

76 Sekunden bis zum Frauenhass

Ein Blick in TikToks Empfehlungssystem

SRF



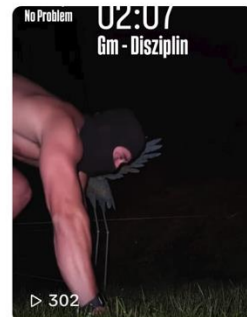
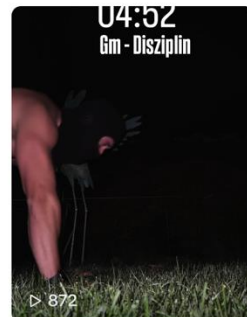
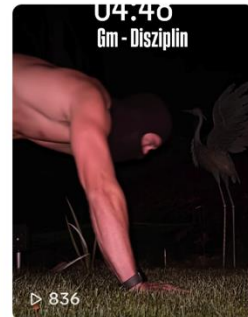
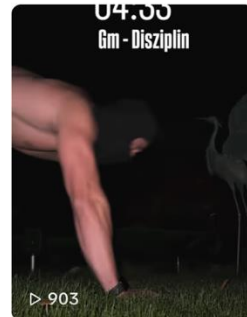
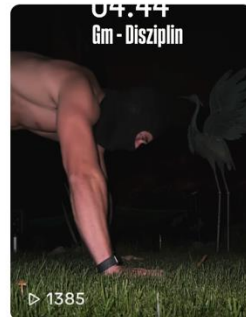
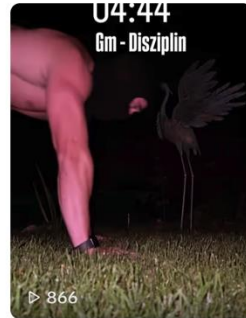
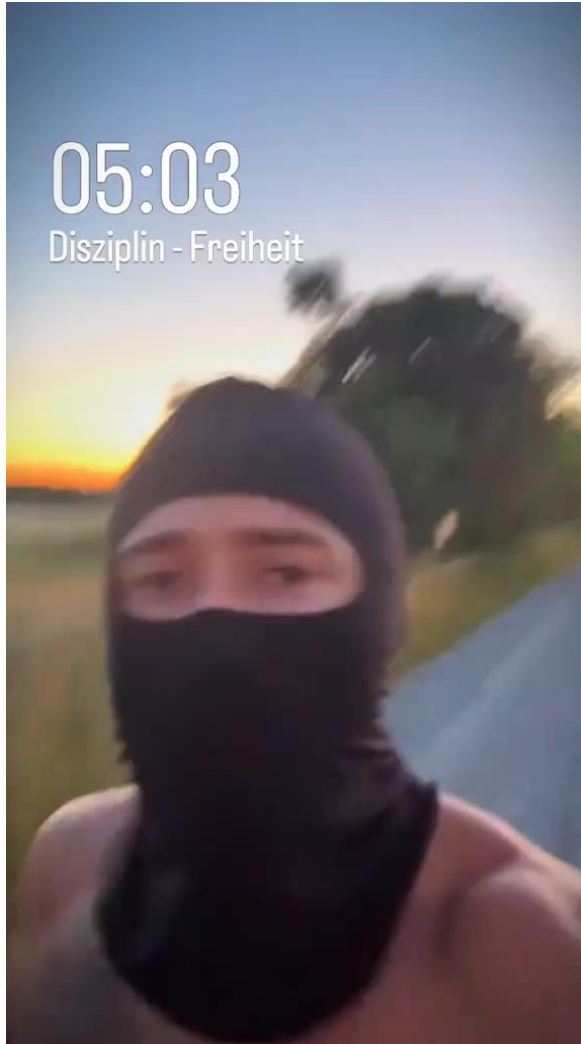
SRF DATA

Wir zeigen euch heute

1. Wie wir die problematischen Seiten von TikToks Empfehlungsalgorithmen durchleuchtet haben.
2. Warum wir dabei auf das Thema Frauenhass / Manosphere setzten.
3. Wie wir dafür ein eigenes Experiment aufgesetzt haben.
4. Wie uns LLMs geholfen haben, 120'000 Videos zu klassifizieren.
5. Wie man selbst solche Recherchen angehen kann.

Am Ende bleibt Zeit für eure Fragen.

Bonus: Einblick in die «Manosphere»



Recherchieren zu KI und Algorithmen

**KI/ML/Algo
Pipeline**

Trainings-Daten

Modell-Training

Modelle

Anwendungen

Themen

**Verlust von Privatsphäre
Überwachungs-
Kapitalismus, Arbeiter-
Ausbeutung,
Urheberrecht**

**Energie/Wasser
Nutzung, Mineralien
Extraktion,
Emissionen,
Geopolitik**

**Automatisierung,
Bias,
Halluzination /
Fehler**

**Überwachung, Job-
Verlust,
Diskriminierung,
Desinformation,
Polarisierung**

Akteure

**Firmen
Staaten**

**Firmen
Staaten
Investoren**

**Firmen
Talent
Investoren**

**Individuen
Branchen
Staaten**

Betroffene

**Individuen
Arbeiter**

**Firmen
Staaten
Bürger**

**Individuen,
Communities,
Branchen,
Unternehmen**

**Communities
Konsumenten
Demokratien
Arbeiter**

KI/ML/Algo
Pipeline

Trainings-Daten

Modell-Training

Modelle

Anwendungen

Themen

Überwachung, Job-
Verlust,
Diskriminierung,
Desinformation,
Polarisierung

Akteure

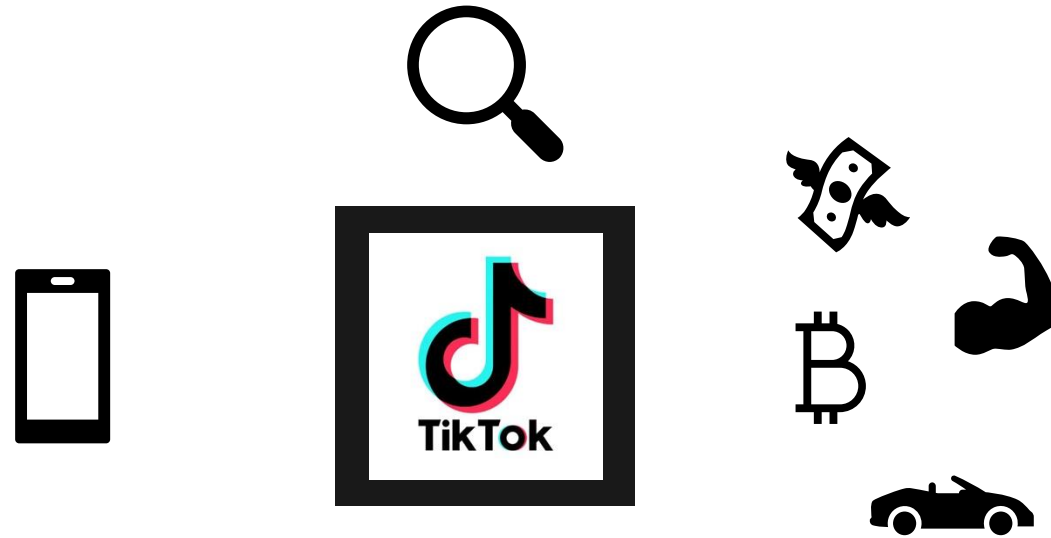
Individuen
Branchen
Staaten

Betroffene

Communities
Konsumenten
Demokratien
Arbeiter



Worum gehts?



Blackbox Algorithmus

durchleuchten und *erzählbar* machen

#algorithmic-accountability

#investigate-ai

#big-tech

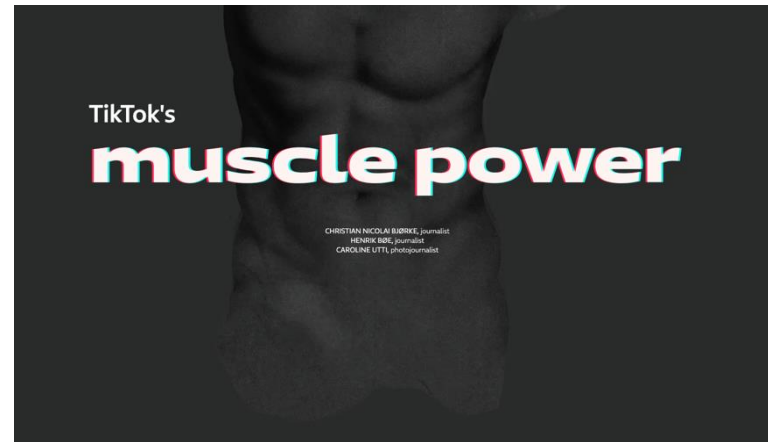
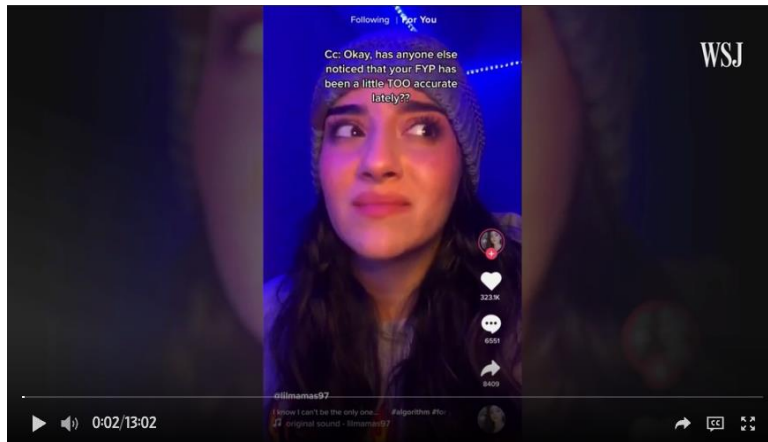
Problemstellung bei Empfehlungsalgos

- Empfehlungen können User **unbewusst und ungefragt** in Sphären locken.
- Fragmentierung der gemeinsamen Öffentlichkeit: Jeder bekommt andere Inhalte zu sehen. **Keine Übersicht.**
- Maximierung von Verweildauer = Bevorzugung extremerer Inhalte
- **Keine Transparenz** bei Gewichtung und Entscheiden.
- TikTok behauptet: Jugendliche werden besonders geschützt. **Stimmt das wirklich?**

Wir prüfen das mit einem Experiment und zeigen die Probleme auf.

Inspiration

- WSJ (2021): How TikTok's Algorithm Figures Out Your Deepest Desires
- NRK (2022): TikToks's Muscle Power
- ZDF (2023): Sex sells – auch für TikTok
- DCU Anti-Bullying Centre (2024): Recommending Toxicity Study



Die «Manosphere»

Verschiedene, sich überlappende Online-Sphären, in denen sich viel um männliche Dominanz und Anti-Feminismus dreht.



SRF

Forschungsfragen

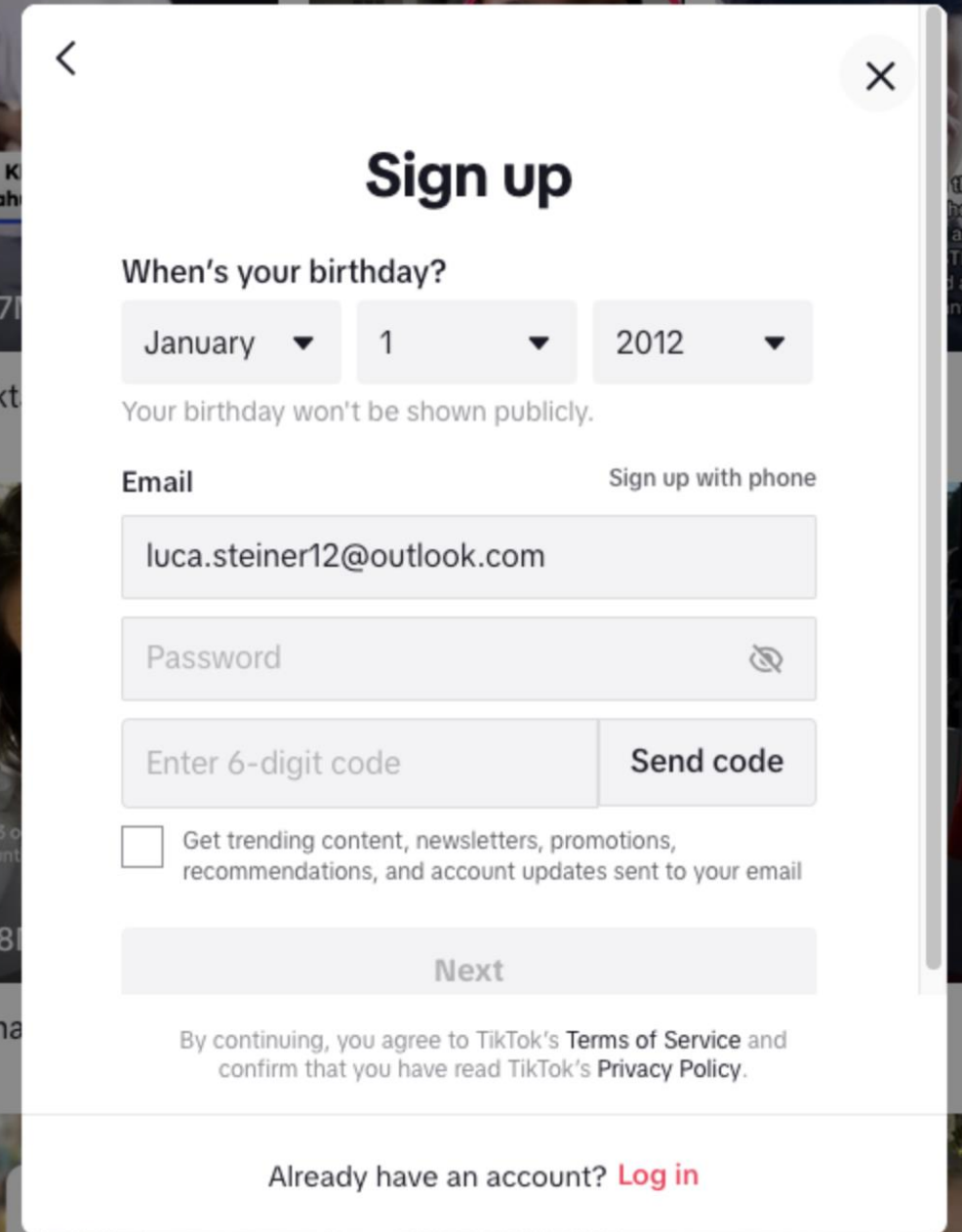
- **Wie schnell** geht es, bis *minderjährigen* Schweizern Manosphere-Inhalte empfohlen werden?
- Wenn sich die Jungs für die Inhalte *empfänglich* zeigen: Wie schnell geht es, bis die **Mehrzahl der Inhalte** aus der Manosphere stammen?
- Werden die Inhalte **über Zeit extremer**?

Experiment: In 3 Schritten zu 120'000 TikTok- Videos

Tiktoks sammeln

Sockpuppet-Accounts

- Accounts aufsetzen (manuell)
- 1 Mail-Postfach mit Aliassen (StartMail)
- Angabe von Interessen & einmalige Suche nach "sixpack" oder "reich werden"

A screenshot of the TikTok sign-up interface. The screen is titled "Sign up" and features a back arrow in the top left and a close button in the top right. The first section asks "When's your birthday?" with three dropdown menus showing "January", "1", and "2012". Below this, it states "Your birthday won't be shown publicly." The next section is for "Email" with a text input field containing "luca.steiner12@outlook.com" and a "Sign up with phone" link. Below the email field is a "Password" field with a visibility toggle icon. There are two buttons: "Enter 6-digit code" and "Send code". A checkbox is present with the text "Get trending content, newsletters, promotions, recommendations, and account updates sent to your email". At the bottom of the form is a "Next" button. Below the button, there is a disclaimer: "By continuing, you agree to TikTok's Terms of Service and confirm that you have read TikTok's Privacy Policy." At the very bottom of the screen, there is a link: "Already have an account? Log in".

Sign up

When's your birthday?

January 1 2012

Your birthday won't be shown publicly.

Email [Sign up with phone](#)

luca.steiner12@outlook.com

Password

Enter 6-digit code [Send code](#)

Get trending content, newsletters, promotions, recommendations, and account updates sent to your email

[Next](#)

By continuing, you agree to TikTok's [Terms of Service](#) and confirm that you have read TikTok's [Privacy Policy](#).

Already have an account? [Log in](#)

Automatisierung mit Appium

- **Gestensteuerung**
- **Interesse bekunden (Hashtags)**
 - **Kein Match: weiterscrollen**
 - **1 Match: 15-25 Sekunden schauen**
 - **> 1 Match 25-45 Sekunden schauen + Like**

appium/**appium**

Cross-platform automation framework for all kinds of apps, built on top of the W3C WebDriver protocol



👤 355
Contributors

🏠 7k
Used by

💬 428
Discussions

★ 21k
Stars

🍴 6k
Forks



github.com/appium/appium

```
# Topics to watch for – bot watches longer and likes if caption matches any keyword
KEYWORDS = ["topic_a", "topic_b"]
MAX_VIDEOS = 100

# Screen resolution of the physical device – run `adb shell wm size` to get yours
SCREEN_H, SCREEN_W = 2400, 1080

# Connect to the device over USB/WiFi via a local Appium server, then open TikTok
driver = connect_to_device(udid="emulator-5554", appium_port=4723)
launch_tiktok(driver)

# Find the like button once upfront and cache its pixel coordinates for reuse
like_x, like_y = find_like_button(driver)

for i in range(MAX_VIDEOS):

    # Read the current screen – extracts the video caption from the UI tree
    description = get_video_description(driver)

    # Apply your logic here – keyword match, LLM call, classifier, DB lookup, ...
    if is_relevant(description, KEYWORDS):
        like_video(driver, like_x, like_y) # tap like to signal interest
        watch_seconds = 30                # stay longer to train the algorithm
    else:
        watch_seconds = 3                 # uninteresting – move on quickly

    # — insert your own data collection / logging here —
    print(f"[{i+1}] ({watch_seconds}s) {description[:80]}")

    time.sleep(watch_seconds)            # simulate actually watching the video
    swipe_to_next_video(driver, SCREEN_H, SCREEN_W) # advance to next video

driver.quit() # release the device connection cleanly
```



Tiktoks extrahieren

Watchlist

- Export von Watchlists (manuell)
- Enthält Liste mit Zeitstempel & Link (ID)

https://www.tiktok.com/support/faq_detail?id=7543597460594285112&category=web_account

```
"Watch History": {  
  "VideoList": [  
    {  
      "Date": "2025-08-27 09:59:06",  
      "Link": "https://www.tiktokv.com/share/video/7541078978136640790/"  
    },  
    {  
      "Date": "2025-08-27 10:01:31",  
      "Link": "https://www.tiktokv.com/share/video/7541078978136640790/"  
    },  
    {  
      "Date": "2025-09-10 06:38:00",  
      "Link": "https://www.tiktokv.com/share/video/7545219626888645910/"  
    },  
    {  
      "Date": "2025-09-10 06:39:43",  
      "Link": "https://www.tiktokv.com/share/video/7537317397896662328/"  
    },  
    {  
      "Date": "2025-09-11 06:51:28",  
      "Link": "https://www.tiktokv.com/share/video/7547092554446097686/"  
    },  
    {  
      "Date": "2025-09-11 06:52:42",  
      "Link": "https://www.tiktokv.com/share/video/7547459497808252182/"  
    },  
    {  
      "Date": "2025-09-12 08:48:01",  
      "Link": "https://www.tiktokv.com/share/video/7548202465624100109/"  
    },  
    {  
      "Date": "2025-09-12 08:50:53",  
      "Link": "https://www.tiktokv.com/share/video/7541465899929324822/"  
    },  
    {  
      "Date": "2025-09-15 11:00:14",  
      "Link": "https://www.tiktokv.com/share/video/7518498191910636813/"  
    },  
    {  
      "Date": "2025-09-15 10:59:55",  
      "Link": "https://www.tiktokv.com/share/video/7546903592351845654/"  
    }  
  ]  
}
```

Tiktok scrapen

- Metadaten
- Original-Videos (MP4)
- Untertitel & Captions

- Lokale Datenbank (SQLite)
- Achtung: zeitnah verarbeiten!

dfreelon/pyktok



A simple module to collect video, text, and metadata from Tiktok.

10 Contributors 2 Issues 456 Stars 56 Forks



github.com/dfreelon/pyktok

Code-Beispiel: Metadaten

[Code auf Github](#)

```
def hydrate_videos(video_urls, db_conn, db_cursor):
    # skip URLs we already have – same video may appear across multiple participants
    db_cursor.execute("SELECT url FROM video_metadata")
    already_scraped = set(row[0] for row in db_cursor.fetchall())
    urls_to_scrape = [u for u in video_urls if u not in already_scraped]

    for url in urls_to_scrape:
        # fetch raw JSON from TikTok's internal API
        tt_json = pyk.alt_get_tiktok_json(url)

        # navigate to the video detail object
        data_slot = tt_json["__DEFAULT_SCOPE__"]["webapp.video-detail"]["itemInfo"]["itemStruct"]

        # pyktok parses the slot into a flat DataFrame row with all key fields:
        # video_id, author, description, play/like/share counts, hashtags, ...
        row = generate_data_row(data_slot).iloc[0].to_dict()

        db_cursor.execute(
            "INSERT OR IGNORE INTO video_metadata VALUES (?, ?, ?, ...)",
            [url, row.get("video_id"), row.get("video_description"), ...]
        )
        db_conn.commit()
        time.sleep(random.uniform(1, 3)) # avoid rate limiting
```

Übrigens, das Tiktok-JSON...

```
itemStruct = { # video
  "id", "desc", "createTime",
  "locationCreated", # country where video was uploaded (not creator's location)
  "textLanguage", "textTranslatable", # auto-detected language of the description
  "IsAigc", "AIGCDescription", # whether TikTok flagged this as AI-generated content
  "diversificationLabels", # TikTok's own content categories e.g. ["Random Shoot", "Comedy"]
  "CategoryType", "suggestedWords", # internal content taxonomy + search terms TikTok associates with it
  "indexEnabled", # whether TikTok is actively distributing/recommending the video

  # engagement
  "stats": {"diggCount", "shareCount", "commentCount", "playCount", "collectCount"},

  # author
  "author": { "uniqueId", "nickname", "verified", "signature" }, # signature = bio text
  "authorStats": { "followerCount", "heartCount", "videoCount", ... }, # heartCount = total likes across all videos

  # content
  "challenges": [{ "title", "desc" }], # full hashtag objects incl. descriptions
  "textExtra": [{ "hashtagName", "start", "end" }], # hashtag positions as char offsets in desc
  "music": { "title", "authorName", "original", "isCopyrighted" }, # original=True means creator's own audio
  "stickersOnItem", "effectStickers", # on-screen text stickers and AR effects used

  # video file
  "video": {
    "duration", "ratio", "format", "VQScore", # VQScore = TikTok's internal video quality score 0-1
    "subtitleInfos", # auto-generated subtitle files per language (downloadable URLs)
    "claInfo", # caption/transcript URLs – useful for text analysis without transcribing
    "playAddr", "downloadAddr",
  },

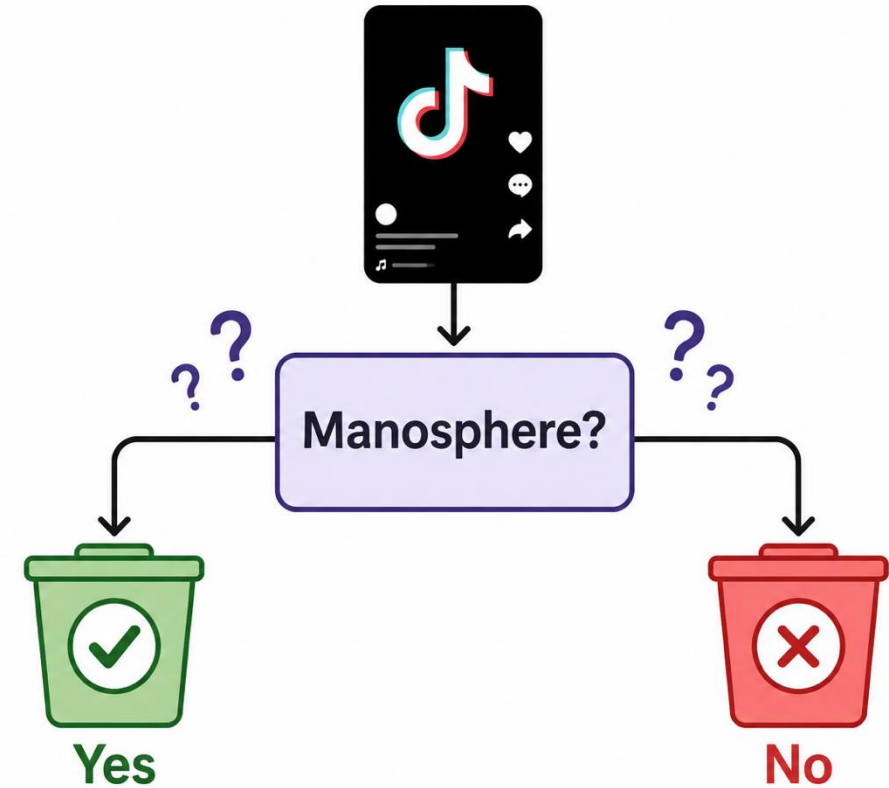
  # distribution
  "duetEnabled", "stitchEnabled", "shareEnabled", "privateItem",
```

Tiktoks analysieren

120'000 Videos... und jetzt?

Klassifizierung

- Welches Videos gehört zur Manosphere?
- Verschiedene Ebenen zur Verfügung:
Bild, Audio, Beschreibung, Captions, ...



Der Prompt

- Vorhandenen Kontext mitgeben (Beschreibung, Bilder, Untertitel, etc.)
- Prompt mit Domänenwissen und Beispielen
- Alles Wissen, das man gesammelt hat, dem Modell weitergeben
- Auch nützlich: Begründung inkl. Bildbeschreibung verlangen ("Chain-of-thought Prompting").

[Code auf Github](#)

```
# BACKGROUND
```

```
You are analyzing social media videos to identify content related to the Manosphere.
```

```
The "Manosphere" refers to online communities focused on men's issues, including:
```

- Red pill/blue pill ideology and dating advice
- Men's rights activism (MRA) and criticism of feminism
- Pickup artistry (PUA) and dating strategies
- MGTOW (Men Going Their Own Way) philosophy
- Alpha/beta/sigma male hierarchies and masculinity discussions
- Discussions of hypergamy, sexual market value (SMV), or "body count"
- Content from figures like Andrew Tate, Fresh & Fit, Jordan Pererson, Hamza Ahmed, Adin Ross, Sneaky Kevin Samuels, Nino & Elias Haralambidis (ChampLife), Karl Ess, Maximilian Pütz, Drache des Westens influencers
- Hustle/grindset mentality with hyper-masculine framing
- Emphasis on discipline, hardness, pain, pressure, "growing through pain"
- Incel and blackpill ideology
- Tradwife/traditional gender roles advocacy
- Sigma male content and lone wolf masculinity
- Traditional gender roles which see the man as provider for his wife
- Success in life defined through the accumulation of wealth, shown to the public through status symbols, watches, supercars, stashes of money, mega yachts and objectivized representations of women
- The ideas that men have to be hard and "real men" according to traditional gender roles lie at the core of the manosphere

```
"Manosphere" content on social media is often, but does not need to be characterized by:
```

- showing of status symbols, including women as status symbols (signifying a goal which can be reached through hard work)
- showing "role models" which are exceptionally "beautiful" (e.g. looksmaxing)/"rich"/"successful"/"wealthy"
- showing excessive training (e.g. running in the rain / working out outdoors in the dark)
- dark visual moods and distorted music

```
---
```

```
# GUIDELINES
```

```
## Analysis Process
```

```
Analyze both the visual content in the images as well as the following textual information: 1. the video content 2. the stickers above which represent textual elements visible in the video (if available) 3. the video description (if available).
```

```
Score the social media video on a manosphere-score from 1-5 (see details below). Use the information provided in the manosphere as well as any other knowledge about "manosphere" that you have. When scoring, take the following steps: First, check the visuals provided. Make a first evaluation. Second, fine-tune the evaluation based on the context information provided. Third, consolidate the evaluation into an output as described below.
```

```
## Classification Guidelines
```

```
**Classify as SCORE 1 (Not related at all) when:**
```

- Content has absolutely no connection to masculinity, gender dynamics, or men's issues
- Examples: cooking recipes, travel vlogs, pet videos, nature documentaries, gaming, sports commentary, DIY projects, music performances, comedy skits (unless gender-related), dance videos, fashion hauls, relationship advice for women (unless explicitly mentioning tradwife issues or traditional role models)
- The content could be created by or for anyone regardless of gender

Wieso LLM?

(Kontext!)



PNAS BRIEF REPORT POLITICAL SCIENCES OPEN ACCESS

ChatGPT outperforms crowd workers for text-annotation tasks

Fabrizio Gilardi¹, Moysam Alizadeh¹, and Maël Kubli¹
Edited by Mary Waters, Harvard University, Cambridge, MA; received March 27, 2023; accepted June 2, 2023

Many NLP applications require manual text annotations for a variety of tasks, notably to train classifiers or evaluate the performance of unsupervised models. Depending on the size and degree of complexity, the tasks may be conducted by crowd workers on platforms such as MTurk as well as trained annotators, such as research assistants. Using four samples of tweets and news articles ($n = 6,183$), we show that ChatGPT outperforms crowd workers for several annotation tasks, including relevance, stance, topics, and frame detection. Across the four datasets, the zero-shot accuracy of ChatGPT exceeds that of crowd workers by about 25 percentage points on average, while ChatGPT's intercoder agreement exceeds that of both crowd workers and trained annotators for all tasks. Moreover, the per-annotation cost of ChatGPT is less than \$0.003—about thirty times cheaper than MTurk. These results demonstrate the potential of large language models to drastically increase the efficiency of text classification.

ChatGPT | text classification | large language models | human annotations | text as data

Many NLP applications require high-quality labeled data, notably to train classifiers or evaluate the performance of unsupervised models. For example, researchers often aim to filter noisy social media data for relevance, assign texts to different topics or conceptual categories, or measure their sentiment or stance. Regardless of the specific approach used for these tasks (supervised, semisupervised, or unsupervised), labeled data are needed to build a training set or a gold standard against which performance can be assessed. Such data may be available for high-level tasks such as semantic evaluation (1). More typically, however, researchers have to conduct original annotations to ensure that the labels match their conceptual categories (2). Until recently, two main strategies were available. First, researchers can recruit and train coders, such as research assistants. Second, they can rely on crowd workers on platforms such as Amazon Mechanical Turk (MTurk). Often, these two strategies are used in combination: trained annotators create a relatively small gold standard dataset, and crowd workers are employed to increase the volume of labeled data. Trained annotators tend to produce high-quality data but involve significant costs. Crowd workers are a much cheaper and more flexible option, but the quality may be insufficient, particularly for complex tasks and languages other than English. Moreover, there have been concerns that MTurk data quality has decreased (3), while alternative platforms such as CrowdFlower and Figure Eight are no longer practicable options for academic research since they were acquired by Appen, a company that is focused on a business market.

This paper explores the potential of large language models (LLMs) for text-annotation tasks, with a focus on ChatGPT, which was released in November 2022. It demonstrates that zero-shot ChatGPT classifications (that is, without any additional training) outperform MTurk annotations at a fraction of the cost. LLMs have been shown to perform very well for a wide range of purposes, including ideological scaling (4), the classification of legislative proposals (5), the resolution of cognitive psychology tasks (6), and the simulation of human samples for survey research (7). While a few studies suggested that ChatGPT might perform text-annotation tasks of the kinds we have described (8, 9), our work provides a systematic evaluation. Our analysis relies on a sample of 6,183 documents, including tweets and news articles that we collected for a previous study (10) as well as a new sample of tweets posted in 2023. In our previous study, the texts were labeled by trained annotators (research assistants) for five different tasks: relevance, stance, topics, and two kinds of frame detection. Using the same codebooks that we developed to instruct our research assistants, we submitted the tasks to ChatGPT as zero-shot classifications, as well as to crowd workers on MTurk. We then evaluated the performance of ChatGPT against two benchmarks: i) its accuracy, relative to that of crowd workers, and ii) its intercoder agreement, relative to that of crowd workers as well as of our trained annotators. We find that across the four datasets, ChatGPT's zero-shot accuracy is higher than that of MTurk for most tasks. For all tasks, ChatGPT's intercoder

Author affiliations: ¹Department of Political Science, University of Zurich, Zurich 8050, Switzerland

Author contributions: F.G., M.A., and M.K. designed research; performed research; analyzed data, and wrote this paper. The authors declare no competing interest.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

To whom correspondence may be addressed: Email: gilardi@piz.unizh.ch.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2305016120/-DCSupplemental>.
Published July 18, 2023.

PNAS 2023 Vol. 120 No. 30 e2305016120 <https://doi.org/10.1073/pnas.2305016120> 1 of 3

<https://doi.org/10.1073/pnas.2305016120>

From Walls to Windows: Creating Transparency to Understand Filter Bubbles in Social Media

Luka Bekavac¹, Kimberly Garcia¹, Jannis Strecker¹, Simon Mayer¹ and Aurelia Tamò-Larrieux²

¹University of St. Gallen, St. Gallen, Switzerland

²University of Lausanne, Lausanne, Switzerland

Abstract

Social media platforms play a significant role in shaping public opinion and societal norms. Understanding this influence requires examining the diversity of content that users are exposed to. However, studying filter bubbles in social media recommender systems has proven challenging, despite extensive research in this area. In this work, we introduce SOAP (System for Observing and Analyzing Posts), a novel system designed to collect and analyze very large online platforms (VLOPs) data to study filter bubbles at scale. Our methodology aligns with established definitions and frameworks, allowing us to comprehensively explore and log filter bubbles data. From an input prompt referring to a topic, our system is capable of creating and navigating filter bubbles using a multimodal LLM. We demonstrate SOAP by creating three distinct filter bubbles in the feed of social media users, revealing a significant decline in topic diversity as fast as in 60min of scrolling. Furthermore, we validate the LLM analysis of posts through an inter- and intra-reliability testing. Finally, we open source SOAP as a robust tool for facilitating further empirical studies on filter bubbles in social media.

Keywords

Filter Bubbles, Social Media, Black-Box testing, Systemic Risks, Deductive Coding, VLOP, DSA

1. Introduction

The measurement of filter bubbles is a critical area of research, particularly given their potential impact on public opinion and societal polarization [1]. Michiels et al. [2] performed a review on empirical studies of filter bubbles, building upon Dahlgren [3] and Pariser [4] to propose and operationalize a systematic and empirically verifiable definition of technological filter bubble as: “a decrease in the *diversity* of a user's *recommendations* over time, in any dimension of diversity, resulting from the choices made by different *recommendation stakeholders*”. In this contribution, we refer to technological filter bubbles simply as filter bubbles.

Policymakers and the public are increasingly aware of the risks associated with filter bubbles, recognizing their potential to influence voter behavior and exacerbate societal divisions. This is fueled by recent investigations, such as the Wall Street Journal's deep dive into TikTok's

NORMalize 2024: The Second Workshop on the Normative Design and Evaluation of Recommender Systems, October 18, 2024, co-located with the ACM Conference on Recommender Systems 2024 (RecSys 2024), Bari, Italy

✉ lukajurelars.bekavac@student.unisg.ch (L. Bekavac); kimberly.garcia@unisg.ch (K. Garcia); jannisstrecker@unisg.ch (J. Strecker); simon.mayer@unisg.ch (S. Mayer); aurelia.tamo-larrieux@unil.ch (A. Tamò-Larrieux)

0009-0009-3598-3012 (L. Bekavac); 0000-0002-4971-2944 (K. Garcia); 0000-0001-7607-8064 (J. Strecker); 0000-0001-6367-3454 (S. Mayer); 0000-0003-3404-7643 (A. Tamò-Larrieux)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

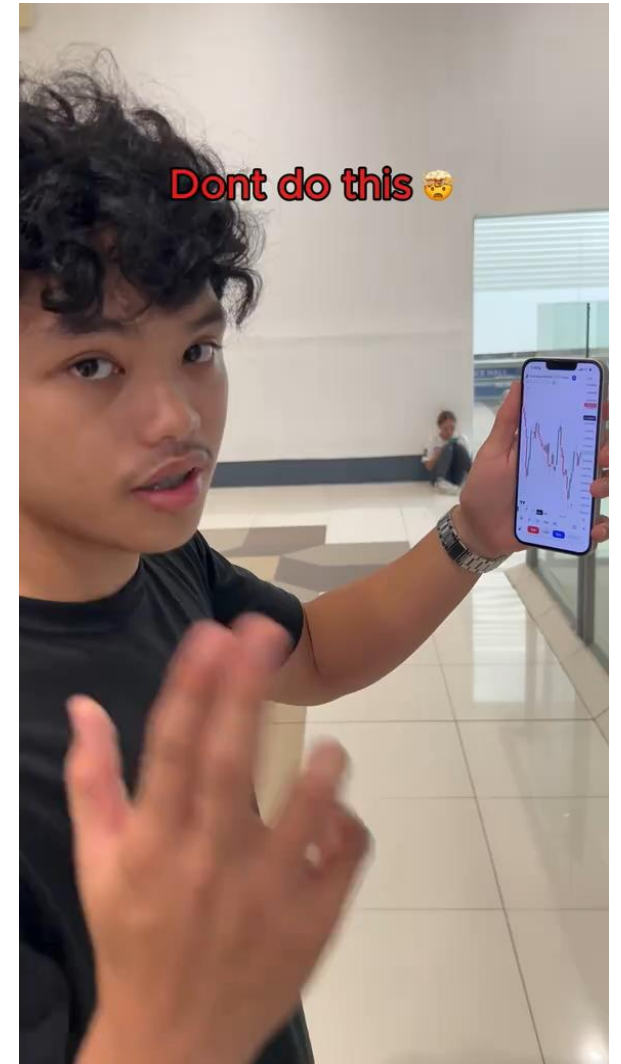
alexandria.unisg.ch/entities/publication/f41060db-3d3b-407...

Baseline: Stichwortsuche

- Beispiel: Suche nach "Alpha"



[@alpha.rizz/video/7522508827770637590](https://www.instagram.com/alpha.rizz/video/7522508827770637590)

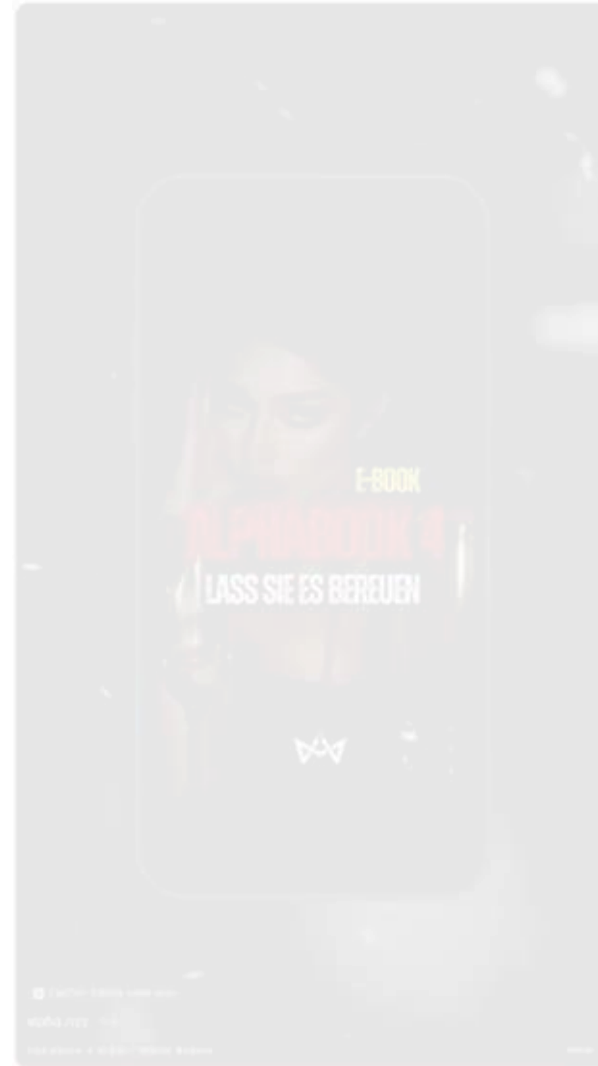


[@gainzalgo/video/7525785204313804054](https://www.instagram.com/gainzalgo/video/7525785204313804054)

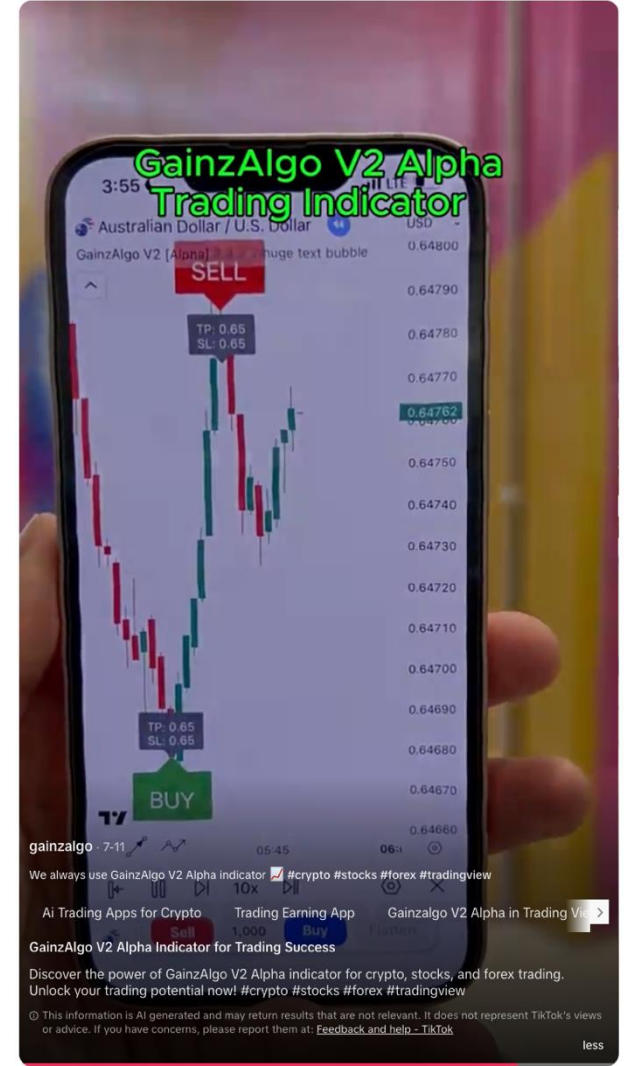
Beispiel: Alpha-Trading

"The video focuses on trading and financial indicators, specifically **showcasing a trading tool** called 'GainzAlgo v2 Alpha Indicator' for analyzing forex, crypto, and stocks. There are **no references to masculinity, gender dynamics, or any Manosphere-related themes**. The content is purely about financial trading and does not include any ideological framing or gendered messaging."

"The video shows a person holding a phone displaying trading charts and indicators. The text overlay mentions 'GainzAlgo v2 Alpha Indicator' and hashtags related to trading (#crypto, #stocks, #forex, #tradingview). There are **no visible status symbols, gendered content, or references to masculinity**."



[@alpha.rizz/video/7522508827770637590](https://www.tiktok.com/@alpha.rizz/video/7522508827770637590)

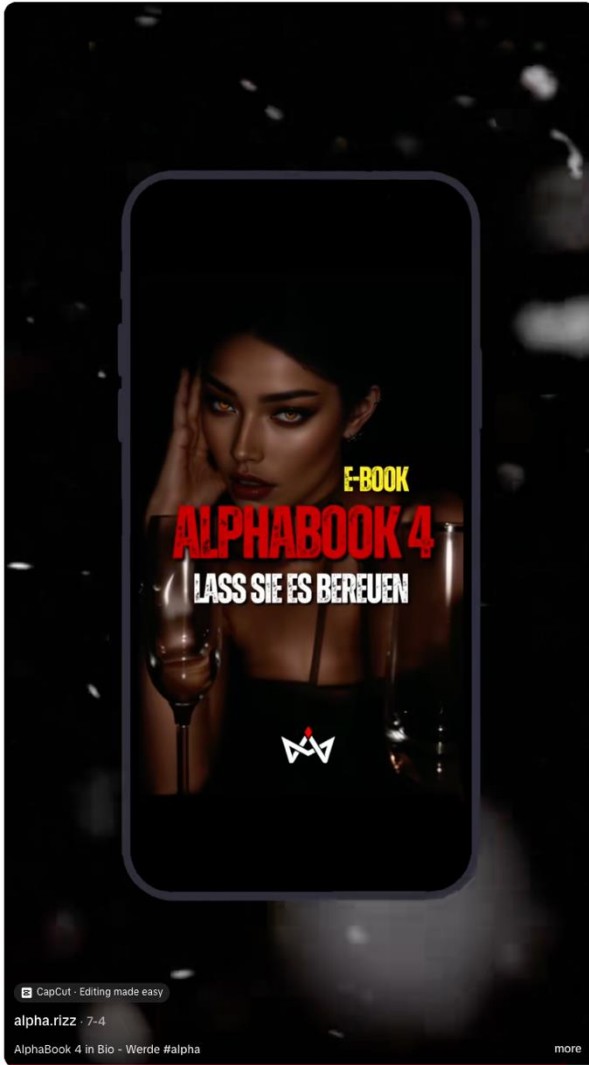


[@gainzalgo/video/7525785204313804054](https://www.tiktok.com/@gainzalgo/video/7525785204313804054)

Beispiel: Werde Alpha

"The video contains strong Manosphere elements, including the use of the term 'Alpha' in the description and subtitles, which is a key concept in the Manosphere. The subtitles **frame women as manipulative and men as victims** who need to become 'Alpha' to avoid being broken by women, which aligns with red pill ideology. The visual elements, such as the dark, moody setting, the lone figure in a desolate environment, and the use of bold text overlays, **reinforce the hyper-masculine and self-improvement themes**. The promotion of the 'Alpha Book' as a solution to become unbreakable by women further ties the content to Manosphere ideology."

"Dark, moody visuals with a lone male figure in a desolate outdoor setting, snow falling, and text overlays in bold red and white. The logo at the end appears to be a crown, which could symbolize dominance or power.



[@alpha.rizz/video/7522508827770637590](https://www.instagram.com/alpha.rizz/video/7522508827770637590)



[@gainzalgo/video/7525785204313804054](https://www.instagram.com/gainzalgo/video/7525785204313804054)

```
{
  "score": 4,
  "reasoning": "The video contains strong Manosphere elements, including the use of the term 'Alpha' in the description and subtitles, which is a key concept in the Manosphere. The subtitles frame women as manipulative and men as victims who need to become 'Alpha' to avoid being broken by women, which aligns with red pill ideology. The visual elements, such as the dark, moody setting, the lone figure in a desolate environment, and the use of bold text overlays, reinforce the hyper-masculine and self-improvement themes. The promotion of the 'Alpha Book' as a solution to become unbreakable by women further ties the content to Manosphere ideology.",
  "key_indicators": ["use of 'Alpha' terminology", "red pill ideology", "manipulative women narrative", "self-improvement to avoid emotional vulnerability", "dark visual mood", "lone male figure"],
  "visual_elements": "Dark, moody visuals with a lone male figure in a desolate outdoor setting, snow falling, and text overlays in bold red and white. The logo at the end appears to be a crown, which could symbolize dominance or power.",
  "confidence": "high",
  "category": "alpha_beta/sigma"
}
```

Können wir dem LLM vertrauen?

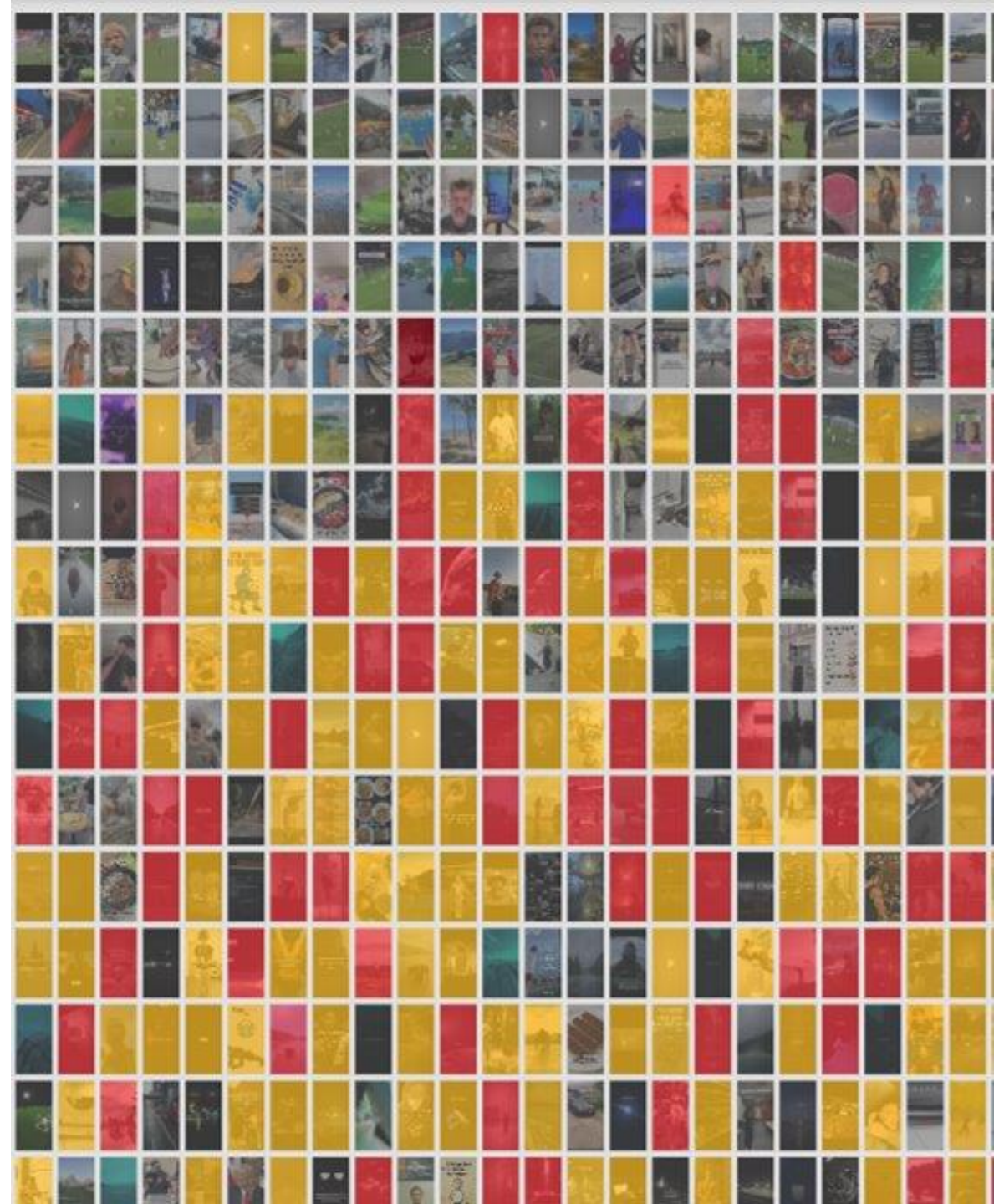
- Temperatur
- Sind wir D'accord mit den Antworten? (Interrater-Reliabilität)
- Antwortet das Modell konsistent? (Intrarater-Reliabilität)
- Begründungen & manuelle Prüfung

[Code auf Github](#)

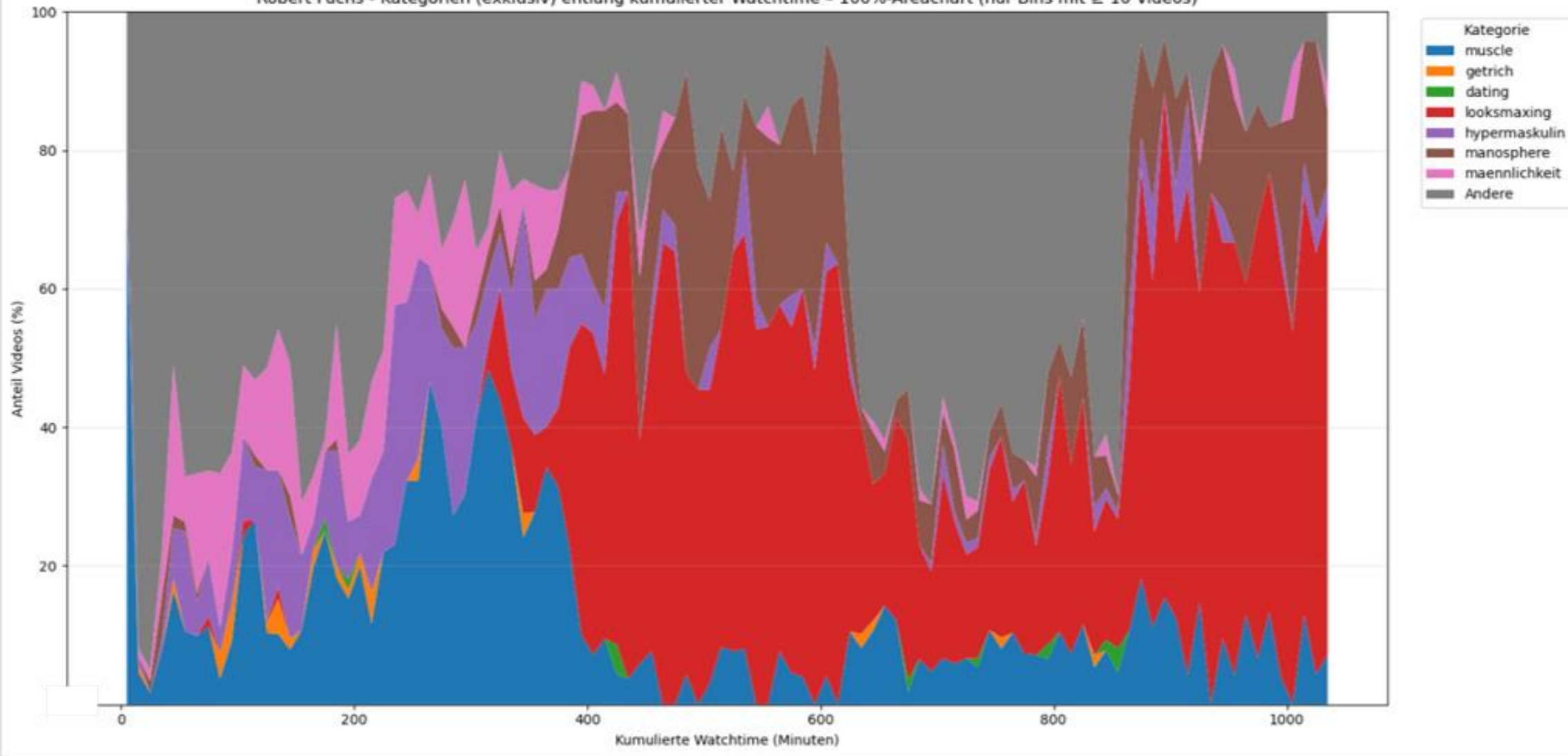
Resultate

Resultate

- Im Schnitt dauert es auf Tiktok 5 Minuten, bis einem Minderjährigen Manosphere-Inhalte serviert werden.
- Wenn er sich empfänglich zeigt, geht zw. 20 Minuten und 5 Stunden, bis der Feed mehrheitlich aus Manosphere-Inhalten besteht.
- Die ersten 500 Videos von Robert: **31% Gateway-Inhalte**, **19% Manosphere**.
- Je länger man scrollt, desto extremer werden die Inhalte.

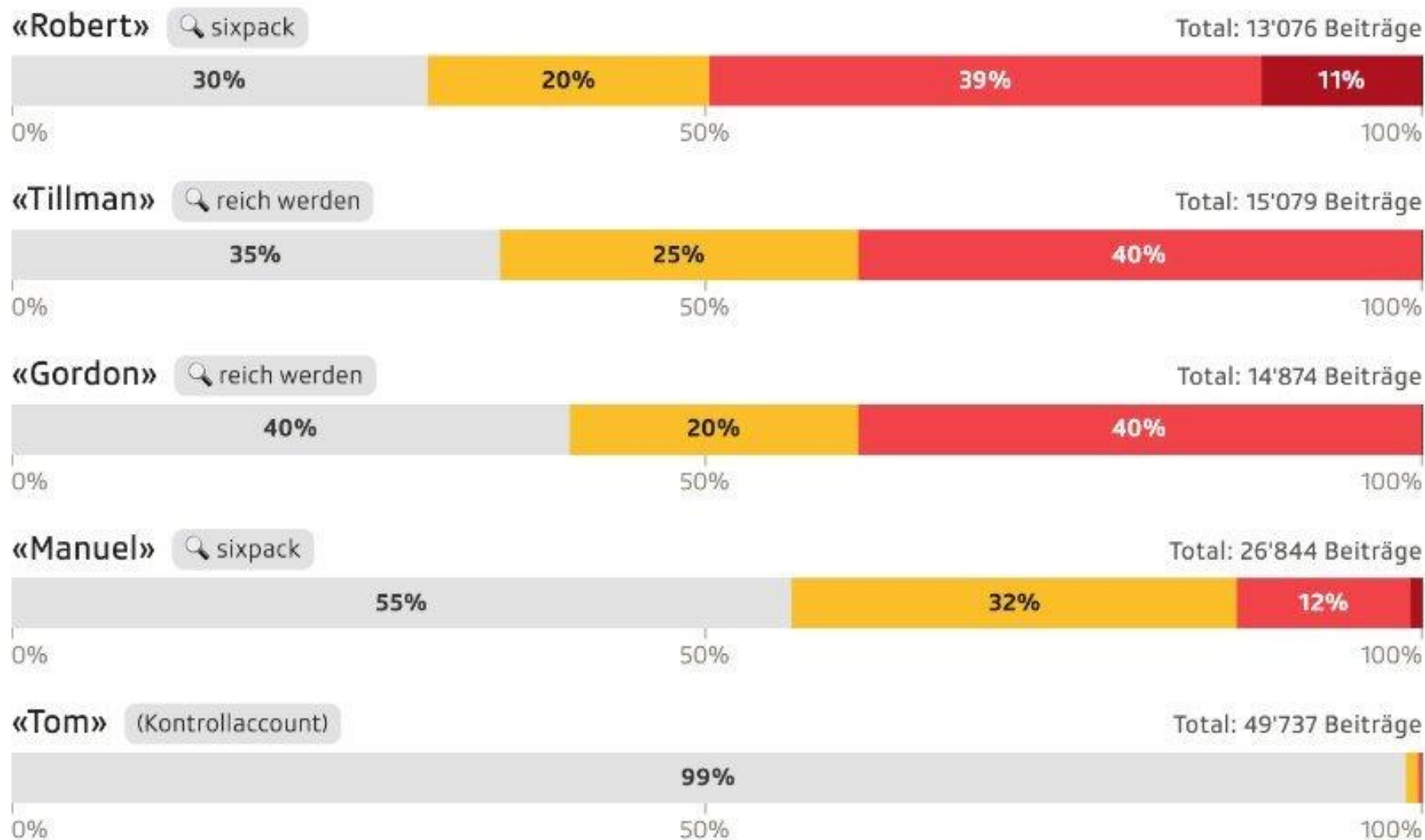


Robert Fuchs - Kategorien (exklusiv) entlang kumulierter Watchtime - 100%-Areachart (nur Bins mit ≥ 10 Videos)



Hälfte der Inhalte mit Manosphere-Bezug

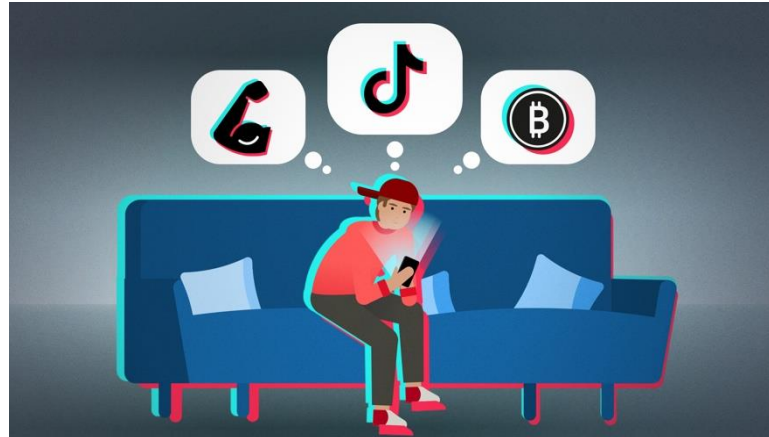
Anteile der Inhalte **ohne Bezug** zur Manosphere, **Einstiegsthemen**, **moderat-implizite** Manosphere-Inhalte, **typische** und **stark abwertende oder extreme** Inhalte pro Account.



Grafik: SRF Data. Klassifizierung mit Hilfe einer KI (Modell: GPT-4o, OpenAI).

Was ist herausgekommen?

- Podcast, Teil einer Serie zur Manosphere
- Interaktives Online-Stück
- TV-Beiträge



**Eure Fragen zu unserer
Recherche.**

Vielen Dank für die Aufmerksamkeit

srf.ch/data

julian.schmidli@srf.ch

pascal.albisser@srf.ch

SRF