politischer Schachzug bereits im Amerikanischen Bürgerkrieg 1779 genutzt, hat der Term auch in der US-Wahl 2016 wieder eine besondere Rolle gespielt ([Cite - A comprehensive Overview of fake news detection](#)).

Sie können die Form von Text, Fotos, Filmen oder Audio annehmen und sind dementsprechend auf jeder Plattform auffindbar, die die Verbreitung nicht unterbindet. Die momentan populärste Plattform zum Teilen der Fake News ist WhatsApp ([Cite - An Analysis and Identification of Fake News using Machine Learning Techniques](#)).

Fake News definieren sich wie folgt:

Fake News sind eine Form der Online-Desinformation.

Sie bestehen aus vollständig oder teilweise falschen Inhalten, die mit der bewussten Absicht erstellt werden, ein bestimmtes Publikum zu täuschen und/oder zu manipulieren.

Dabei ahmen sie gezielt das Format klassischer Nachrichten oder Reportagen nach, um so Glaubwürdigkeit zu erlangen.

Die enthaltenen Falschinformationen können – müssen aber nicht – in Zusammenhang mit realen Ereignissen stehen.

Charakteristisch ist eine opportunistische Struktur, die sich in auffälligen Titeln, emotionalen Bildern und einem strategisch gestalteten Inhalt zeigt.

*Ziel ist es, die Aufmerksamkeit der Leser*innen zu gewinnen und sie zur Glauben an die Falschinformation zu bewegen – um dadurch mehr Klicks und Teilungen zu erzeugen und letztlich finanzielle Einnahmen (z. B. durch Werbung) oder ideologische Vorteile zu erzielen. ([Cite - Understanding Fake News Consumption- A Review](#))*

Zur Identifizierung von Fake News gibt es fünf verschiedene Ansätze:

Kategorie	Kurzbeschreibung	Beispiele / Methoden
Language Approach	Analyse sprachlicher Merkmale (Wortwahl, Grammatik, Syntax).	<ul style="list-style-type: none">- Bag of Words (n-Gramme)- Semantische Analyse- Deep Syntax (PCFG, Parse Trees)
Topic-Agnostic	Bewertet formale Eigenschaften statt Inhalte.	<ul style="list-style-type: none">- Viele Werbeanzeigen- Lange, reißerische Headlines- Autorennamen vorhanden
Machine Learning	Einsatz trainierter Algorithmen auf Fake-News-Datensätzen.	<ul style="list-style-type: none">- Twitter Crawler- Rumor Identification Framework- Datensätze via Crowdsourcing
Knowledge-Based	Abgleich mit externen Faktenquellen zur Validierung.	<ul style="list-style-type: none">- Expert Fact Checking- ClaimBuster Tool- Crowdsourcing über Kiskkit

Hybrid Approach	Kombination mehrerer Ansätze (Text, Quelle, Reaktionen).	<ul style="list-style-type: none"> - CSI-Modell (Capture, Score, Integrate) - Kombi aus menschlicher & KI-Analyse
------------------------	--	---

(Cite - [Approaches to Identify Fake News: A Systematic Literature Review](#))

Im Folgenden wird ausschließlich der sogenannte „Machine Learning“-Ansatz behandelt.

Relevante Literatur zu diesem Bereich ist unter anderem hier zu finden:

- (Cite - [Fake News technisch begegnen – Detektions- und Behandlungsansätze zur Unterstützung von NutzerInnen](#))
- (Cite - [A comprehensive overview of fake news detection on social networks](#))
- (Cite - [An Analysis and Identification of Fake News using Machine Learning Techniques](#))
- (Cite - [Fake News Detection using Machine Learning](#))

Verschiedene Arbeiten zu diesem Thema im deutschsprachigen Raum sind:

- TH Ingolstadt - (Cite - [Comparison of deep learning and word embedding approaches for German fake news detection](#))
- University of Applied Sciences Upper Austria - (Cite - [Detecting Fake News and Performing Quality Ranking of German Newspapers Using Machine Learning](#))
- RWTH Aachen - (Cite - [FANG COVID](#))
- TU Darmstadt - (Cite - [2021_HartwigReuter_FakeNewstechnischbegegnen_WahrheitundFake](#))

Aus Betrachtung dieser geht hervor, dass die allgemeine Aufarbeitung dieses Themas bereits umfassend bearbeitet wurde. Es existieren verschiedenste „Machine Learning“-Modelle zum Erkennen von Fake News mit teils sehr guten Metriken. Auch stehen verschiedene Datensätze mit deutschen Fake und nicht Fake Artikeln zur Verfügung.

Im Paper der University of Applied Sciences Upper Austria wird die Qualität verschiedener deutscher Nachrichtenmedien mit „Machine Learning“-Modellen getestet. Das Ergebnis zeigt, dass Spiegel, Die Zeit und Süddeutsche die besten Medien sind und Express, BZ-Berlin und Bild die schlechtesten - also auch am meisten Fake News verbreiten.

Das Paper der TU Darmstadt stellt das White-Box Browser Plugin TrustyTweet vor. Dieses Plugin unterstützt Twitter-Nutzer*innen beim Erkennen von Falschinformationen. Es gibt politisch neutrale, transparente und verständliche Hinweise, ohne das eigene Urteilsvermögen einzuschränken. Ziel ist es, die Medienkompetenz der Nutzer*innen langfristig zu stärken.

Hauptziel der Arbeit

Motiviert durch die Arbeiten der University of Applied Sciences Upper Austria und der TU Darmstadt würde ich in dieser Arbeit gerne die Entwicklung eines weiteren Tools dokumentieren. Dieses Tool soll wie auch das Browser Plugin TrustyTweet eine Unterstützung zum Erkennen von Fake News anbieten.

Ob dieses Tool als Browser Plugin oder als eine andere Form der Software implementiert wird, steht zum jetzigen Zeitpunkt noch nicht fest.

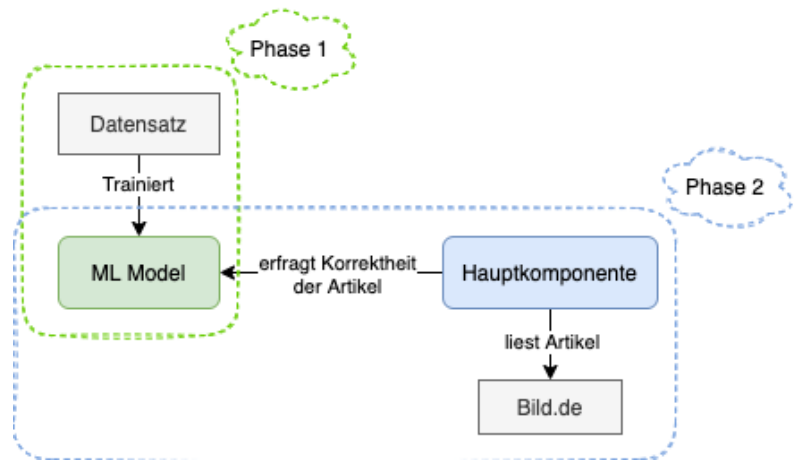
Auch ob eine Black Box oder White Box - das heißt, kann der User zum Beispiel sehen, warum der Artikel als Fake News deklariert wird - Architektur genutzt wird, wird im Laufe der Arbeit entschieden.

Ziel ist es, dass das Tool nicht wie TrustyTweet auf Twitter eingesetzt wird, sondern auf Bild.de. Bild.de ist ein deutschsprachiges Nachrichtenportal des Axel-Springer-Verlags. Gemäß Zahlen der IVW ist Bild.de das meistbenutzte Nachrichtenportal in Deutschland, vor n-tv.de und Der Spiegel (online) (Cite - [Wikipedia](#)).

Es ergeben sich zwei wichtige Softwarekomponenten:

(1) Ein „Machine Learning“-Model, welches durch in „Fake“ und „nicht Fake“ klassifizierte, deutsche Artikel trainiert wird.

(2) Eine Hauptkomponente, welche die Schnittstelle zum Endnutzer in Form einer UI anbietet und außerdem mit dem „Machine Learning“-Model kommuniziert um die zu analysierenden Bild.de Artikel auszuwerten.



Datensatzwahl

Folgende Datensätze enthalten deutsche Fake News Artikel:

Datensatz	Quelle	Anzahl Zeilen	Anteil Fake (%)	Besonderheiten / Einschränkungen	Eignung
Fake News Dataset German	University of Applied Sciences Upper Austria	63.868	7,24 %	Geringer Anteil an Fake-News → unausgewogene Klassenverteilung	Weniger geeignet
GermanFake NC	Fraunhofer-Institut für Sichere Informationstechnologie SIT	489	100 %	Enthält nur Referenzen, keine Texte → keine direkte Textauswertung möglich	Nicht geeignet
FANG-COVID	Association for Computational Linguistics	41.242	31,97 %	Ausgewogene Klassen, vollständige Texte, viele Metadaten (Titel, Header, Label)	Sehr geeignet
DeFaktS	FZI Forschungszentrum Informatik	–	–	Kein Zugang → Nutzung ausgeschlossen	Nicht verfügbar

Für das Trainieren meines Models kommt am ehesten der FANG-COVID-Datensatz in Frage. Dieser bietet mit über 41.000 Einträgen eine gute Datengrundlage. Außerdem hat dieser mit einem Fake-News-Anteil von ca. 32 % eine im Vergleich zu anderen verfügbaren Datensätzen deutlich ausgewogenere Klassenverteilung. Ob diese Klassenverteilung ausreicht, kann zum jetzigen Zeitpunkt noch nicht bestätigt werden. **Deshalb ist diese Festlegung auf diesen Datensatz erst einmal temporär.**

Darüber hinaus enthält der Datensatz neben der Label-Kennzeichnung („fake“ oder „real“) auch Titel, Header und den gesamten Beitrag der Artikel, was eine tiefergehende textanalytische Auswertung ermöglicht.

Im „German News Dataset“ (<https://www.kaggle.com/datasets/pqbsbk/german-news-dataset>). Sind 19862 Artikel der Bild.de enthalten. Diese erleichtern das Testen der Hauptkomponente während der Entwicklung.