



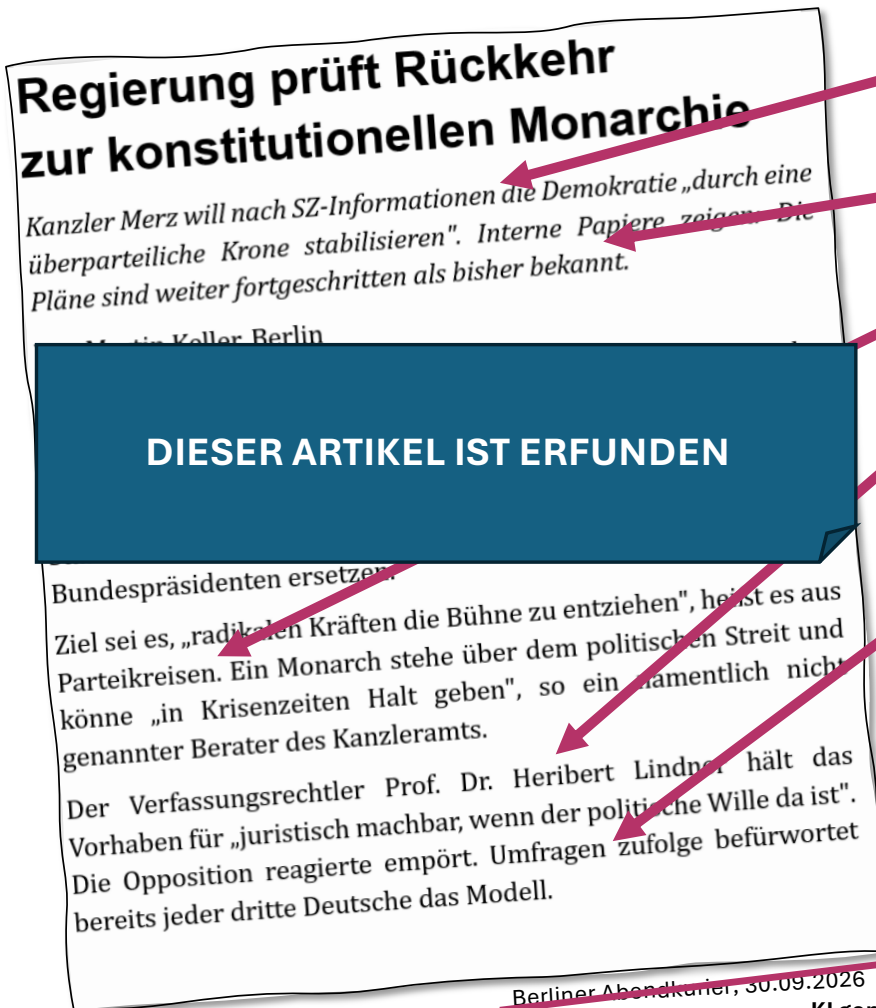
KI & Social Media

Gefahr für unsere Demokratie?

Thomas Schmidhuber
dreivirtuell® UG (haftungsbeschränkt)

Thomas.Schmidhuber@email.de

Neulich in den Medien



➤ Geliehene Glaubwürdigkeit

Verweis auf ein bekanntes und zuverlässiges Medium.

➤ Unprüfbare Quelle

Wer kann schon ein „internes Papier“ bestätigen oder widerlegen?

➤ Anonyme Stimmen

„Parteikreise“ und „ein namentlich nicht genannter Berater“.

➤ Erfundener Experte

„Prof. Dr.“ wirkt seriös. Eine kurze Recherche zeigt, dass es diese Person nicht gibt.

➤ Zahlen ohne Beleg

„Umfragen zufolge“ nennt weder Institution noch Datum. Die Zahl soll Normalität suggerieren.

➤ Paradoxe Logik in seriösem Ton

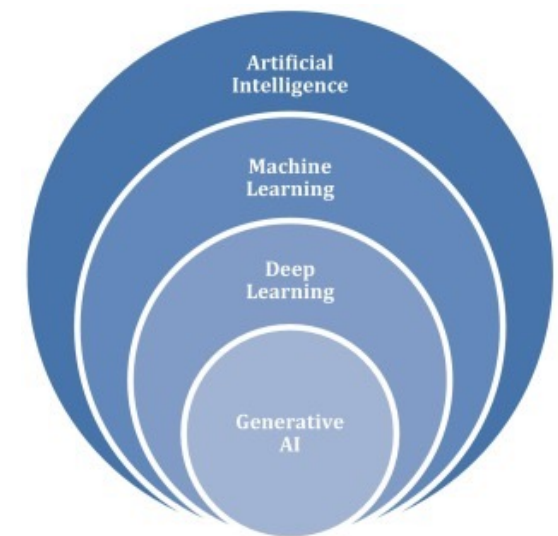
Die Idee, Demokratie einschränken, um Sie zu schützen wirkt im Nachrichtenton weniger absurd, als sie ist.

Zugegeben: Ein sehr konstruiertes Beispiel

- Solche Falschmeldungen in Printmedien, TV oder Radio zu verbreiten kostet enorme Mengen Geld und Zeit
- Inhalte werden oft umfangreich redaktionell aufbereitet und lektoriert
- Unsere Printmedien (bzw. Webseiten) haben eine hohe Glaubwürdigkeit
- **Der Aufwand zur Verbreitung von Falschmeldungen ist enorm**

Was ist eigentlich Künstliche Intelligenz?

- Was heute als „KI“ (in Form von Agenten, Assistenten und Chatbots) bezeichnet wird, ist nur ein kleiner Teil eines riesigen Forschungsgebiets
- Generative Text-KI
Mathematische / statistische Methoden zur Erstellung von natürlich klingender Sprache
- Ziel der Forschung: möglichst schnell, möglichst plausibel klingende Inhalte zu erstellen



Die Illusion der Plausibilität

- Sprachmodelle werden anhand von Textdaten trainiert
- Das Internet ist voll von diesen Textdaten: Bücher, Blogs, Zeitungsartikel, öffentliche Archive, Gesetzestexte, Wikipedia, ...
- Ein aktuelles Sprachmodell ist trainiert auf ca. 50 Billionen Token Text
(zum Vergleich: Harry Potter umfasst ca. 1,5 Millionen Token; Herr der Ringe ca. 650.000 Token)
- Das Sprachmodell hat dabei erlernt, **wie Sprache funktioniert**

Die Illusion der Plausibilität (II)

➤ Was gehört zusammen? Was taucht häufig zusammen auf?

Himmel → blau

Kirsche → rot

Gras → grün

Haribo macht Kinder froh. Und Erwachsene ebenso.

Markdorf. Markant. Lebendig. Mit besten Aussichten.

Die Illusion der Plausibilität (III)

➤ Seit tausenden von Jahren galt eine einfache Regel:

Je flüssiger und müheloser sich ein Text lesen lässt, desto eher halten wir ihn für wahr. Falschnachrichten verraten sich oft durch holprige Sprache und Fehler.

➤ **KI-Systeme heben diese Faustregel aus den Angeln**

KI-Systeme sind keine Wahrheitsmaschinen

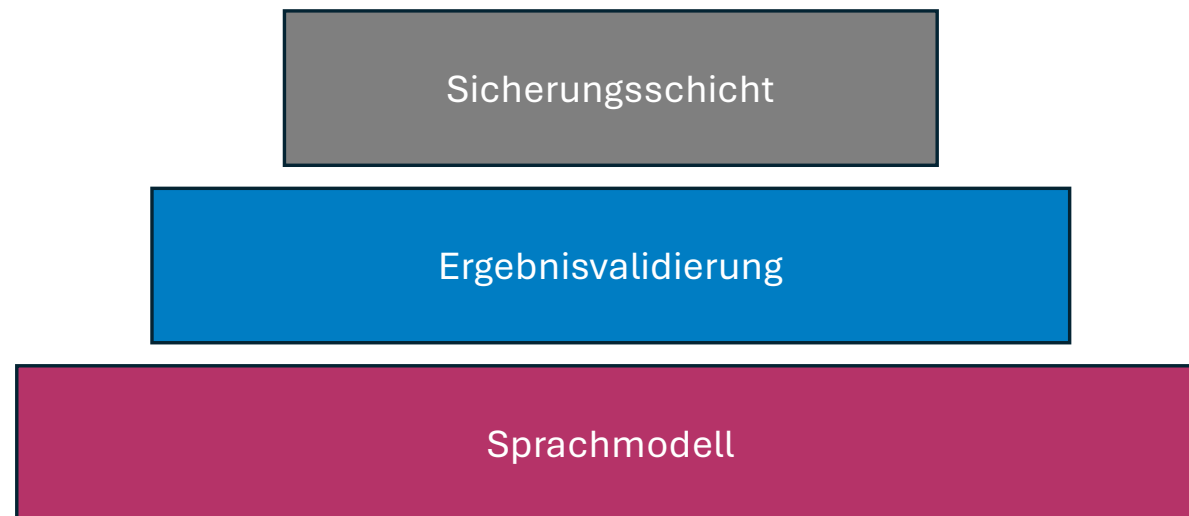
Wahrheit: Haribo macht Kinder froh. Weil es Unmengen an Zucker enthält.

Gelernt: Haribo macht Kinder froh. Und Erwachsene ebenso.

- Was bedeutet das für unsere Demokratie?
 - KI kann die Definition von Demokratie wiederholen und umformulieren
 - KI kann den Widerspruch zwischen Demokratie und Monarchie in Texten aufdecken
 - KI kann den Zusammenhang zwischen Texten zur Demokratie herstellen
- **Aber: KI versteht nicht, was Demokratie ist**

KI-Systeme sind keine Wahrheitsmaschinen (II)

- Durch das Training kennt KI das öffentlich zugängliche „Weltwissen“
- KI besitzt aber keine Moral-/Wertevorstellung hierzu
- „Moral“ und „Werte“ werden durch die Sicherungsmechanismen ergänzt



Was macht KI gefährlich?

- KI kann mit einfachen Anweisungen sehr schnell unzählige Texte erstellen
 - Gleicher Inhalt mit kleinen Variationen
 - Unterschiedlicher Sprachstil (per-Du, per-Sie, Jugendsprache, etc.)
 - Die Namen der Quellen und Autoren werden mit erfunden
- Bei lokal gehosteten Modellen können die Sicherungsschichten problemlos ausgeschalten/umgangen werden
- **Falschnachrichten können in bester Qualität schnell erstellt werden**

Was macht KI gefährlich? (II)

➤ Über Text-KI hinausgehend: Deepfakes

- Bilder, Videos und Tonaufnahmen von Politikern mit gezielten Falschaussagen.

➤ Mikrotargeting

- KI-gestützte Analysen von persönlichen Daten, um Inhalte auf individuelle Ängste und Vorlieben zuzuschneiden.

➤ Vertrauensverlust

- Wenn alles gefälscht sein könnte, wer glaubt dann noch den echten Beweisen?

➤ Algorithmische Diskriminierung

- KI-Systeme können Vorurteile aus Trainingsdaten übernehmen.

Was sind Soziale Medien?

- Soziale Medien wurden als „Demokratisierung der Medien“ bezeichnet
- Jeder kann seine eigenen Inhalte weltweit verbreiten

- Einige Beispiele

Facebook

Twitter/X

Discord

Instagram

TikTok

Twitch

WhatsApp

LinkedIn

YouTube

Telegram

Xing

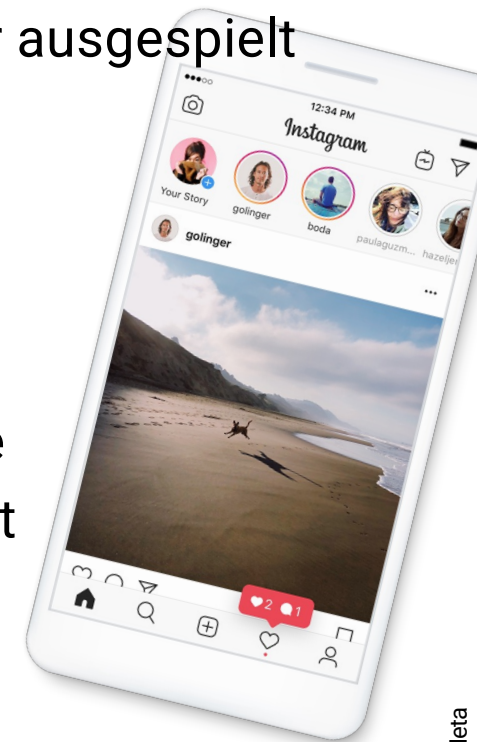
Pinterest

- **Die Inhaltsbeschränkungen erfüllen den gesetzlichen Mindeststandard**

(meist nicht deutschen oder europäischen)

Wie verbreiten sich Inhalte auf Social Media?

- Inhalte werden von einem Algorithmus bewertet
- Je besser die Bewertung, umso öfter wird der Inhalt an Nutzer ausgespielt
- Ziel: Verbreitung populärer Inhalte → Werbeeinnahmen
- Diese Algorithmen sind neben der Rezeptur von Coca-Cola die wohl am **besten gehüteten Geschäftsgeheimnisse** dieser Welt



Wie gelangen die Inhalte in die Netze?

- Benutzer erstellen die Inhalte am Handy oder am PC
- Fast alle Sozialen Medien besitzen Programmierschnittstellen, um eine große Menge von Inhalten schnell einzuspielen (z.B. Zeitungen)
- Diese Programmierschnittstellen sind oftmals kostenfrei nutzbar
- **Beispiel Instagram:** Kostenlos für jedermann: 200 Posts / Stunde
 Beahlt für „Agenturen“: 100.000 Posts / Stunde

Wenn der Algorithmus anspringt

- Alle Social Media Algorithmen funktionieren in einem Punkt identisch:
 - Kommen in kurzer Zeit eine große Menge inhaltlich ähnlicher Nachrichten, scheint das Thema wichtig zu sein!
 - Die Folge: Diese Inhalte werden überproportional oft ausgespielt.
- Was das für eine große Anzahl Falschnachrichten bedeutet, ist wohl klar...

Aber die Nutzungsbedingungen...

- Die Nutzungsbedingungen der meisten Plattformen verbieten die Flutung durch automatisierte Systeme (Bots)
- Ebenso ist die massenhafte Nutzung unter verschiedenen Identitäten auf den meisten Plattformen untersagt
- **Aber: wer sich um die Demokratie nicht schert, der schert sich sicher auch nicht um die Nutzungsbedingungen von Online-Diensten!**

Soziale Medien – Gefahren

➤ Echokammern

- Die Algorithmen zeigen Nutzern gezielt Inhalte, die deren Meinung bestätigen.

➤ Polarisierung

- Empörende und emotionale Inhalte erzeugen mehr Klicks → stärkere Verbreitung.

➤ Radikalisierung

- Aufgrund der Anonymität nutzen extremistische Gruppen Soziale Medien.

➤ Verdrängung

- Werbeeinnahmen wandern an Online-Netzwerke, dadurch sterben Lokalzeitungen und kritischer, qualitativer Journalismus.

Soziale Medien – Gefahren (II)

➤ Macht der Plattformbetreiber

- Die Algorithmen sind ein gut gehütetes Geschäftsgeheimnis; somit entscheiden die Plattformbetreiber, welche (politischen) Inhalte sichtbar werden.

➤ Vereinfachung komplexer Themen

- Teilweise bestehen Längenbeschränkungen für Inhalte, somit werden komplexe Themen sehr stark vereinfacht dargestellt.

➤ Datensammlung und Manipulation

- Die Plattformen sammeln große Mengen an persönlichen Daten, diese können gezielt zur Manipulation der Nutzer herangezogen werden.

Künstliche Intelligenz & Social Media

- Was lässt sich hier tun?
 - Warten / hoffen auf bessere Regulierung?
 - Bessere technische Inhaltskontrollen?
 - Verbot?

Was kann ich denn tun?

- Erst nachdenken, dann teilen – Social Media lebt von der Verbreitung
 - Wenn ich kommentiere „das ist so nicht richtig“, verbreite ich den Originalinhalt ebenfalls weiter
- Quellen prüfen
 - Gibt es überhaupt Quellen? Woher stammt die Information?
- Faktenchecks
 - CORRECTIV, ARD-Faktenfinder, dpa-Faktencheck, ...
- Deepfakes erkennen lernen
 - Manche Bilder enthalten offensichtliche Fehler, manche sind einfach zu perfekt

Was kann ich denn tun? (II)

- Qualitätsjournalismus unterstützen
 - Lokale Zeitungen sind normalerweise besser recherchiert als Social Media Posts
- Algorithmen bewusst steuern
 - „Nicht interessiert“ klicken oder den Account blockieren/melden
- Melden statt ignorieren
 - Der Digital Services Act verpflichtet die Plattformen, Meldungen zu bearbeiten
- Datensparsamkeit
 - Privatsphäre-Einstellungen prüfen und überlegen, welche Daten wo stehen

Vielen Dank für die Aufmerksamkeit!

Und wer beginnt mit der ersten Frage?