

Mehrwerte aus unstrukturierten Online-Daten mit dem Data Knowledge Hub .

Clara Ruthardt | Bertelsmann Stiftung
Data Week Leipzig | 11.06.2025



Elon Musk ✅ X
@elonmusk

Subscribe

...

And no one is even trying to assassinate
Biden/Kamala 😐



brihallofficial #ad Bringing in the end of
this year 💔 . Time to go even harder next
year with a little help from @specialk -
Who's ready to join me and the
#FuelYourResolution challenge?

Load more comments

vanillab24 Bri hive ❤️ ★★★★

autumnvera94 Bri Hive ❤️❤️

curlycaulin Bri Hive ❤️

zuludemipeach Literally my fave cereal

sinnabun I just found out, and I am
beyond happy for you!! Like girl!!!! 😊 For
all the hard work, dedication, and pure
diligence that you have put into
everything you do, you deserve every
ounce! Thank you and I'll never forget
how you showed me how to make lipstick



PNP

DogeDesign Why they wan 8:41PM · 9/15/24

Posts

Classic

Sutra discussion-III.31 kürma-nādyām sthairyam
Posted by u/yogibattle 11 days ago

Mat Comparison Megathread
Posted by u/kalanya ashtangi /FAQBot 23 hours ago

Oh so helpful..
Posted by u/the_irish_yogi 1 hour ago

YSK about this sweet deal: Get \$300 off Google Pixel 3. With select trade-in. Terms



INTERNATIONAL
Première attaque ukrainienne
aussi loin du front

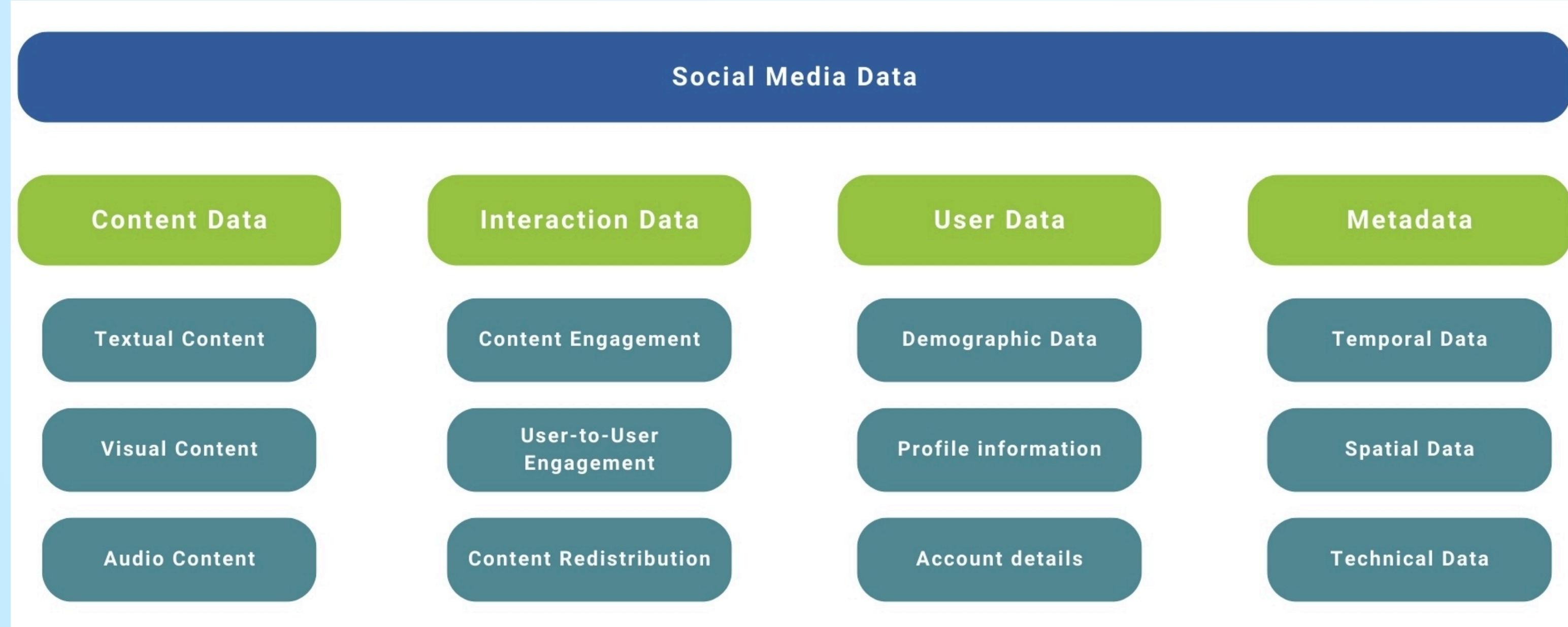


Follow @HopeForBoston

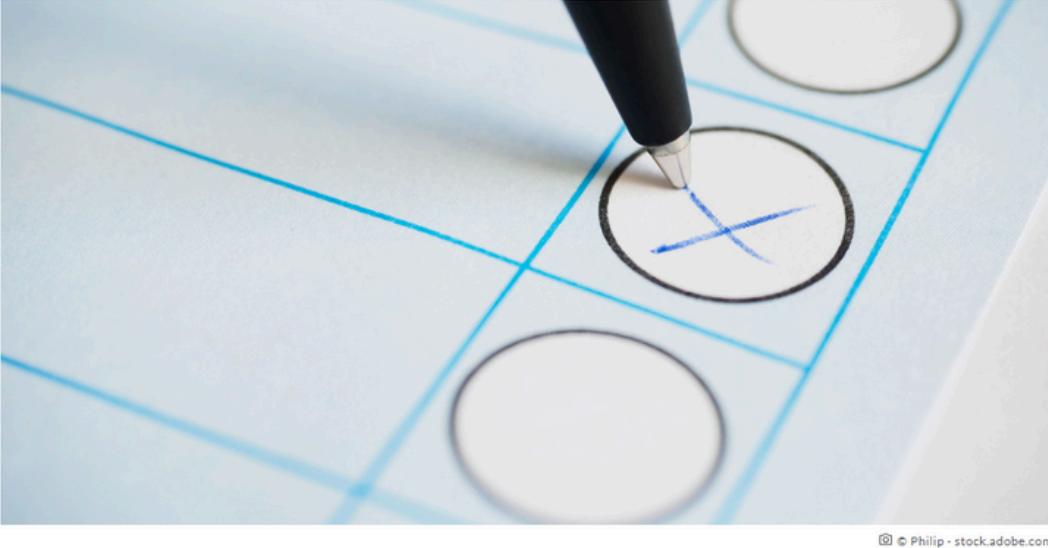
R.I.P. to the 8 year-old boy who died in Boston's explosions,
while running for the Sandy Hook kids. #prayforboston
<http://t.co/Xmv2E81Lsb>

April 16, 2013 12:18 am via web Reply Retweet Favorite

Zeit für etwas Struktur: Um welche **Datentypen** geht es?



Quelle: What types of data can be collected for research on social media?, CorrelAid e.V. | Data Knowledge Hub



16.12.2024

Superwahljahr 2024: Ein Superjahr für Desinformation?

2024 gab es mehr als 70 Wahlen weltweit. In fast allen kam es zu Versuchen, mit Desinformationen Einfluss auf den Wahlausgang zu nehmen. In mehr als der Hälfte der Fälle kamen KI-generierte Inhalte zum Einsatz. Was sagen diese Zahlen über den Zustand unserer Demokratien?

RUSSISCHE DESINFORMATION

Russische Einflussnahme läuft weiter – Bundesregierung bisher machtlos

Russland setzt seine Versuche fort, den Wahlkampf vor der Bundestagswahl zu beeinflussen. Jetzt nimmt eine von CORRECTIV kürzlich aufgedeckte Desinformationskampagne auch Friedrich Merz mit einem gefälschten Artikel ins Visier. Die Bundesregierung sieht dagegen kaum Handlungsmöglichkeiten.

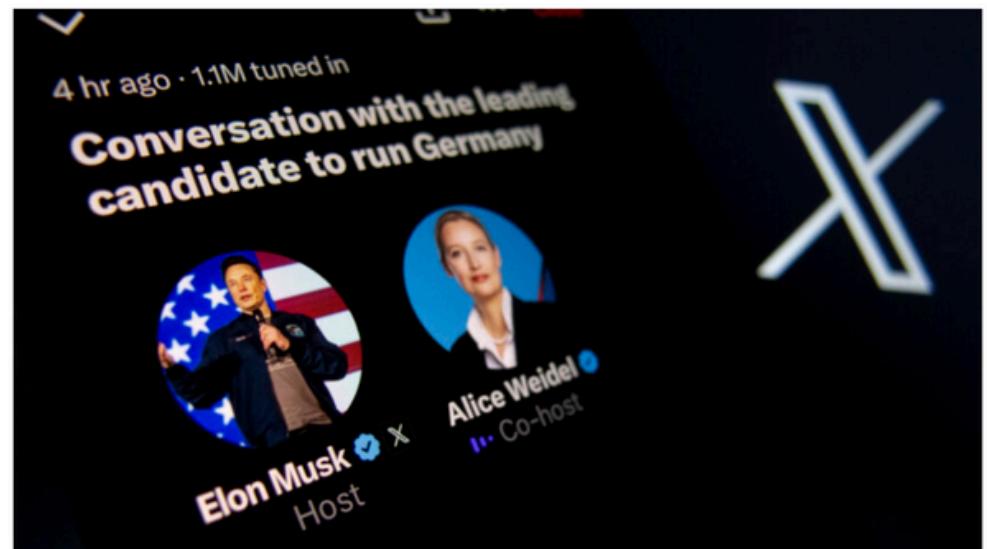
von Alexej Hock, Max Bernhard, Sarah Thust, Till Eckert

05. Februar 2025

Studien zu Algorithmen TikTok und X pushen rechte Parteien

Gleich zwei Untersuchungen zeigen, dass die rechtsradikale AfD von den Algorithmen bei TikTok und X überproportional profitiert. Zudem stellen andere Recherchen fest, dass die AfD auf TikTok von anonymen Accounts unterstützt wird, welche die Partei mit einfachen, oft KI-generierten Inhalten hochjubeln.

21.02.2025 um 10:15 Uhr - Markus Reuter - in Öffentlichkeit - eine Ergänzung

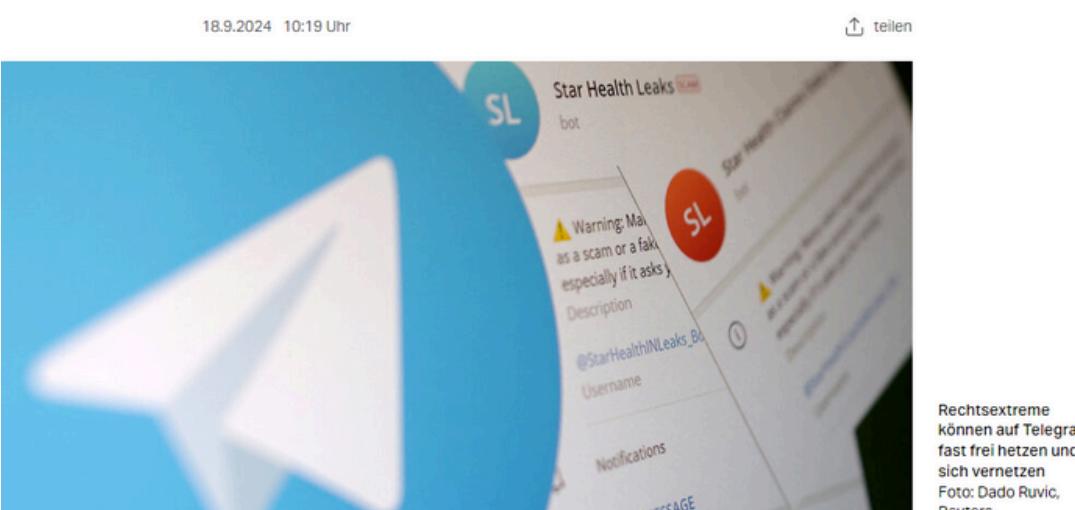


Elon Musks Haltung zugunsten rechter Parteien und Bewegungen ist bekannt.
– Alle Rechte vorbehalten IMAGO / ZUMA Press Wire

Zwei voneinander unabhängige Recherchen zeigen, dass die Plattformen X und TikTok mit ihren Algorithmen rechte Parteiinhalte überproportional pushen. So hat eine [Untersuchung von ZDF Frontal zusammen mit einem Forscherteam aus Dublin](#) ergeben, dass Elon Musks Plattform X sowohl die AfD wie auch das BSW überproportional oft im „Für Dich“-Feed anzeigt.

Neonazis im Internet Rechtsextremer Problemfall Telegram

Der Verfassungsschutz warnt in einer Analyse, wie vielfältig Rechtsextreme das Internet für sich nutzen. Vor allem Telegram bereitet ihnen Sorge.



Rechtsextreme können auf Telegram fast frei hetzen und sich vernetzen
Foto: Dado Ruvic, Reuters



Konfrontation mit EU

Stoppt Meta den Faktencheck auch in Europa?

von Anne Stadtfeld, Brüssel

08.01.2025 | 19:48

Der Meta-Konzern schafft unabhängige Faktenchecks in den USA ab. Experten befürchten, mehr Desinformation oder hasserfüllte Beiträge. Was bedeutet das für die EU?





Studien zu Algorithmen TikTok und X pushen rechte Parteien

Gleich zwei Untersuchungen zeigen, dass die rechtsradikale AfD von den Algorithmen bei TikTok und X überproportional profitiert. Zudem stellen andere Recherchen fest, dass die AfD auf TikTok von anonymen Accounts unterstützt wird, welche die Partei mit einfachen, oft KI-generierten Inhalten hochjubeln.

21.02.2025 um 10:15 Uhr - Markus Reuter - in Öffentlichkeit - eine Ergänzung



FIGURE C

Global risks ranked by severity over the short and long term

"Please estimate the likely impact (severity) of the following risks over a 2-year and 10-year period."

16.12.2024

Superwahljahr 2024: Ein Superjahr für Desinf

2024 gab es mehr als 70 Wahlen weltweit. In fast allen kam es zu Versuchen, mit Wahlausgang zu nehmen. In mehr als der Hälfte der Fälle kamen KI-generierte Zahlen über den Zustand unserer Demokratien?

RUSSISCHE DESINFORMATION

Russische Einf weiter – Bunde machtlos

Russland setzt seine Versuche, die Bundestagswahl zu beeinflussen, fort. Kürzlich aufgedeckte Desinformationen sind in einem gefälschten Artikel ins Verteilernetz gelangt und haben kaum Handlungsmöglichkeiten.

von Alexej Hock, Max Bernhard, Sarah Thust, Till Eckert

05. Februar 2025

Risk categories

- Economic
- Environmental
- Geopolitical
- Societal
- Technological

2 years

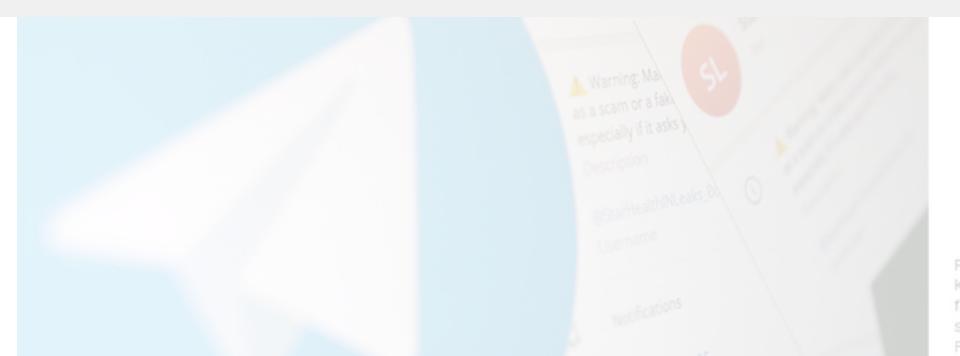
| | |
|------------------|---|
| 1 st | Misinformation and disinformation |
| 2 nd | Extreme weather events |
| 3 rd | State-based armed conflict |
| 4 th | Societal polarization |
| 5 th | Cyber espionage and warfare |
| 6 th | Pollution |
| 7 th | Inequality |
| 8 th | Involuntary migration or displacement |
| 9 th | Geoeconomic confrontation |
| 10 th | Erosion of human rights and/or civic freedoms |

10 years

| | |
|------------------|--|
| 1 st | Extreme weather events |
| 2 nd | Biodiversity loss and ecosystem collapse |
| 3 rd | Critical change to Earth systems |
| 4 th | Natural resource shortages |
| 5 th | Misinformation and disinformation |
| 6 th | Adverse outcomes of AI technologies |
| 7 th | Inequality |
| 8 th | Societal polarization |
| 9 th | Cyber espionage and warfare |
| 10 th | Pollution |

Source

World Economic Forum Global Risks Perception Survey 2024-2025.



Rechtsextreme können auf Telegram fast frei hetzen und sich vernetzen.
Foto: Dado Ruvic, Reuters



Faktencheck auch in Europa?

08.01.2025 | 19:48

Abhängige Faktenchecks in den USA ab. Experten können oder hasserfüllte Beiträge. Was bedeutet das für

Von Anekdoten hin zu langfristiger Evidenz.

Desinformation als „systemisches Risiko“ für die Demokratie. Es braucht kontinuierliches Plattformmonitoring, um:

- fortlaufend verlässliche Daten zu gewinnen,
- manipulative Dynamiken frühzeitig zu erkennen
- und deren Auswirkungen systematisch zu analysieren.

Nur belastbare Evidenzen bieten Grundlage für:

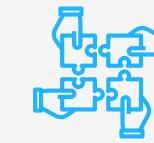
- faktenbasierte Regulierung
- wirksame Gegenmaßnahmen
- und eine stärkere Rechenschaftspflicht von Plattformen.



Aber: Es gibt zu wenig umfassende Plattformforschung, da zahlreiche Barrieren Forschende daran hindern, digitale Plattformen zu untersuchen.



Datenzugang



Fehlende Kooperation



Fehlendes Wissen



Rechtliche
Unsicherheiten



Schnell und konstant wandelndes
Informationsökosystem



Fehlende Ressourcen und Serverkapazitäten

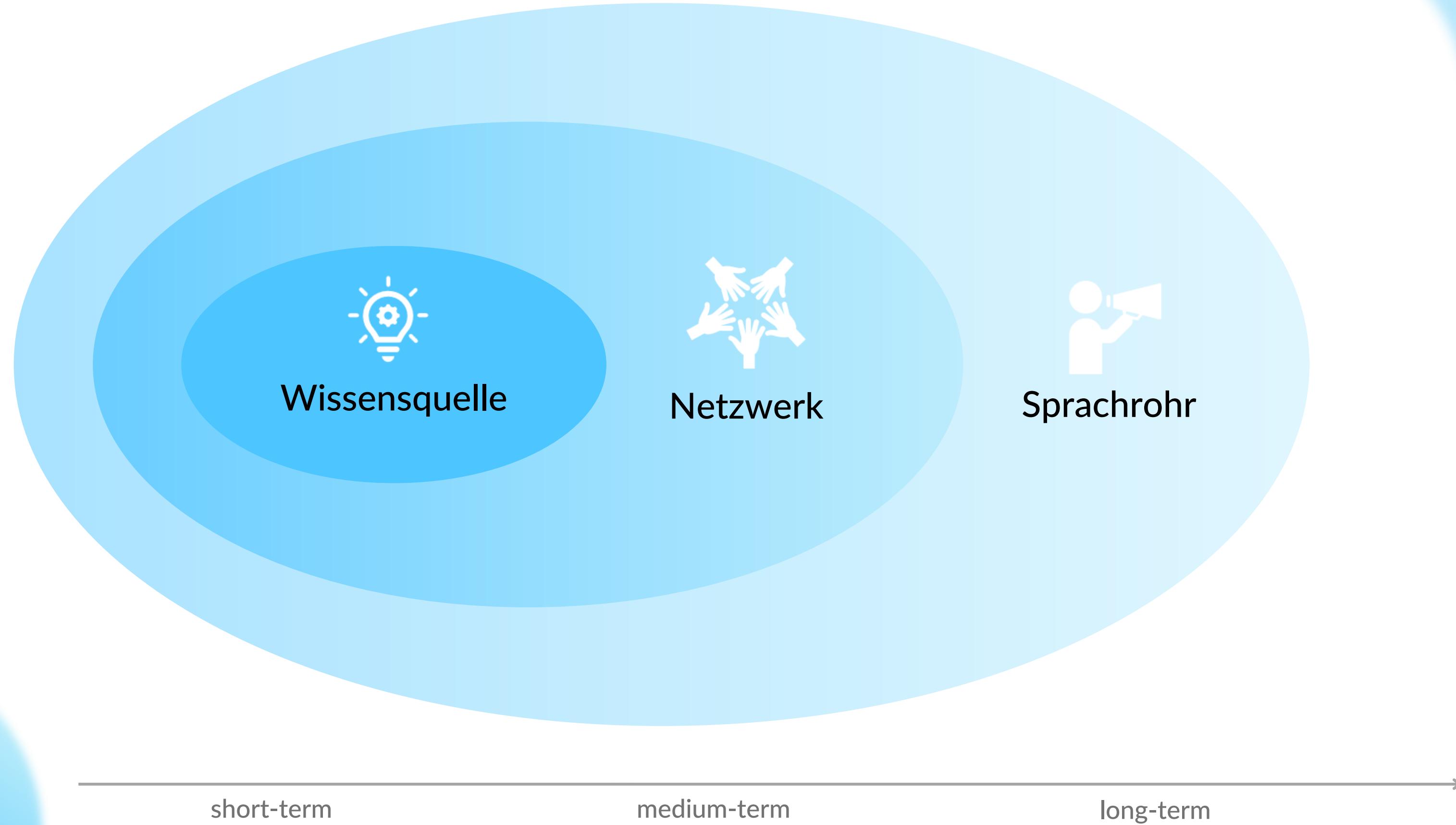
The Data Knowledge Hub

for Researching Online Discourse

Mit Social Media Daten für eine resiliente Demokratie

Data Week Leipzig | 11.06.2025

Unsere Vision: ein Data Knowledge Hub, drei Funktionen.



Was bedeutet das konkret?

Eine kollaborative Plattform, die Wissen bereitstellt, Austausch ermöglicht und als Sprachrohr für unabhängige Plattformforschung dient.



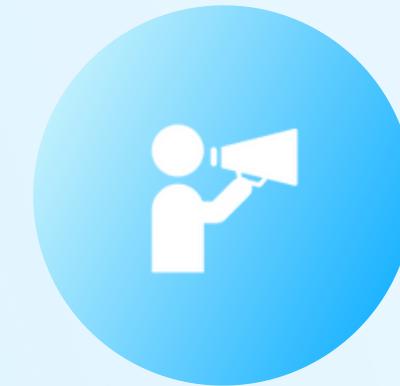
Wissensquelle

- Open-Source Plattform aufgebaut (GitHub mit Website-Spiegelung)
- Enthält Informationen über:
 - rechtliche Grundlagen und ethische Standards
 - Zugang zu Plattformen und Daten
 - Methodisches und technisches Fachwissen
 - Codebeispiele, Tutorials und Tools
 - Beispielforschung



Netzwerk

- Beiträge werden crowd-sourced
- Netzwerkaufbau
- Transfer-Workshops und Schulungen
- Gemeinsame Forschungsprojekte
- Aktuell: Koordinierter Austausch zur BTW25



Sprachrohr

- Agenda-Setting und Sprachrohr für Monitoring und Forschung:
 - Positions papier zu Datenzugang (DSA Delegated Act)
 - Austausch mit Digital Services Coordinator (BNetzA)
- Aktuell: Empfehlungen zu Desinformationsmonitoring an neue Bundesregierung



short-term

medium-term

long-term



Wie sieht das in Praxis aus?

1. Preprocessing
Text

2. Text Analysis

3. Test against
Quality Criteria

Quelle: Overview: How to analyse social media data, LMU | Data Knowledge Hub

| Method | Description | Tutorials |
|-------------------------------------|---|--|
| Dictionary | Lists of relevant features to identify things like sentiments | Tutorial R Tutorial Python |
| Semantic Network Analysis | Network analysis to analyse and visualise the co-occurrence of features | Tutorial R Tutorial Python |
| Supervised Machine Learning | Training of algorithms on manually annotated texts to classify known variables in new texts | Tutorial R Tutorial Python |
| Topic Modelling | Using feature co-occurrences to identify previously unknown topics in texts | Tutorial R Tutorial Python |
| Large Language Models (LLMs) | Using prompts to identify known variables in new texts using little or no training data | Tutorial R Tutorial Python |

Quelle: Overview: How to analyse social media data, LMU | Data Knowledge Hub

Scraping Data on TikTok with puppeteer

For this analysis we will leverage the **web version of TikTok** accessible with any browser on TikTok.com.

While there are various "unofficial APIs" or libraries available that allow you to scrape content from TikTok, the purpose of this chapter is to build a scraper from scratch.

Below, you will find a basic scraping built on **puppeteer** to open the web version of the FYP.

Tech Setup

To run this project, you should be able to open a command line on your computer and install software. For the first part you need to have **nodejs** installed. After you have created a folder and navigated to it with the console run **npm install puppeteer** in it.

A **basic scraper** that opens the TikTok website should look like this:

```
const puppeteer = require("puppeteer");

async function run() {
  // Setting up the browser
  const browser = await puppeteer.launch({
    executablePath: "/usr/bin/google-chrome", //we want to run chrome. You need to adjust this
    headless: false, // we want to see what puppeteer is doing, headless:true would mean the browser runs in the background
    defaultViewport: null,
  });
  const page = await browser.newPage();
  await page.goto("https://www.tiktok.com");
  const content = await page.content();
  console.log(content);
  await browser.close();
}
```

Quelle: Data Collection and Analysis of TikTok Hashtags with JavaScript and Phyton, SNV | Data Knowledge Hub



Mach mit.

Der Data Knowledge Hub freut sich immer über neue Daten, Codes und schriftliche Inhalte und fördert ein kollaboratives Umfeld für alle Mitwirkenden.

Please note

Right now, we would be particularly interested in including and discussing chapters on:

- Data access and ethics:
 - Data access rights beyond the European Union and the U.S.
 - How to deal with dark socials?
- Data collection: sock puppet, snowball sampling and other innovative approaches
- Examples of data collection: Facebook, Instagram, Reddit, Fediverse and others
- Examples of data analysis: Topic modelling, sentiment analysis, geospatial analysis, infrastructure as code, and others
- Additional aspects that benefit from monitoring as a research method





Mach mit.

Der Data Knowledge Hub freut sich immer über neue Daten, Codes und schriftliche Inhalte und fördert ein kollaboratives Umfeld für alle Mitwirkenden.

Schaut vorbei: data-knowledge-hub.com

Unterstützt uns: <https://github.com/bertelsmannstift/data-knowledge-hub>

Meldet euch direkt bei mir: clara.ruthardt@bertelsmann-stiftung.de

