Diskursreport | 5

Pilotprojekt

Decoding Antisemitism: Eine KI-gestützte Untersuchung von Hassrede und -bildern im Internet

Zentrum für Antisemitismusforschung
Technische Universität Berlin

Decoding Antisemitism

April 2023

Decoding
Antisemitism

Principal Investigator:
Dr. Matthias J. Becker
Zentrum für Antisemitismusforschung, Technische Universität (TU) Berlin

Co-Investigator:
Prof. Dr. Helena Mihaljević
Hochschule für Technik und Wirtschaft (HTW) Berlin

Forschungsteam TU & HTW Berlin:
Dr. Laura Ascone
Dr. Matthew Bolton
Alexis Chapelan
Dr. Jan Krasni
Karolina Placzynta
Milena Pustet
Marcus Scheiber
Elisabeth Steffen
Hagen Troschke
Chloé Vincent

Project Manager:
Prof. Dr. Uffa Jensen
Zentrum für Antisemitismusforschung, TU Berlin

In Zusammenarbeit mit dem HateLab der Universität Cardiff und dem Digital Humanities Department am King's College London

Gefördert durch die Alfred Landecker Foundation

Projektkoordination:
Dr. Susanne Beer

Sekretariat:
Andrea Rellin

Studentische Hilfskräfte:
Pia Haupeltshofer
Alexa Krugel
Victor Tschiskale

Technische Universität Berlin
Zentrum für Antisemitismusforschung (ZfA)
Kaiserin-Augusta-Allee 104 – 106
10553 Berlin

Kontakt: info@decoding-antisemitism.eu
Web: decoding-antisemitism.eu

Zitiervorschlag:

Chapelan, Alexis; Ascone, Laura; Becker, Matthias J.; Bolton, Matthew; Haupeltshofer, Pia; Krasni, Jan; Krugel, Alexa; Mihaljević, Helena; Placzynta, Karolina; Pustet, Milena; Scheiber, Marcus; Steffen, Elisabeth; Troschke, Hagen; Tschiskale, Victor; Vincent, Chloé (2023). *Decoding Antisemitism: An AI-driven Study on Hate Speech and Imagery Online.* Diskursreport 5. Berlin: Technische Universität Berlin. Zentrum für Antisemitismusforschung.

DOI: https://doi.org/10.14279/depositonce-17106

Wissenschaftlicher Beirat

Inhaltsübersicht

Zusammenfassung

1. Dieser Report untersucht antisemitische Internetkommentare im Kontext von drei internationalen Medienereignissen: die antisemitischen Äußerungen von Kanye West (Fokus auf Großbritannien und Frankreich), die antisemitischen Vorfälle während der Fußballweltmeisterschaft 2022 in Katar (Großbritannien) und die israelischen Parlamentswahlen (Deutschland).

2. Seit Jahren fällt Kanye West durch seine rechtsextremen und antisemitischen Äußerungen auf. Die antisemitischen Haltungen des Rappers haben sich seit Oktober 2022 noch weiter radikalisiert. Unsere Untersuchung nimmt insgesamt 3.953 Kommentare in den Blick, die als Reaktion auf Wests Äußerungen gepostet wurden. Ein Ergebnis war, dass antisemitische Reaktionen in Frankreich (14 %) häufiger vorkamen als in Großbritannien (11 %) und dass die Prozentsätze in Bezug auf die kommunizierten antisemitischen Konzepte je nach Medium stark variierten.

3. Die Analyse zeigt auch, dass britische und französische User*innen in ähnlicher Weise reagieren: Sie tendieren dazu, die antisemitischen Äußerungen von West entweder zu BESTÄTIGEN, zu RELATIVIEREN oder zu LEUGNEN. Ebenso äußern sie Vorstellungen von JÜDISCHER MACHT und antisemitische VERSCHWÖRUNGSTHEORIEN. Zur Unterstützung des Musikers greifen die User*innen zunehmend auf Umwegkommunikation zurück oder neigen dazu, den Fokus der Debatte zu verschieben, indem sie entweder eine OPFERROLLE durch eine unterstellte Zensur einnehmen oder den ANTISEMITISCHEN CHARAKTER des Vorfalls schlichtweg LEUGNEN.

4. Während der FIFA-Fußballweltmeisterschaft 2022 kam es zu verschiedenen antisemitischen Vorfällen, wie bspw. Zwischenrufen von Fußballfans, während israelische Journalist*innen Interviews führten. In Großbritannien wurden diese Ereignisse vor allem auf *Twitter* diskutiert. 10 % der 1.250 analysierten Kommentare beinhalten antisemitische Äußerungen, in denen Nutzer*innen häufig die Zuschreibungen APARTHEID-ANALOGIE, LEUGNUNG DES JÜDISCHEN SELBSTBESTIMMUNGSRECHTS und das Stereotyp des ÜBLES hervorbringen.
5. Der Sieg von Benjamin Netanjahu bei den israelischen Wahlen 2022 hat international für große Aufregung gesorgt. Über dieses Ereignis wurde in Deutschland ausführlich berichtet und es löste antisemitische Reaktionen in den Kommentarbereichen der wichtigsten deutschen Medienwebseiten sowie auf deren *Facebook*-, *Twitter*- und *YouTube*-Profilen aus. Das Korpus, das aus 2.111 Kommentaren besteht, enthält 7 % antisemitische Kommentare. Wir haben ein breites Spektrum an antisemitischen Konzepten identifiziert, wie bspw. NS- oder APARTHEID-ANALOGIEN, welche die Vorstellung von Israel als einer Entität des BÖSEN schüren.
6. Unsere Kolleginnen aus dem Data Science-Bereich an der HTW Berlin haben bestehende Ansätze zur automatischen Erkennung antisemitischer Beiträge untersucht. Der bestehende Webdienst *Perspective API*, der auf die Erkennung toxischer Sprache abzielt, wurde gegenüber bestimmten Schlüsselwörtern als voreingenommen bewertet. Gleichzeitig wurde deutlich, dass er codierte Formen von Antisemitismus nicht ohne Weiteres erkennen kann. Die Autorinnen stellen diesem Befund ihren eigenen, auf dem Transfer-Lernen beruhenden Ansatz gegenüber und präsentieren erste Ergebnisse. Sie diskutieren außerdem die Frage, wie Machine Learning-Modelle (darunter auch jenes, mit dem sie arbeiten) verwendet werden können.

1. Einleitung

Decoding Antisemitism ist ein transnationales und interdisziplinäres Forschungsprojekt, das Inhalt, Struktur und Häufigkeit von Antisemitismus in digitalen Räumen untersucht. Ziel ist es, Judenfeindschaft im Internet in ihrer ganzen Komplexität und Diversität zu verstehen und die automatische Erkennung antisemitischer Äußerungen mit Hilfe von Algorithmen deutlich zu verbessern. Im Zuge eines dynamischen Ansatzes, bei dem Antisemitismus mit Hilfe eines auf der IHRA-Definition basierenden Instrumentariums untersucht wird,[1] verfolgt das Forscher*innenteam die Entwicklung judenfeindlicher Debatten im Kontext realer Ereignisse.
Die Korpora (bzw. Datensätze) umfassen Kommentare von Internetnutzer*innen, die auf verschiedene nationale sowie internationale Ereignisse in den sozialen Medien reagieren. Dies ermöglicht es, das breite Spektrum antisemitischer Stereotype und Topoi in einer Vielzahl von Diskursen zu erfassen und detailliert zu analysieren. Unsere Studien belegen die grundlegende Plastizität und Anpassungsfähigkeit des antisemitischen Diskurses, dessen symbolische Grammatik sich ständig weiterentwickelt, um dem gestiegenen gesellschaftlichen Bewusstsein in Bezug auf Hassrede zu entgehen, aber auch um neue Anhänger*innen zu mobilisieren. Diese umfangreichen und vielgestaltigen Datensätze fließen wiederum in den Arbeitsbereich maschinellen Lernens ein, was in Zukunft zu wesentlich breiter angelegten Analysen von Online-Debatten bei gleichzeitig steigender Genauigkeit führen wird.[2] Die halbjährliche Veröffentlichung unserer Diskursreports gibt Einblicke in die Fortschritte und Zwischenergebnisse unserer Korpusanalysen.

1 – Siehe Diskursreport 2: https://decoding-antisemitism.eu/publications/second-discourse-report.

2 – Einzelheiten zu unserem Forschungsdesign finden Sie unter https://decoding-antisemitism.eu/about.

In diesem fünften Diskursreport konzentrieren wir uns auf die Auswirkungen von drei großen Ereignissen Ende 2022, über die britische, französische und deutsche Medien berichtet haben. Außerdem freuen wir uns, die ersten umfassenden Ergebnisse aus dem Prozess der KI-Entwicklung und -Prüfung vorstellen zu können, die von unseren Kolleginnen an der Hochschule für Technik und Wirtschaft (HTW) in Berlin an unseren bisher codierten Daten durchgeführt wurden. Diese vorläufigen Ergebnisse haben sowohl das große Potenzial als auch die erheblichen Herausforderungen aufgezeigt, die bei der Entwicklung maschineller Lernfähigkeiten zur Imitation der Entscheidungsprozesse von Seiten menschlicher Expert*innen bestehen.

Der erste Teil des Reports konzentriert sich auf die antisemitischen Äußerungen des Rappers Kanye West im Herbst 2022. Unsere Analyse britischer und französischer Korpora zeigt, dass Wests Artikulation klassisch antisemitischer Vorstellungen von JÜDISCHER MACHT oder VERSCHWÖRUNGSTHEORIEN von beträchtlichen Teilen der Nutzer*innen (11 bzw. 14 %) aufgegriffen und erweitert wurde – oft in Verbindung mit neueren Topoi wie KRITIKTABU oder LEUGNUNG und RELATIVIERUNG VON ANTISEMITISMUS.

Der Report analysiert darüber hinaus die Online-Reaktionen auf antisemitische Vorfälle während der FIFA-Fußballweltmeisterschaft 2022 in Katar, wie unter anderem die Anfeindung israelischer Journalist*innen. Unsere Analyse der *Twitter*-Reaktionen ergab, dass mehr als 10 % von diesen antisemitisch waren, wobei eine Reihe von israelbezogenen antisemitischen Topoi – von der APARTHEID-ANALOGIE über die Verweigerung des RECHTS AUF SELBSTBESTIMMUNG bis hin zur INSTRUMENTALISIERUNG VON ANTISEMITISMUS – besonders hervorstachen.

Bei den israelischen Parlamentswahlen im November 2022 konnte die konservative Likud-Partei von Benjamin Netanjahu einen Sieg verbuchen. Das Wahlergebnis sorgte international für große Aufregung. In den westlichen Medien und eben auch in Deutschland wurde die Sorge vor einem demokratischen Rückschritt laut. Unsere Untersuchung eines deutschsprachigen Korpus zeigt, dass User*innen diese angespannte Debatte als Gelegenheit nutzten, um antisemitische Narrative zu verbreiten. Hierbei wurde deutlich, dass die politische Geschichte Israels oft durch die Brille von NS- oder APARTHEID-ANALOGIEN gelesen wird, die den Staat Israel dämonisieren und radikal delegitimieren.

Im letzten und umfangreichsten Kapitel diskutieren unsere Kolleginnen von der HTW Berlin – Helena Mihaljević, Milena Pustet und Elisabeth Steffen – Ansätze zur automatischen Erkennung von antisemitischen Beiträgen. Sie untersuchen die Möglichkeiten und Grenzen von *Perspective API*, einem Webdienst zur Erkennung von toxischer Sprache, im Umgang mit antisemitischen Äußerungen. Ihre Untersuchungen zeigen, dass die Effektivität des Dienstes durch eine Voreingenommenheit gegenüber bestimmten Schlüsselwörtern beeinträchtigt wird und Schwierigkeiten vor allem bei der Erkennung verdeckter Formen von Antisemitismus bestehen. Sie liefern auch erste Ergebnisse aus ihren Experimenten mit modernen Ansätzen zur Textklassifizierung und erläutern, was Transfer-Lernen bedeutet und wie es sich von klassischen Zugängen unterscheidet. Die Autorinnen gehen auch auf die Zielkonflikte ein, die sich aus den unterschiedlichen Anwendungsmöglichkeiten dieser Machine Learning-Modelle ergeben.

2. Die antisemitischen Äußerungen von Kanye West im Herbst 2022

Der amerikanische Musiker, Modedesigner und Prominente Kanye West, der sich selbst jüngst als Ye bezeichnet ist eine der bekanntesten Figuren der Unterhaltungsindustrie mit Millionen Follower*innen weltweit. Sein Publikum setzt sich aus jüngeren, weniger politisierten Bevölkerungsgruppen zusammen, die seinen Botschaften direkt ausgesetzt sind. Seit Jahren sind die Beiträge des Rappers mit rechtsextremen und antisemitischen Äußerungen gespickt; die Bewunderung für Adolf Hitler und der angebliche Plan, sein Album 2018 nach diesem zu benennen, sind wohl die explizitesten Beispiele.[3] Seit Oktober 2022 ist seine antisemitische Haltung noch radikaler geworden, was sich in mehreren Verweisen auf antisemitische Verschwörungsmythen und Topoi sowie dem folgenden, inzwischen gelöschten Tweet vom 8. Oktober letzten Jahres ausdrückt:

„I'm a bit sleepy tonight but when I wake up I'm going death con 3 On JEWISH PEOPLE. The funny thing is I actually can't be Anti Semitic because black people are actually Jew also You guys have toyed with me and tried to black ball anyone whoever opposes your agenda."

Ungeachtet der Tatsache, dass West an einer bipolaren Störung leidet, die sein sprunghaftes Verhalten verursacht haben könnte, ist die Aussage selbst unverhohlen antisemitisch – einschließlich des durch ein Wortspiel ausgedrückten Todeswunsches, der LEUGNUNG VON ANTISEMITISMUS und des Stereotyps der MEINUNGSKONTROLLE. Es überrascht nicht, dass diese und ähnliche Behauptungen verschiedene Plattformen wie *Instagram* und *Twitter* dazu veranlasst haben, seine Beiträge zu verurteilen und seine Konten zu sperren. Ebenso haben *Adidas* und *Balenciaga*, zwei der Marken, die mit West zusammenarbeiteten, ihre Verträge gekündigt, und das Madame Tussauds-Museum in London hat seine Wachsfigur entfernt.

Die Eskalation des Rappers rief in Großbritannien und in Frankreich eine umfangreiche Medienberichterstattung hervor und löste antisemitische Reaktionen aus, die altbekannte Elemente antijüdischen Denkens wie JÜDISCHE MACHT oder VERSCHWÖRUNGSTHEORIEN, aber auch neuere Topoi wie KRITIKTABU oder die LEUGNUNG und RELATIVIERUNG VON ANTISEMITISMUS reproduzierten (s. Abb. 1). Um ihre Unterstützung für den Prominenten zum Ausdruck zu bringen, griffen die Nutzer*innen zunehmend auf Umwegkommunikation zurück oder versuchten, die Debatte neu zu gestalten, indem sie entweder einen durch die Zensur herbeigeführten OPFERSTATUS unterstellten oder den antisemitischen Charakter des Vorfalls rundheraus leugneten. In den folgenden Unterkapiteln stellen wir die Ergebnisse unserer qualitativen Analysen dieser konzeptuellen Merkmale in den britischen und französischen Medien vor.

3 – Für weitere Einzelheiten zu Kanye Wests Äußerungen siehe Wilson 2022 und Solomon 2023.

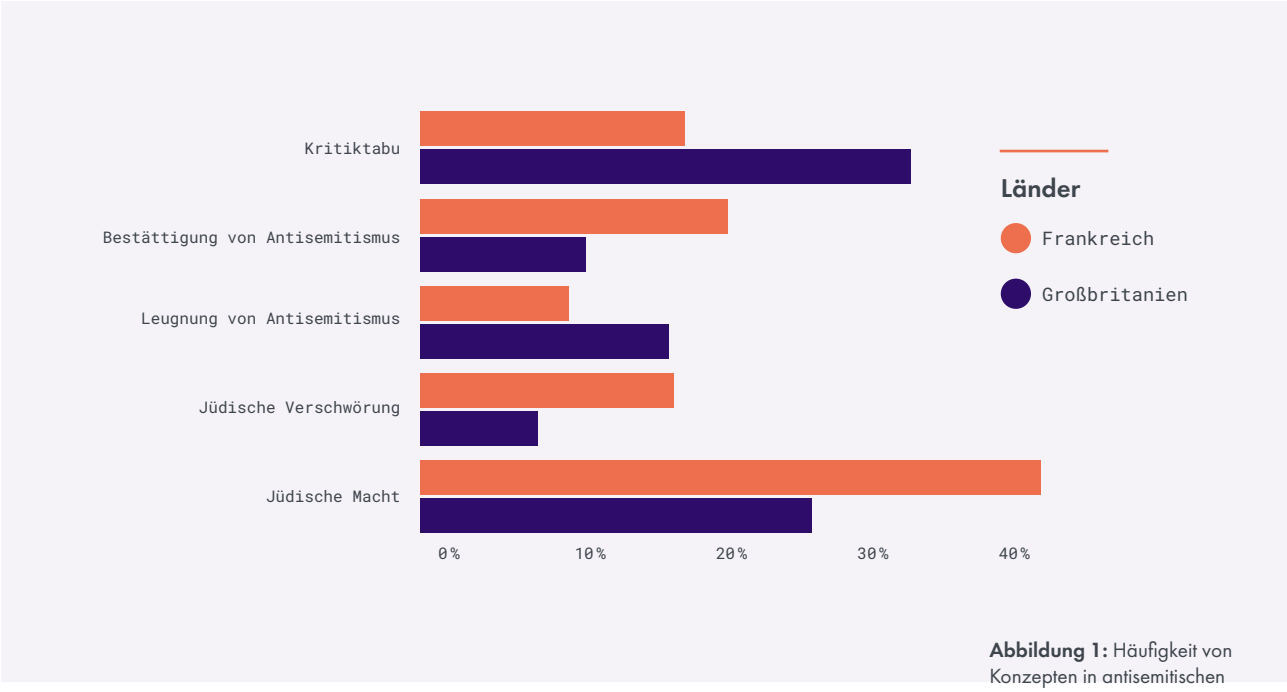

Abbildung 1: Häufigkeit von Konzepten in antisemitischen Kommentaren in den Kanye West-Korpora in Frankreich und Großbritannien.

2.1 Vereinigtes Königreich

Die Kommentar-Threads im britischen Korpus wurden von den offiziellen Social-Media-Seiten (hauptsächlich *Facebook*, mit Ausnahme eines *Twitter*-Threads der *Daily Mail*) von zehn großen britischen Medien des gesamten politischen Spektrums gesammelt, die über Wests Äußerungen berichteten – *BBC News*, *Daily Mail*, *The Guardian*, *The Independent*, *Metro*, *The Mirror*, *The Sun* und *The Times* – sowie von zwei nicht-politischen Unterhaltungsmedien, *OK!* und *Vice*. Wir wählten zwanzig Threads aus, die zwischen Oktober und November 2022 veröffentlicht wurden, und analysierten je 100 Kommentare aus einem Thread. Von den 2.000 Kommentaren wurden 11 % als antisemitisch eingestuft.

Die Unterschiede zwischen den verschiedenen Medien sind groß: Antisemitische Kommentare machten 29 % des *Twitter*-Threads der *Daily Mail*, 15 % der drei *Facebook*-Threads des *Independent*, 12 % der beiden *Facebook*-Threads von *Vice* und jeweils etwa 11 % der *Facebook*-Threads des *Guardian* und des *Daily Mirror* aus. Im Gegensatz dazu waren nur 1 % der Kommentare im *Facebook*-Threads von *The Sun* und 4 % im *Facebook*-Threads von *Metro* antisemitisch. Wir können daher vorläufig davon ausgehen, dass antisemitische Reaktionen und Unterstützung für West in *Twitter*-Diskussionen, linksliberalen Medien und kulturorientierten Medien stärker ausgeprägt waren als anderswo. Im gesamten Korpus waren die häufigsten antisemitischen Konzepte die BESTÄTIGUNG oder aber LEUGNUNG DES (von West hervorgebrachten) ANTISEMITISMUS, die Unterstellung eines KRITIKTABUS und der Vorwurf einer JÜDISCHEN MACHT sowie die Vorstellung einer JÜDISCHEN VERSCHWÖRUNG. Wie bei der prozentualen Verteilung der antisemitischen Äußerungen gab es auch bei den häufigsten Konzepten Unterschiede zwischen den verschiedenen Medien, wobei das Stereotyp der JÜDISCHEN MACHT in den Beiträgen des *Guardian* und des *Independent* besonders stark vertreten war.

Die BESTÄTIGUNG von Wests ANTISEMITISMUS kam in 42 % der antisemitischen Kommentare vor und wurde wiederholt auf zwei Arten zum Ausdruck gebracht. Die erste bezog sich direkt auf die vermeintliche Richtigkeit von Wests Worten: „Sie können ihn nicht schnell genug canceln. Er sagt zu viel Wahres für sie!“ [„They can’t cancel him fast enough. He’s speaking too much truth for them!!“] (INDEP-FB[20221009]); „Ich unterstütze, was er sagt, ich habe gerade ein ganzes Interview gesehen und es gab nichts, dem ich nicht zugestimmt habe“ [„I support what he says

I just watched a whole interview and there was nothing I didn't agree with"] (MIRROR-FB[20221021]). Diese Art der direkten Unterstützung ohne die Wiedergabe antisemitischer Stereotype bzw. ohne Erwähnung von Antisemitismus war besonders häufig vertreten. Die zweite Art der BESTÄTIGUNG waren heroische Darstellungen von West. In diesem Zusammenhang wird seine Person derart überhöht, dass Wests Antisemitismus seinem Ansehen nicht etwa schadet, sondern im Gegenteil als weitere Rechtfertigung für den Wahrheitsgehalt seiner antisemitischen Äußerungen herangezogen wird: „Helden tragen nicht immer einen Umhang. Manchmal sind sie Milliardäre" [„Heroes don't always wear capes. They sometimes are billionaires"] (MIRROR-FB[20221021]).
In einigen Kommentaren wurde die Kritik an West als absichtliche Ablenkung vom Wahrheitsgehalt seiner Behauptungen oder als zynische politische Waffe abgetan, mit der er diffamiert und zum Schweigen gebracht werden sollte. Äußerungen dieser Art nahmen oft die Form von Slogans an, wie diese sprachliche Umkehrung: „Wahrheit klingt wie Hass, für diejenigen, die die Wahrheit hassen 😏" [„truth sounds like hate, to those that hate the truth 😏"] (DAILY-FB[20221019]), oder „Sie nennen ihn antisemitisch, aber sie nennen ihn nicht einen Lügner ! 🤔" [„They call him Anti-Semitic but they dont call him a liar ! 🤔"] (DAILY-TW[20221029]). Diese Umdeutung kommt dem Begriff des politischen Antisemitismus aus dem 19. Jahrhundert in seiner ursprünglich positiven, nicht pejorativen Konnotation sehr nahe.

Die BESTÄTIGUNG VON ANTISEMITISMUS wurde häufig von dessen LEUGNUNG oder RELATIVIERUNG begleitet. Kommentare, in denen die Frage aufkam, warum West des Antisemitismus beschuldigt wurde, haben wir nicht als antisemitisch kodiert – hingegen aber solche, die den eindeutigen antisemitischen Inhalt seiner Aussagen explizit leugneten, bspw. in „Es ist nicht antisemitisch, darauf hinzuweisen" [„It's not antisemitic to point that out"] (DAILY-TW[20221029]) oder den Antisemitismus in Frage stellten bzw. negierten, indem sie diesen als „seine Meinung!!! Lasst den Mann leben!!!" [„his opinion!! Let the man live!!"] (SUN-FB[20221020]) umdeuteten. Einige Nutzer*innen stellten Wests Argumente als wahr dar:

„Wie kann es antisemitisch sein, darauf hinzuweisen, dass Juden die Medien kontrollieren oder die Macht in diesen haben? Ich denke, es ist seit Jahren allgemein bekannt, dass sie das tun"

[„How is it anti-semitic to point out that Jews control or have power in the media? I think that it has been common knowledge for years that they do"] (DAILY-TW[20221029]).

Andere versuchten, die gegen West erhobenen Antisemitismusvorwürfe als durch Anti-Schwarzen Rassismus motiviert darzustellen – womit sie sowohl implizit ihren antisemitischen Inhalt leugneten als auch explizit ein weiteres Stereotyp in den Diskurs einführten: „Seht euch all die weißen Leute hier an, die sich darüber beschweren, dass ein Schwarzer eine Meinung über reiche, korrupte jüdische Geschäftsleute hat. Verdammt rassistisch" [„Look at all the white people in here complaining about a black man having an opinion about rich corrupt Jewish businessmen. Racist af"] (GUARD-FB[20221026]). An anderer Stelle schlug ein*e Nutzer*in ironisch eine alternative Überschrift für den kommentierten Medienbeitrag vor: „Der Schwarze, der sich gegen die Juden ausspricht, muss ausgepeitscht werden" [„The black man that speaks out against the Jews needs to be flogged"] (INDEP-FB[20221027]).

Etwa ein Fünftel (21 %) der antisemitischen Kommentare drehte sich um Stereotype der JÜDISCHEN MACHT und des JÜDISCHEN EINFLUSSES, was sich aus Wests Fokus auf diese Zuschreibung heraus erklärt. Viele Nutzer*innen äußerten die Ansicht, dass Jüdinnen*Juden bestimmte gesellschaftliche Bereiche, insbesondere die Medien und das Finanzwesen, beherrschen würden: Sie würden demnach Unternehmen besitzen, Monopole halten, das Bankensystem kontrollieren und „die Schecks für alle unterschreiben" [„sign[ing] the checks for everyone"] (GUARD-FB[20221026]). Einige erweiterten entsprechende Anschuldigungen, indem sie behaupteten, es gebe eine „signifikante Überrepräsentation von Juden, zum Beispiel in der Pornografie, im Bankenwesen, im aktuellen US-Kabinett, in Hollywood, in der Justiz usw." [„significant overepresentation of Jews in, for example, pornography, banking, the current US cabinet, hollywood, law, etc."] (DAILY-TW[20221029]) – oder aber von einer totalisierenden JÜDISCHEN MACHT sprachen (im Folgenden unter Verwendung der drei Klammern, die in rechtsextremen Online-Milieus für eine implizite Bezugnahme auf Jüdinnen*Juden stehen): „(((Sie))) kontrollieren alles" [„(((They))) control everything"] (GUARD-FB[20221026]). Letztere werden dargestellt, als wären sie „sogar für die Worte, die du sagen sollst" [„even the words you're to say"] (INDEP-FB[20221029]), verantwortlich.

In einigen Kommentaren wurde der Topos des KRITIKTABUS aufgegriffen: „Je mehr er gecancelt wird, desto mehr wird sein Standpunkt bestätigt 🐑🐑🐑" [„The more he's cancelled, the more his point is validated 🐑🐑🐑"] (INDEP-FB[20221027])]), wobei mit dem Schaf-Emoji indirekt die Öffentlichkeit kritisiert wird, die scheinbar gedankenlos einer Medienagenda folgt. Oder User*innen führten ein bekanntes Zitat an, das ursprünglich aus der extremen Rechten stammt, auch wenn es oft fälschlicherweise Voltaire zugeschrieben wird: „Wenn du wissen willst, wer über dich herrscht, sieh dir an, wen du nicht kritisieren darfst" [„If you want to know who rules over you, look at who you are not allowed to criticize"] (GUARD-FB[20221026]). Andere wiesen auf eine unterstellte Doppelmoral bei der Behandlung von Jüdinnen*Juden oder Israel im Vergleich zu anderen Gruppen hin: „Nun, ich denke, wir können sagen, dass sie klargestellt haben, dass NUR Antisemitismus untragbar ist, der Rest kann aus ihrer Sicht zur Hölle fahren. Doppelte und dreifache Standards wie üblich" [„well, i guess we can say they made it clear ONLY antisemitism is intolerable, the rest can go to hell for all they care.Double triple standards as usual"] (BBC-FB[20221025]) – manchmal versehen mit einer konkreten Benennung: „Er konnte über George Floyd sagen, was er wollte, und was auch immer... aber in dem Moment, in dem er jüdische Menschen erwähnte, wurde er von allem gecancelt. Warum ist das so?!" [„He was free to say whatever he wanted about George Floyd and whatever else.. but the moment he mentioned Jewish people he's cancelled from everything. Why is that?!"] (METRO-FB[20221025]). Diese und andere Anschuldigungen wurden oft mit antizionistischen Äußerungen verbunden. In solchen Kommentaren war von einer „zionistischen Lobby" [„ Zionist lobby"] oder „zionistischen Kabale" [„Zionist cabal"] die Rede (DAILY-TW[20221029]), und es wurde gefragt:

„Glauben die Zionisten, dass sie die Richter dieser Welt sind?"

[„Do Zionists think are are the Judges of this world ??"] (TIMES-FB[20221026]).

Diese Zuschreibungen gipfeln in der Vorstellung einer JÜDISCHEN VERSCHWÖRUNG gegen Nichtjuden, erkennbar in 10 % der als antisemitisch annotierten Kommentare, wobei hier ein impliziter Träger Verwendung findet: „Die ‚J' mögen es nicht, wenn man sie darauf hinweist, dass sie alle Medien und die Banken kontrollieren, denn das bedeutet, dass sie erwischt wurden. Sie wollen, dass alle, die nicht zu den ‚J' gehören, gegeneinander kämpfen, ohne zu wissen, dass der Grund für all ihre Probleme von den ‚J' ausging" [„The ‚J' don't like it when you call them out for controlling all the media and the banks because that means they have been caught. They want everyone who isn't ‚J' to be fighting with each other not knowing the reason for all their problems was started by the ‚J'"] (INDEP-FB[20221009]). Andere Kommentare bezeichnen Jüdinnen*Juden als „fanatische Puppenspieler" [„fanatic puppeteers"] [(GUARD-FB[20221025])]) und verstärken das Bild einer mächtigen Gruppe, die im Verborgenen operiert, des „Meisters" [„master"], dem man „gehorchen" [„obey[ed]"] sollte (INDEP-FB[20221029]), oder „der Widerlinge, die die Welt regieren" [„the creeps who run the world"] (INDEP-FB[20221029]). Einige Kommentator*innen lobten West dafür, dass er die vermeintlich verborgene Wahrheit der JÜDISCHEN MACHT aufgedeckt habe: „Kanye spricht klare, überprüfbare Wahrheiten aus. So einfach ist das. Die Ironie ist, dass die Reaktion seinen Standpunkt bestätigt hat. Sie haben sich selbst entlarvt, und das nicht zu knapp. Kanye zerschlägt die Kontrollmatrix buchstäblich in Stücke 👊😄😎" [„Kanye is speaking plain verifiable truths. Simple as that. The irony is that the reaction has proved his point. They have unmasked themselves, and then some. Kanye is literally smashing the control Matrix to pieces 👊😄😎"] (INDEP-FB[20221027]). Andere meinten, dass er dadurch geschädigt werden könnte und stellten ihn als Märtyrer dar: „Er könnte verrückt sein, weil er immer noch die Dinge sagen will, die er sagt, obwohl er das Ergebnis bei Leuten gesehen hat, die es in der Vergangenheit versucht haben" [„he might be crazy for still wanting to say the things he says even tho he has seen the result from ppl who tried in the past"] (VICE-FB[20221019]). Häufig wird durch verschlüsselte Sprache auf jene Gruppe verwiesen, welche angeblich für diese Unterstellungen verantwortlich sei: „weil sein Tod bevorsteht, die Illuminati werden ihn töten" [„bcz his death is on the way illuminati will kill him"] (MIRROR-FB[20221103]); „DIE KHAZARISCHE MAFIA VERSUCHT KANYE ZU ZERSTÖREN" [„THE KHAZARIAN MAFIA IS TRYING TO DESTROY KANYE"] (TIMES-FB[20221026]). Der Verweis auf die Chasaren erinnert an einen immer häufiger anzutreffenden antisemitischen Ursprungsmythos, der die heutigen jüdischen Gemeinden als „Fälscher" oder „Betrüger" darstellt, was zu Wests eigenen, von den Black Hebrew Israelites beeinflussten Aussagen passt.[4] Andere Kommentare riefen zum Handeln gegen die angebliche Kontrolle auf: „Es ist an der Zeit, dass die Leute sich bewusst machen, was sie vorhaben, damit sich die Massen gegen sie auflehnen können" [„Its about time people started raising awareness about what they get up to so the masses can revolt against them"] (DAILY-FB[20221020]).

4 – In einem inzwischen gelöschten Tweet hat West behauptet, dass er „eigentlich nicht antisemitisch sein kann, weil Schwarze eigentlich auch Juden sind" (8. Oktober 2022). Ähnlich äußerte er sich in einem früheren *Instagram*-Post („ein Jude wie alle sogenannten Schwarzen", 6. Oktober 2022) und in einem *Fox News*-Interview („Wenn ich Jude sage, meine ich die 12 verlorenen Stämme Judas, das Blut Christi, die Menschen, die als die Rasse der Schwarzen bekannt sind, wirklich sind", 6. Oktober 2022). Siehe bspw.: https://www.timesofisrael.com/black-people-are-actually-jews-the-origins-of-kanye-wests-inflammatory-remarks (letzter Zugriff am 27. Februar 2023).

2.2 Frankreich

Das französische Korpus umfasst Threads, die wir den *Facebook* und *Twitter*-Profilen von zehn Medien entnommen haben. Darunter fallen sowohl politische Nachrichten-Outlets (*Le Point, Le Monde, Le Figaro, Le Parisien, Le Nouvel Obs, BFMTV, LCI, Les Echos, TF1*) als auch Popkultur- und Boulevardmedien (*QG, Les Inrockuptibles, French Rap US*). 1.953 Kommentare wurden im Rahmen von vier Analysegruppen untersucht. Die erste Gruppe von Medienbeiträgen befasst sich mit Wests antisemitischen Äußerungen, die zweite mit den öffentlichen Reaktionen und der Trennung der Markenpartner vom Musiker, die dritte mit Wests anschließender Eskalation in einer Reihe von Äußerungen, in denen er sich öffentlich für Adolf Hitler und das NS-Regime aussprach, und die vierte mit seiner verspäteten Entschuldigung. Insgesamt wurden 14 %, also 279 Kommentare, als antisemitisch eingestuft.

User*innen, die West gegenüber wohlgesonnen auftraten, drückten ihre Unterstützung auf vielfältige Weise aus (wobei sie oft durch Gegenrede anderer Nutzer*innen herausgefordert wurden). Anders als in unseren früheren Analysen der Reaktionen auf den französischen Komiker Dieudonné M'bala M'bala, bei denen deutlich wurde, dass dessen Verteidiger*innen häufig Insider-Witze aus den Nummern und Shows des Komikers verwendeten (siehe Becker et al. 2021), scheinen Wests Unterstützer*innen keine einheitliche „sich abgrenzende Gemeinschaft" zu bilden (vgl. Proust et al. 2020). Ähnlich wie bei den Reaktionen in den britischen Threads wurden Unterstützungsbekundungen und die BESTÄTIGUNG VON ANTISEMITISMUS oft auf einfache, direkte Weise hervorgebracht, wie etwa in „Meine volle Unterstützung für Kanye" [„Tout mon soutien a Kanye"] (LCI.F-FB[20221026]) oder „Mehr Macht für ihn 💯💯👏👏👏" [„force à lui 💯💯👏👏👏"] (BFMTV-FB[20221027]). Zahlreich verwendete Emoticons und Symbole wie Herzen oder klatschende Hände deuteten ebenfalls auf Zustimmung und Lob hin. Wie im britischen Korpus hoben einige User*innen Wests künstlerisches Talent hervor, um ihn als unverstandenen Visionär und Rebellen darzustellen, der besser als jede*r andere die inneren Mechanismen der Gesellschaft verstünde: „Mehr Macht für Kanye West. Ein unverstandenes Genie gegen die Gedankenkontrolle" [„Force a Kanye West. Un génie incompris face au dicta de la pensée"] (BFMTV-FB[20221027]);

„Er ist ein wahrer Künstler, der sich nicht der politischen Korrektheit beugt, weil er ein großartiger Mann ist"

[„C est un véritable artiste s il ne s inscrit pas dans la bien-pensance c est qu il est un grand homme"] (BFMTV-FB[20221030]).

Die antisemitischen Äußerungen Wests wurden als Weigerung interpretiert, seine „Würde" für Geld und Ruhm zu verkaufen. Ein charakteristisches Merkmal im französischen Korpus war dementsprechend der Vergleich von Wests Verhaltens mit der UNTERWÜRFIGKEIT, die französischen Mainstream-Entertainer*innen und Künstler*innen unterstellt wurde, die bestimmten Minderheiten angehören: „Wenigstens hat sich Kanye nicht wie diese nassen Lumpen gebückt, wie [der Komiker Jamel] Debbouze, usw." [„Au moins Kanye a su garder son pantalon contrairement à tt ces serpillières debouze etc"] (BFMTV-FB[20221027]). Wests vermeintliche Nonkonformität verärgere die ‚Machthaber' und setze ihn Vergeltungsabsichten aus, da er eine VERSCHWÖRUNG anprangern würde: „Ye's Meinungs- und Gedankenfreiheit ist ein Hindernis für die Verschwörung" [„Cette liberté d'expression et d'esprit de YE dérange la théorie du complot"] (BFMTV-FB[20221027]). Einige Kommentator*innen verwenden den Slogan „Je suis Kanye" in Anlehnung an „Je suis Charlie" und stellen West als Märtyrer der Meinungs- und Gewissensfreiheit dar – und seine Gegner*innen indirekt als Vertreter*innen oder Handlanger*innen von Brutalität und Terrorismus. Auf diese Weise kommt es zu einer impliziten AFFIRMATION VON Wests ANTISEMITISMUS.

West als Ikone der freien Meinungsäußerung darzustellen, die vom Establishment angegriffen wird, entspricht dem antisemitischen Stereotyp des KRITIKTABUS. Ein*e User*in stellt fest, dass

„seltsamerweise nur diejenigen, die die J * kritisieren, gecancelt und dabei als noch schlimmer als Mörder oder echte Kinderschänder dargestellt werden!!!“**

[„Bizarrement il n’y a que ceux qui critiquent les j**** qui finissent au placard, présenté comme des assassins pire que les vrais violeurs de gosses !!!“] (LEPOI-FB[20221025]).

Das Verb „kritisieren“ legitimiert die antisemitischen Äußerungen und deutet sie indirekt um, während das Adverb „seltsam“ vage verschwörungstheoretische Assoziationen auslöst: Der*die Autor*in gibt vor, über die vermeintliche Bevorzugung von Jüdinnen*Juden gegenüber anderen Minderheiten überrascht zu sein. Implikaturen, Schein-Naivität und rhetorische Fragen stellen traditionelle Elemente des verschwörungstheoretischen Diskurses dar. Die jüdische Gemeinschaft wird selten direkt, sondern regelmäßig durch vage Formulierungen erwähnt, wie etwa das „auserwählte Volk“, „die Unberührbaren“ oder im Rahmen der Behauptung, dass „es Bürger erster Klasse gibt und solche, die diese nicht ansehen oder über sie sprechen dürfen“ [„Il y a les citoyens de première zone et ceux qui ne peuvent les regarder ou parler d’eux“] (LEFIG-FB[20221025]).

Diese Übertreibung soll eine radikale Asymmetrie der Macht hervorheben und nährt die populistische Dichotomie zwischen einer imaginierten korrupten (möglicherweise jüdischen) Elite und dem reinen Volk, das im Dunkeln gehalten und seiner Würde und Rechte beraubt werde. In Anspielung auf die geopolitische Lage wird Kanye West sogar mit dem russischen Präsidenten Wladimir Putin verglichen, der ebenfalls von einem korrupten, jüdisch geführten Mediensystem dämonisiert und zum Paria gemacht worden sein solle: „Als Putin seine Meinung sagte, wurde er gehasst und wird immer noch gehasst. Er wird von demselben Moloch zermalmt. Eine einzige Gemeinschaft will über alles herrschen“ [„Quand poutine a dit cela on l’a détesté et le déteste actuellement YE ne fait que subir ce même rouleau compressor Une seule communauté veut faire le dictat“] (BFMTV-FB[20221027]).

Wie beim britischen Korpus gipfeln diese Konzepte in weitreichenden VERSCHWÖRUNGSTHEORETISCHEN Behauptungen:

„Der Typ hat sich gegen die Weltordnung ausgesprochen, er wird von den Machthabern unserer Machthaber niedergemacht. Diejenigen, die kontrollieren und dennoch behaupten, die Geschädigten in der Geschichte zu sein, nicht wahr?“

[„Le mec a parlé contre l’ordre mondial, il se fait abattre par les dirigeants des dirigeants. Ceux qui contrôlent, et qui, paraît ils d’ailleurs sont ceux insultés dans l’histoire non?“] (BFMTV-FB[20221027]).

Der biblische Topos des „auserwählten Volkes“ wird erneut beschworen, um angebliche jüdische Weltherrschaftspläne zu unterstützen: „Das passiert, wenn man versucht, die Menschheit zu kontrollieren, indem man sie für Götter oder das auserwählte Volk hält!...Nun, man muss damit rechnen, dass die Menschen revoltieren!“ [„Lorsque on essaye de contrôler l’humanité, de se prendre pour dieu ou pour le peuple élu... ! Beh il faut s’attendre à ce que les gens se révolte !“] (LEPOI-FB[20221025]). Jüdinnen*Juden wird vorgeworfen, die Medien in der Hand zu haben, aber auch die Kulturindustrie zu kontrollieren und so die öffentliche Meinung fast vollständig zu beherrschen. Wests Ausschluss lässt sich daher leicht mit der Tatsache erklären, dass „sie es sind, die die Welt, die Industrien, Hollywood, die Banken, alles [...] sogar die sozialen Medien kontrollieren“ [„c’est eux qui contrôlent le monde les industries Hollywood les banques tous quoi [...] même des réseaux sociaux !!“] (LEPOI-FB[20221025]). Die Nutzer*innen verlassen sich stark auf Anspielungen, um eine antisemitische Bedeutung zu vermitteln, bspw. durch den deiktischen Begriff „sie“, der häufig von Verschwörungsgläubigen verwendet wird, um mächtige, schattenhafte Kräfte heraufzubeschwören. Erneut weisen typografische Praktiken andere „eingeweihte“ User*innen darauf hin, dass sich der Ausdruck auf Jüdinnen*Juden bezieht, wie in dem folgenden Kommentar: „Und dann sagt man uns, es sei nicht wahr, dass (((sie))) nicht alles kontrollieren“ [„Et après on nous dit que ce n’est pas vrai, (((ils))) ne contrôlent pas tout“] (BFMTV-FB[20221027]).

Antisemitische Anspielungen finden sich auch in der Popkultur, etwa in Comics oder Mangas. So werden Jüdinnen*Juden beispielsweise wiederholt als „himmlische Drachen“ perspektiviert – ein Bezug zu einer Gruppe böser Übermenschen in Eiichiro Odas Manga-Serie *One Piece*: „Er hat Recht, er hat nichts Böses gesagt, und das weißt du auch, aber da sie himmlische Drachen sind ...“ [„Il a raison ya rien de grave et tu le sais mais comme ce sont les dragons céleste... “] (FRENC-TW[20221026]). Anspielungen dieser Art bergen das Risiko, dass Antisemitismus an eine neue, jüngere Generation weitergegeben wird, die zunächst nicht besonders politisiert ist. Ein weiterer Code, der im französischen politischen Diskurs auftaucht, ist die rhetorische Frage „Aber WER?“ („Mais QUI?“) – ein Kommunikationsmuster, das während der Covid-19-Pandemie aufkam, um das Bild einer von Jüdinnen*Juden ausgehenden Inszenierung (zugunsten von Bevölkerungskontrolle und Profit) zu suggerieren (vgl. Ascone et al. 2022). Seitdem ist diese Formulierung zu einem universellen Schlagwort geworden, das im informellen Diskurs auf angebliche jüdische Machenschaften anspielt. Einige User*innen wandten sie auf die West-Affäre an, beizeiten in Kombination mit anderen dehumanisierenden antisemitischen Stereotypen: „Hunde und Ratten arbeiten offenbar in allen westlichen Gesellschaften zusammen. Eine Elite von Pädokriminellen kontrolliert unsere Regierungen und unsere Medien... Sie alle gehorchen den Befehlen... Aber wessen???“ [„Les chiens et les rats se mettent d’accord dans toutes les sociétés occidentales apparemment. Une élite de pedocriminels qui décident dans nos gouvernements et nos médias..... ils sont tous ‚aux ordres‘ .... mais de qui ????“] (BFMTV-FB[20221027]). Der Verweis auf Pädophilie fügt sich in das breitere QAnon-Narrativ ein, das seinerseits auf subtile Weise Elemente der RITUALMORDLEGENDE aufgreift (vgl. Friendberg 2020).

3. Antisemitische Vorfälle bei der FIFA-Fußball-Weltmeisterschaft 2022 (Vereinigtes Königreich)

Die palästinensische Flagge war während der Fußballweltmeisterschaft 2022 in Katar unübersehbar. Auf den Tribünen schwenkten die Fans – insbesondere, aber nicht ausschließlich, aus der umliegenden arabischen und nordafrikanischen Region – die Flagge und sangen pro-palästinensische Lieder und Sprechchöre. Auf dem Spielfeld zeigte die marokkanische Mannschaft die palästinensische Flagge, als sie ihre beiden Siege feierte. Außerhalb des Stadions gab es zahlreiche Berichte über israelische Journalist*innen, die von den Fans belästigt und angefeindet wurden. Die Vorfälle wurden auf *Twitter* heftig diskutiert, wo auf Tweets von Journalist*innen und unabhängigen Kommentator*innen hunderte, teils sogar tausende von Antworten folgten. Die zehn Threads, die wir für die Analyse identifizierten, stammten von einer Vielzahl von Konten mit hoher Follower*innenzahl: darunter der ESPN.co.uk-Chef-Fußballjournalist Mark Ogden (268.000 Follower*innen), die palästinensische Politikanalystin Dr. Yara Hawari (82.000 Follower*innen) und der amerikanisch-israelische Philanthrop Adam Milstein (171.000 Follower*innen). Wir kodierten die ersten 125 Kommentare jedes Threads, was ein qualitativ analysiertes Korpus von 1.250 Kommentaren ergibt. Im gesamten Korpus wurden 10 % der Kommentare als antisemitisch kategorisiert. Die häufigsten antisemitischen Begriffe waren APARTHEID-ANALOGIE, LEUGNUNG DES JÜDISCHEN SELBSTBESTIMMUNGSRECHTS und das Stereotyp des ÜBELS.

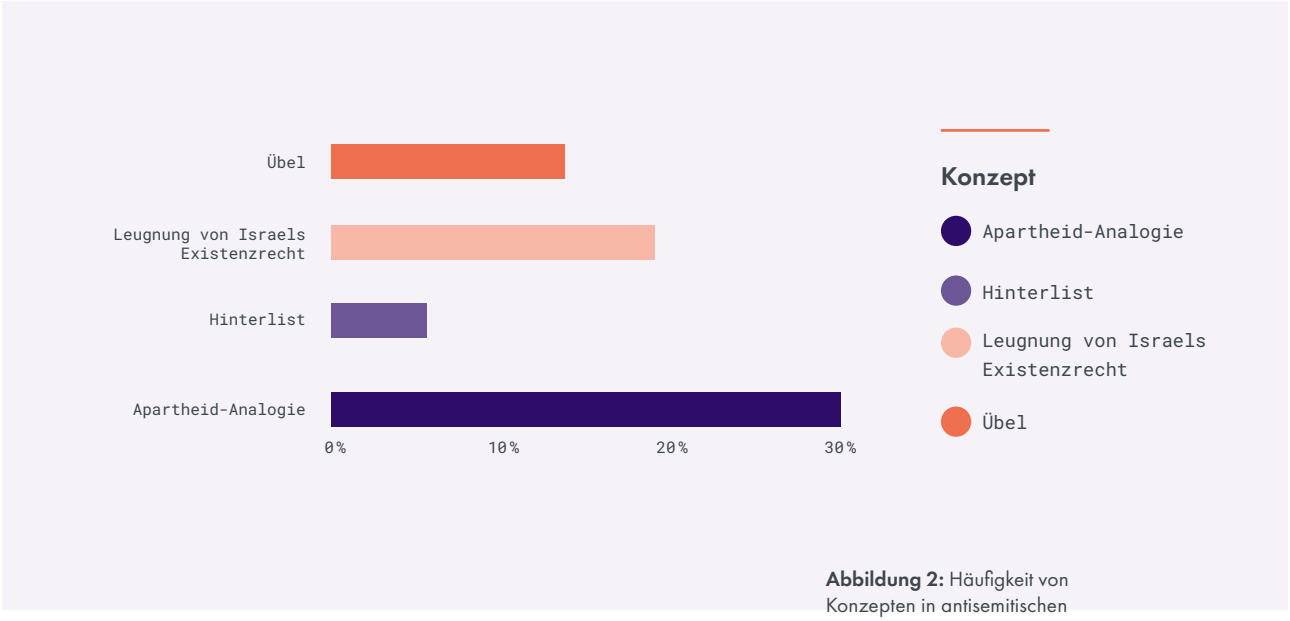

Abbildung 2: Häufigkeit von Konzepten in antisemitischen Kommentaren aus dem FIFA-Korpus in Großbritannien.

In den Threads wurde Israel häufig als „Apartheid-Regime“ [„apartheid regime“] (TW-JENNINE[20221126]), „zionistische Apartheid“ [„Zionist apartheid“] (TW-OGDEN[20221206]) oder „blutrünstiger Apartheidstaat“ [„bloodthirsty apartheid state“] (TW-JASKOLL[20221126]) bezeichnet. Mehrere Kommentare versuchten, der APARTHEID-ANALOGIE mehr Nachdruck zu verleihen, indem sie sagten, die Ansicht werde von der „allgemeinen Öffentlichkeit“ [„the general public“] geteilt, die „das Apartheid-Israel von ganzem Herzen ablehnt“ [„wholeheartedly dislikes apartheid Israel“] (TW-JASKOLL[20221126]), oder sich auf die Autorität der internationalen Gemeinschaft und ihrer Institutionen beriefen: „Israel ist ein systemisches Apartheidregime, das sich nicht von der weißen Herrschaft Südafrikas unterscheidet. 98 Länder stimmen in der jüngsten UN-Resolution zu“ [„Israel is a systemic apartheid regime no different to White South Africa rule. 98 countries agree at the recent UN resolution“] (TW-JASKOLL[20221126]). Wenn auf den Antisemitismus solcher Aussagen hingewiesen wurde, war die Reaktion oft Verachtung oder Feindseligkeit: „Es ist seltsam, dass du überrascht bist, dass niemand außerhalb ihrer Apartheid-Blase Tyrannen und Massenmörder mag“ [„It’s weird that your surprised that no one outside their apartheid bubble likes bullies and mass murderers“] (TW-JENNINE[20221126]), hier noch verstärkt durch die Aktivierung der NS-ANALOGIE:

„Du verlierst den Verstand, wenn du versuchst, ein völkermörderisches Apartheid-Regime zu verteidigen... verlasse deine zionistische Echokammer für eine Sekunde und du wirst erkennen, dass du die Meinung der Öffentlichkeit längst verloren hast.😂Ihr seid die Unterdrücker und niemand mag oder respektiert euch. Die Nazis von heute, nur schwächer“

[„You’re losing your mind trying to defend a genocidal apartheid regime.. step out of your Zionist echo chamber for a second and you’ll realize that you’ve long lost the public’s opinion. 😂You’re the oppressor and nobody likes or respects you. Modern day Nazis, but weaker“] (TW-JASKOLL[20221126]).

Die Infragestellung der Legitimität Israels vollzog sich, indem das Land als „zionazistische Entität“ [„zionazi entity“] (TW-HARAWI[20221127]), „Israelische GmbH“ [„Israeli LLC company“] (TW-MILSTEIN[20221127]), „nicht mal ein Land“ [„not even a country“] (TW-CARTER[20221210]) oder – in Begleitung einer Drohung – „eine 73 Jahre alte Kolonie, die es nicht mehr geben wird, bevor sie 100 Jahre alt wird“ [„a 73 years old colony that won’t be a thing before it turns 100“ (TW-AMRO[20221206]) bezeichnet wurde.

In anderen Kommentaren wurde Israel als prototypisch BÖSE und als Feind der Menschheit dargestellt: „Der Hass auf Israelis kommt automatisch, wenn man ein menschliches Wesen ist" [„Hatred for Israelis will come automatically if you are a human being"] (TW-MILSTEIN[20221127]). In einigen Kommentaren wurde der jahrhundertelalte Vorwurf des KINDERMORDES erhoben – mit einem unverhohlenen „Babykiller-Schlampen" [„baby killer bitches"] (TW-OGDEN[20221206]), einem sarkastischen „Du wurdest als Kind schikaniert, also befürwortest du die Ermordung von Babys. Wie niedlich herr kleine schwanze" [„u Got bullied in kid so u endroce killing of babies. How cute mr kleine schwanze"][5] (TW-CARTER[20221204]) oder trotzig kommuniziert mittels „Kindermord und illegale Besatzung sind falsch und ihr könnt mich als ALLES bezeichnen, was ihr wollt, weil ich das sage" [„Murder of children and illegal occupation is wrong and you can label me ANYTHING YOU WANT for saying that"] (TW-JENNINE[20221126]).

Gelegentlich wurde diese Behauptung mit weiteren Analogien zwischen Israel und dem Nationalsozialismus kombiniert: „Schauen dir an, was Israel jeden Tag kleinen Kindern und Familien antut, die nicht in Frieden leben können, und überlege, wer die wahren Nazis sind" [„Look at what Israel is doing to little children and families every day that can't live in peace then think who the real nazis are"] (TW-MILSTEIN[20221127]); „Wie kann man das Töten von Tausenden von Kindern verteidigen. Sie aus ihren Häusern zu vertreiben und sie auf der Straße abzuschlachten. [...] Juden sind schlimmer als Nazis. Holocaust des 21. Jahrhunderts" [„How do you defend killing thousands of children. Displacing them from homes and butchering them in the streets. [...] Jews are worst than Nazis. 21st century holocaust"] (TW-AMRO20221206]).

Mit Blick auf jene Fans, die es ablehnten, sich von israelischen Reporter*innen interviewen zu lassen, stellten die Kommentator*innen letztere oft als HINTERLISTIG dar und behaupteten, dass „diese israelischen Journalisten für Clickbait da sind. Das sind genau die Interaktionen und Antworten, nach denen sie suchen, damit sie das Opfer spielen und den Teufel an die Wand malen können" [„these israeli journalists are there for the clickbait. These are the exact interactions and answers they're looking for so that they can play victim and cry wolf"] (TW-HARAWI[20221127]). Auch wird behauptet, sie seien „auf einer Mission [...], über Antisemitismus zu berichten, um ihre eigene Agenda zu unterstützen und weiter voranzutreiben" [„[t]hey're on a mission to report antisemitism to support and further push their own agenda"] (TW-JENNINE[20221126]). Das Ziel sei hierbei also, „ein antisemitisches Narrativ nach dem Motto ‚die ganze Welt hasst uns' zu kreieren" [„create an entire ‚the whole world hates us' anti-Semitic narrative"] und „körperliche Angriffe gegen sich selbst zu provozieren" [„try to get themselves physically assaulted"] (TW-HARAWI[20221127]) – mit anderen Worten: die Journalist*innen suchen laut User*innen über manipulatives Verhalten, ANTISEMITISMUS für ihren eigenen Vorteil zu INSTRUMENTALISIEREN.

5 – Die deutschsprachige Beleidigung am Ende des Kommentars spielt mit dem Namen des Nutzers, auf den er antwortet. Charakteristisch für den Diskurs ist die Vermischung von schwerwiegenden antisemitischen Anschuldigungen mit persönlichen Beleidigungen, während direkte antisemitische Beleidigungen weitgehend fehlen (möglicherweise als Folge der Moderation).

4. Die israelischen Wahlen im November 2022 (Deutschland)

Am 1. November 2022 fanden die Parlamentswahlen in Israel statt, bei denen die konservative Likud-Partei unter der Führung von Benjamin Netanjahu die Mehrheit der Stimmen erhielt. Während die deutschen Medien lange vor diesem Ereignis über die Wahlen in Israel berichteten, skizzierten jene Artikel, welche nun die meisten Kommentare generierten, die Aussicht, dass sich in Israel eine rechte oder sogar rechtsradikale Regierung formiert. Die Schlagzeilen der Medien konzentrierten sich häufig auf die Ideologie der erfolgreichen Parteien mit Ausdrücken wie „rechtskonservativ", „Rechtsruck", „national-religiöse Allianz" – all dies Beschreibungen, die auch in den meisten israelkritischen Kommentaren auftauchten, einschließlich jenen mit antisemitischer Schlagseite.

Wir nahmen Threads der Medien-Webseiten *Zeit, Spiegel, ZDF, DW, Arte, Welt* und *Tagesschau*, sowie deren Social Media-Profile bei *Facebook*, *YouTube* und *Twitter* in den Blick. Das Korpus bestehend aus 19 Threads und insgesamt 2.111 annotierten Kommentaren enthielt fast 7 % antisemitische Kommentare. Außerdem gab es einen relativ hohen Anteil von 4 % an Kommentaren, in dem Gegenrede geäußert wurde und der sich offensichtlich auf vorausgehende Kommentare bezog, die allerdings von den jeweiligen Moderator*innen bereits gelöscht wurden. Die hohe Anzahl gelöschter Kommentare auf der einen und Kommentare mit Gegenrede auf der anderen Seite führt indirekt vor Augen, in welchem Umfang dieses Diskursereignis antisemitische (oder allgemein hasserfüllte) Kommentare auslöst. In den annotierten Threads ließen sich insbesondere folgende antisemitische Konzepte finden: Israel als das BÖSE, die UNFÄHIGKEIT, AUS DER VERGANGENHEIT ZU LERNEN, KRITIKTABU, NS- und APARTHEID-ANALOGIEN und die Behauptung, JUDEN SEIEN AM ANTISEMITISMUS SELBST SCHULD.

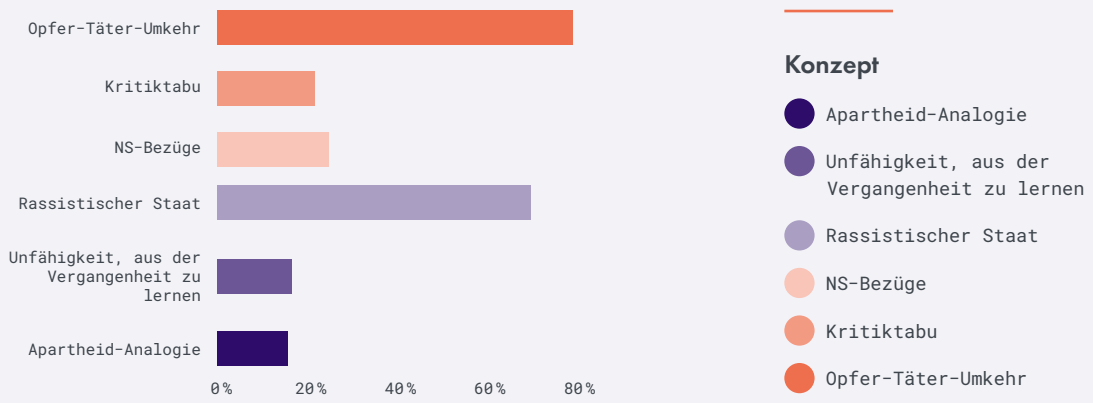

Abbildung 3: Häufigkeit von Konzepten in antisemitischen Kommentaren aus dem Korpus der israelischen Wahlen in Deutschland.

Das antisemitische Stereotyp des BÖSEN ist mit dem Bild verbunden, dass Jüdinnen*Juden blutrünstig seien und ohne jeden rationalen Grund entsprechend handeln würden. Im Einklang mit diesem Konzept heißt es in einem Kommentar: „Naja, Israel ist ja auch derzeit derart beschäftigt wieder mal Bömbchen en masse zu werfen und deren Abwürfe zu coachen, zu überwachen und zu begleiten“ (TAGESFB[20221215]). Darüber hinaus erscheint in Kommentaren die Vorstellung des BÖSEN in Kombination mit diversen Strategien – wie im Folgenden mit einer impliziten OPFER-TÄTER-UMKEHR:

„Ghettos werden gebaut und es wird je nach Tageslaune entschieden, wie viel Schrott und Entwicklung ins Ghetto fließt. Wenn aufgemuckt wird, wird der Nachschub abgeschnitten und ein paar Häuser werden als Beispiel bombardiert“ (ZEIT[20221101]).

In diesem Kommentar wird implizit behauptet, dass Israel die Verbrechen NS-Deutschlands wiederholen würde. Diese Vorstellung wird bei den Leser*innen durch Rückgriff auf Weltwissen aktiviert und einerseits mittels der Anspielung „Ghetto“, andererseits durch Formulierungen wie „Versorgung abgeschnitten“ und „Häuser als Vergeltung bombardiert“ ausgelöst. Eine vermeintlich unverhältnismäßige Reaktion auf Rebellion evoziert zusätzlich das Stereotyp der JÜDISCHEN RACHSUCHT.

In anderen Kommentaren wird das Konzept des BÖSEN mit der Vorstellung des KRITIKTABUS verbunden, das von Jüdinnen*Juden und jüdischen Institutionen ausgehen würde. Entsprechend verknüpft der folgende Kommentar die in diversen Medienartikeln geübte Kritik an ultraorthodoxen Parteien mit einer stereotypen Behauptung, bei der die Konzepte des BÖSEN und des KRITIKTABUS zusammentreffen:

„Das gilt natürlich nur für Israel. Dort darf das Böse nicht direkt als ‚böse‘ bezeichnet werden. So wie ‚ultra-religiöse‘ Menschen in anderen Ländern als rückständige, religiöse Fundamentalisten bezeichnet werden“ (ZEIT[20221104]).

Jene Kommentare, mittels welcher das Konzept der UNFÄHIGKEIT, AUS DER VERGANGENHEIT ZU LERNEN, reproduziert wird, überraschen teils mit ihrer expliziten Gestalt: „Nichts aus der eigenen Geschichte ihrer Vorfahren gelernt“ (ZEITIG[20221102]). Teils kommunizieren User*innen über dieses Konzept aber auch ihre Sorge in Bezug auf die ‚wahre‘ Gefahr: „Falls die radikalen Kräfte an die Macht kommen, bleibt nur zu hoffen, dass der Westen und vor allem Amerika die Integrität hat, einen Genozid zu verhindern von dem Volk, was eigentlich historisch wissen sollte, was es bedeutet, aufgrund von absurden Ideologien verfolgt zu werden“ (ZEIT[20221101]).

Ebenso kann sich dieses Stereotyp hinter dem Ausdruck von Mitgefühl und Verwunderung verbergen:

> „Ich wundere mich manchmal über die Israelis, obwohl sie selbst so viel Leid erfahren haben, wieviel Leid sie selbst bringen, auch hier sieht man, wie bei allen Kulturen, dass die Menschen nicht in der Lage sind, aus der Geschichte zu lernen“ (ARTED-YT[20221229]).

Der folgende Kommentar kommuniziert das Stereotyp der JÜDISCHEN SCHULD AM ANTISEMITISMUS, da er einen Zusammenhang zwischen den Wahlergebnissen und dem Hass auf Jüdinnen*Juden herstellt: „Die Zentrale der Antisemitismusmacher ist Netanjahus Büro in Jerusalem“ (SPIEG[20221113]). Hier ist interessant, dass die Argumentation auf einem (von dem*der User*in missverstandenen) Beitrag der Medien beruht. Am 17. November bewarb der *Spiegel* einen Artikel auf *Twitter* mit einem Posting, in dem er erklärte: „Wenn in Israel eine rechtsradikale Regierung an die Macht kommt, droht eine neue Welle des Antisemitismus – gegen Juden in Europa und Deutschland“ (SPIEG-TW[20221113]). In dem Artikel selbst wird behauptet, dass die ideologische Verortung der neuen Regierung als Alibi für das Aufkommen von Judenfeindschaft benutzt werde. Indem der Tweet also eine Verbindung zwischen der israelischen Politik und Antisemitismus in Europa herstellte (ersteres also als Rechtfertigung für letzteres missbraucht wird), bildete er die Grundlage für Interpretationen von Seiten der Nutzer*innen, bei denen Antisemitismus als Folge israelischen – und im weiteren Sinne – jüdischen Verhaltens dargestellt wird.

Die bereits erwähnte Vorstellung des KRITIKTABUS wird sowohl über explizite als auch implizite Muster kommuniziert. Der Vorwurf der falsch verstandenen Kritik ist typisch für dieses Stereotyp: „Weil jede Israel-Kritik gleich als Antisemitismus verstanden wird“ (ZEIT-IG[20221102]). Auch Kommentare, die von imaginierten Ängsten, dem eigenen Opferstatus bzw. dem Wunsch nach „subversiver Kritik“ Zeugnis ablegen, greifen auf die Vorstellung einer Tabuisierung von Kritik zurück und skizzieren – wie im folgenden Beispiel – im Falle einer Kritik an Israel ein drohendes Gefahrenszenario:

> „Keine Ahnung aber wir wollen ja jetzt auch nichts falsches sagen und dann gleich vom Terrorkomando abgeholt werden. Man trifft bei diesem Thema ja dauernd wunde Punkte“ (SPIEG-TW[20221113]).

Der vage konspirative Ton des Kommentars (der dadurch entsteht, dass die Gefahrenquelle nicht klar benannt wird) wird verwendet, um die Möglichkeit einer solchen Gefahr suggestiv noch hervorzuheben. Da gleichzeitig keine direkte Erwähnung von Jüdinnen*Juden oder Israel erfolgt, ist die Bedeutung des Kommentars kontextabhängig. Ähnlichen Mustern folgend können solche Äußerungen erscheinen, die andere antisemitische Konzepte unterstützen, wie bspw. BEZÜGE ZUM FASCHISMUS in Verbindung mit Schuldprojektionen – im vorliegenden Falle Netanjahu: „antsemitismus,,,lassen sie mich raten ist,,,, wenn man einen Faschisten der zufällig jüdischen Glaubens ist,, einen Faschisten nennt!“ (ZEIT[20221101]).

Den jüdischen Staat offen mit NS-Deutschland zu vergleichen oder gar gleichzusetzen – einer Diktatur, in der Jüdinnen*Juden systematisch verfolgt wurden – stellt eine Form der OPFER-TÄTER-UMKEHR sowie eine Form der eklatanten (und entlastenden) Verharmlosung der NS-Verbrechen dar. Um eine Sanktionierung für die Äußerung solcher antisemitischen Ideen zu vermeiden, verschleiern Kommentator*innen ihre Aussagen und verweisen hinsichtlich unterstellter Ähnlichkeiten zwischen historischen nationalsozialistischen (teils auch faschistischen) Szenarien und dem heutigen Israel ausschließlich auf bestimmte Aspekte (im Sinne eines Paralogismus). Die NS-ANALOGIE wird oft in längeren Beschreibungen des deutschen historischen Kontextes versteckt, die wiederum verwendet werden, um indirekt den Vergleich zwischen beiden Szenarien zu rechtfertigen:

> „Das deutsche Volk hatte Hitler auch gewählt ,immerhin nur mit ca 42% und das auch regional sehr unterschiedlich .Das israelische Staatsvolk ist sehr heterogen ,wählt aber seit Jahren mit immer nationalistischer, rassistischer Tendenz“ (WELT[20221102]).

Einige Kommentator*innen aktivierten die Analogie mittels rhetorischer Fragen und offener Anspielungen auf noch indirektere Weise: „Israel sollte sich die Geschichte Deutschlands nochmals vor Augen führen und sich fragen, ob dies der richtige Weg ist?“ (ZEIT[20221216]) (zu Anspielungen und indirekten Sprechakten im Kontext der NS-ANALOGIE s. Becker 2018).

Obwohl manchmal mit Stereotypen wie jenem des KRITIKTABUS kombiniert, kommunizieren User*innen die APARTHEID-ANALOGIE in der Regel sehr deutlich: „einfach mal richtige Kritik üben, wenn Menschen andere Menschen unterdrücken, Israel ist eine Apartheid“ (SPIEG-TW[20221113]). Das Objekt des Hasses in der APARTHEID-ANALOGIE ist oft Israel, aber sie kann sich auch gegen andere jüdische Institutionen bzw. Personen richten. Ebenso findet sie im Rahmen von Handlungsaufforderungen Verwendung, in denen eine Politik der Segregation in Israel präsupponiert wird – wie in diesem auf Englisch verfassten Kommentar: „end apartheid & zionism!“ (ARTED-YT[20221229]). Im Einklang mit dieser Vorstellung wird Israel als vom Rassismus angetrieben perspektiviert: „Hauptsache Rassismus und Apartheid funktionieren. Dann ist ja alles gut“ (ZEIT[20221101]). Die mit Apartheid-Bezügen versehenen Äußerungen nehmen ebenso Bezug auf die Palästinenser*innen und fordern Unterstützung: „Viel Glück und Kraft den Palästinensern beim Widerstand gegen diesen terroristischen Apartheidstaat 🙏“ (ZDF-YT[20221229]).

Unsere Analyse zeigt, dass die oftmals kritische Berichterstattung in den deutschen Medien über die israelische Politik – und insbesondere über Wahlen – in den Threads dämonisierende Zuschreibungen und die Verbreitung antisemitischer Vorstellungen in Bezug auf Israel ausgelöst haben. Kommentatoren berufen sich auf demokratischen Gesellschaften zugeschriebene Werte, während sie gleichzeitig ihr Arsenal an antisemitischen Konzepten – wie institutionalisierten Rassismus, Apartheid, Faschismus oder Nationalsozialismus – auf den israelischen Staat projizieren. Diese Zuschreibungen werden verwendet, um Israel als undemokratische Gesellschaft zu delegitimieren und international auszugrenzen. Es ist wahrscheinlich, dass in Fällen wie dem des *Spiegel*-Tweets die antisemitischen Beiträge durch die Zweideutigkeit der Medienberichterstattung verstärkt wurden.

5. Potenziale und Grenzen der automatischen Erkennung von Antisemitismus online

Im dritten Quartal 2022 meldete *Meta*, gegen 10,6 Millionen Inhalte vorgegangen zu sein, die als Hassrede auf *Facebook* eingestuft worden waren. Von diesen Beiträgen wurden über 90 % durch das Unternehmen selbst identifiziert und bearbeitet, bevor Nutzer*innen sie meldeten (Meta 2022). Angesichts der schieren Menge an Inhalten, die über die sozialen Medien verbreitet werden, ist die automatische Erkennung von Hassrede und anderen beleidigenden Inhalten zu einer Kernfrage für die großen Social Media-Plattformen geworden. Vor ähnlichen Herausforderungen stehen empirisch basierte Forschungsprojekte, das Monitoring durch NGOs sowie die journalistische Analyse politischer Diskurse.

Die technische Grundlage zur Lösung dieser Aufgabe ist die Textklassifizierung, d. h. der Prozess der automatischen Zuordnung von Kategorien oder Klassen zu einem Text. Im Kontext politischer Online-Kommunikation können solche Kategorien bspw. verschiedene Formen von Hassrede, Abwertung und Ausgrenzung sein, wie etwa Frauenfeindlichkeit, Rassismus und Antisemitismus. Die Klassifizierung von Texten ist eine Kernaufgabe des *Natural Language Processing* (NLP), der computergestützten Verarbeitung natürlicher Sprache. In der Vergangenheit wurden spezifisch definierte Regeln, die auf bestimmte Textaspekte abzielen, zur Durchführung von Aufgaben wie der Textklassifikation verwendet. Moderne Ansätze nutzen jedoch das maschinelle Lernen, um verbesserte Ergebnisse zu erzielen. Dabei werden große Datensätze in Algorithmen eingespeist, die Muster in den Texten erkennen sollen, um die Klassen für neue Daten möglichst genau vorherzusagen. Zu den gängigen Anwendungen der Klassifikation, von denen wir einige täglich nutzen, gehören die Sentiment-Analyse, die automatische Spracherkennung und Übersetzung oder die Erfassung von Spam.

Eine der größten Herausforderungen bei der Textklassifikation ist die adäquate Operationalisierung der Aufgabe. Dazu gehört die Festlegung der zu verwendenden Klassen, die Entscheidung darüber, was einen Text ausmacht, wie er vorverarbeitet werden soll und auf welcher Ebene die Klassifizierung durchgeführt werden soll (bspw. auf Dokument-, Absatz- oder Satzebene).

Die Klassifizierung von Texten erfolgt in der Regel in einem überwachten Verfahren (*Supervised Machine Learning*), bei dem ein Algorithmus mit manuell annotierten Daten trainiert wird, um genaue Vorhersagen zu treffen. Die menschlichen Annotationen dienen als ‚Goldstandard' und werden nicht nur dazu verwendet, dem Algorithmus etwas beizubringen, sondern auch die Vorhersagen des erlernten Modells anhand von Standardmetriken zu bewerten. Häufig werden sogenannte *Benchmark*-Datensätze verwendet, um die Leistung verschiedener Lernmodelle, die für eine bestimmte Aufgabe trainiert worden sind, anhand eines gemeinsamen Datensatzes und unter Verwendung aufgabenspezifischer Metriken zu vergleichen. Bemühungen, Benchmark-Datensätze für die automatische Erkennung von Antisemitismus zu erstellen, wurden bisher nur von einer Handvoll Forscher*innen unternommen (Jikeli/Cavar/Miehling 2019, Chandra et al. 2021, Jikeli et al. 2021 und 2022, Steffen et al. 2022). Die daraus entstandenen Datensätze sind noch nicht mit bestehenden Korpora für verwandte Phänomene wie beleidigende und toxische Sprache sowie andere Formen von Hassrede vergleichbar.

Herausforderungen bei der Operationalisierung von Texten

Damit Computer menschliche Sprache verarbeiten können, müssen numerische Repräsentationen der Daten erzeugt werden. Dieser Kodierungsprozess ist anspruchsvoll, denn das Ziel ist es, möglichst viele Informationen aus den Textdaten beizubehalten. Ein einfacher Ansatz, das sogenannte *Bag of words*-Model, besteht darin, einen Text über die jeweilige Anzahl der darin vorkommenden Wörter darzustellen. Das Training von Klassifikationsmodellen auf der Grundlage dieser Textrepräsentation kann zufriedenstellende Ergebnisse liefern, solange die Aufgabe kein tiefergreifenderes Verständnis von Narrativ und Kontext erfordert. In komplexeren Fällen wird es daher eher eingesetzt, um Basismodelle zu erstellen, die als Referenzpunkte für zukünftige Verbesserungen dienen können.

Um die Performanz von Klassifikationsmodellen zu steigern, sind Textrepräsentationen erforderlich, welche die semantische Dimension der menschlichen Sprache, wie z. B. die Ähnlichkeit von Wörtern und Konzepten, und damit Kontextinformationen abbilden können. Die Bedeutung von durch Menschen produzierten Texten zu erfassen ist eine schwierige Aufgabe, insbesondere bei kurzen Nachrichten, wie sie häufig in der Online-Kommunikation und in sozialen Medien zu finden sind. Nutzer*innen können ihre Meinung auf subtile, verschlüsselte, implizite Weise zum Ausdruck bringen, bspw. um ein gewisses Maß an Ambivalenz zu erreichen und so Maßnahmen der Plattform-Moderation zu umgehen. Beispiele dafür finden sich in fragmentierten Glaubensbekundungen im Kontext von Verschwörungstheorien (Steffen et al. 2022), der impliziten Leugnung des Klimawandels (Falkenberg/Baronchelli 2023) oder der Verwendung von Codes in antisemitischen Narrativen. Bezüge zu Weltwissen erschweren es zusätzlich, den Inhalt eines Textes zu ‚verstehen'. Ein extremes Beispiel hierfür ist eine kürzlich getätigte Aussage von Nicholas J. Fuentes, einem politischen Kommentator und Live-Streamer aus der *White Supremacy*-Szene, der den Holocaust leugnete, indem er ‚scherzhaft' die Möglichkeit bezweifelte, dass innerhalb von fünf Jahren sechs Millionen Kekse gebacken werden könnten.[6]

Ein Problem, das eng mit der Erfassung von Informationen aus Texten zusammenhängt, ist die Menge der Daten. Die Annotation von Daten ist in der Regel zeit- und kostenaufwändig und erfordert oft – wie im Projekt Decoding Antisemitism – die Arbeit von Expert*innen. Dies stellt eine große Herausforderung dar, da Algorithmen aus relativ kleinen Datenmengen vielfältige Interaktionen zwischen Wörtern erlernen sollen. In der Praxis führt dies zu wenig zufriedenstellenden Ergebnissen, vor allem wenn die resultierenden Modelle in Szenarien angewendet werden, die sich (leicht) von der Trainingssituation unterscheiden. In unserem Kontext würde das bedeuten, dass ein Modell, das auf das im Projekt annotierte Korpus trainiert wurde, eine (deutlich) geringere Performanz zeigt, wenn es mit Beispielen antisemitischer Äußerungen in einem neuen Diskurs konfrontiert wird.

6 – In einem seiner Livestreams liest Fuentes den folgenden Text vor: „Wenn ich eine Stunde brauche, um eine Ladung Kekse zu backen, und das Keksmonster 15 Öfen hat, die 24 Stunden am Tag arbeiten, jeden Tag, fünf Jahre lang, wie lange braucht das Keksmonster, um 6 Millionen Kekse zu backen?", und verwendet dann die Keks-Analogie für mehrere, den Holocaust leugnende Aussagen. Für die entsprechende Livestream-Szene siehe https://mobile.twitter.com/CalebJHull/status/1189594371030695937 (letzter Zugriff am 23. Februar 2023). Für weitere Informationen siehe bspw. https://www.adl.org/resources/blog/nicholas-j-fuentes-five-things-know (letzter Zugriff am 14. Februar 2023). ChatGPT konnte den antisemitischen Charakter dieser Aussage übrigens *nicht* erkennen.

Transfer-Lernen mit Transformer-Architekturen

In den letzten Jahren haben zwei wichtige Entwicklungen dazu beigetragen, die Herausforderungen des überwachten Lernparadigmas und der kontextarmen Sprachrepräsentationen zu bewältigen: (1) Transfer-Lernen, ein Paradigma, bei dem das bei der Lösung einer Aufgabe erworbene Wissen auf eine andere, möglicherweise schwieriger zu lösende Aufgabe übertragen wird, und (2) Transformer-Architekturen, die in der Lage sind, komplexe Kontextinformationen in Texten zu erfassen.

Beim **Transfer-Lernen** wird ein Modell trainiert, um eine ‚Ausgangsaufgabe' zu lösen, und dann für eine ausreichend ähnliche ‚Zielaufgabe' angepasst. Dies ist besonders nützlich, wenn für die Zielaufgabe nur wenige gelabelte Daten vorliegen (wie oft bei der Klassifizierung von Texten im politischen Bereich), für die Ausgangsaufgabe aber eine große Menge an Trainingsdaten gegeben ist. Im NLP-Bereich wird dies insbesondere durch die Verwendung großer digitaler Korpora wie *Wikipedia* oder *Google News* für die eher generelle primäre Aufgabe der Vorhersage eines nächsten oder fehlenden Wortes aus dem vorangehenden bzw. umgebenden Kontext veranschaulicht. Solche Sprachmodelle,[7] die ohne menschliche Annotation trainiert wurden,[8] sind sogar in der Lage, eine Vielzahl von linguistischen Phänomenen wie Semantiken auf Wort- und Satzebene, syntaktische Strukturen und Phänomene auf Diskursebene, ebenso wie Feinheiten der menschlichen Sprache wie etwa Sarkasmus oder Slang zu erfassen.

Sobald ein Sprachmodell trainiert wurde, kann es für verschiedene Anwendungsfälle optimiert werden, z. B. für die Klassifizierung von Texten in ‚Hassrede' und ‚keine Hassrede'. Im Wesentlichen nutzt das Klassifikationsmodell das allgemeine domänenunabhängige Wissen, das vom Ausgangsmodell erlernt wurde, so dass es nur noch die Besonderheiten der Zielkategorien/-klassen lernen muss. Technisch gesehen lässt es sich so veranschaulichen, dass das Ausgangsmodell um einen vergleichsweise kleinen Satz anwendungsspezifischer Parameter erweitert wird, die aus den Zieldaten gelernt werden müssen.

Neuerdings werden sogenannte **Transformer-Architekturen** eingesetzt, um Sprachmodelle für die Ausgangsaufgabe zu trainieren. Transformer stellen eine paradigmatische Veränderung gegenüber früheren Modellarchitekturen dar. Sie verwenden maßgeblich einen Mechanismus namens (*self-*)*attention*, der es ermöglicht, jedes Wort in Bezug auf den Kontext, beispielsweise den Satz oder Paragraphen, in dem es vorkommt, darzustellen (vgl. Vaswani et al. 2017). Dieser Mechanismus ermöglicht es ihnen zu lernen, wie Wörter miteinander in Beziehung stehen, auch wenn sich diese nicht in unmittelbarer Nähe zueinander im Text befinden.

7 – Ein Sprachmodell berechnet, wie wahrscheinlich es ist, dass eine bestimmte Folge von Wörtern in einer bestimmten Sprache wie Deutsch oder Englisch vokommt. Es kann verwendet werden, um das nächste Wort in einer Folge vorherzusagen und so Text zu erzeugen.

8 – Wenn wir ein Wort aus einem Satz entfernen, dient der Rest des Satzes als Kontext für die Vorhersage des fehlenden Wortes, wodurch ursprünglich ungelabelte Daten (z. B. Sätze aus *Wikipedia*) in gelabelte Daten umgewandelt werden. Diese Art des Lernens wird als selbstüberwachtes Lernen bezeichnet.

Es gibt eine Vielzahl an vortrainierten Sprachmodellen, die für verschiedene Folgeaufgaben, einschließlich der Textklassifikation, angepasst werden können (sogenanntes *fine-tuning*). Die Sprachmodelle unterscheiden sich in Aspekten wie der zum Training verwendeten Korpora (bspw. *Wikipedia* vs. *Twitter*), der Sprache, der Architektur (bspw. Art und Anzahl der Schichten [9]) oder der Vorverarbeitung des Textes (bspw. Kleinschreibung aller Wörter).

Eine der am häufigsten verwendeten Architekturen ist BERT, die aktuell für eine Reihe von NLP-Anwendungen die besten Ergebnisse erzielt. BERT-ähnliche vortrainierte Sprachmodelle werden in der aktuellen Forschung üblicherweise verwendet, um Klassifikatoren für verschiedene Aufgaben zu trainieren, darunter Hassrede (Basile et al. 2019, Aluru et al. 2020, Mathew et al. 2022), beleidigende Sprache (Wiegand/Siegel/Ruppenhofer 2018, Zampieri et al. 2019 und 2020, Mandl et al. 2021) oder (vordefinierte) Verschwörungserzählungen (Pogorelov et al. 2020, Moffitt/King/Carley 2021, Elroy/Yosipof 2022, Phillips/Ng/Carley 2022). Die Mehrheit dieser Benchmark-Datensätze ist in englischer Sprache und konzentriert sich stark auf *Twitter* als Datenquelle (vgl. Poletto et al. 2021), was auf die Popularität der Plattform, aber auch auf den einfachen technischen Zugang zu den Daten zurückzuführen ist. Antisemitismus wurde bisher in nur relativ wenigen Bemühungen zur Textklassifikation berücksichtigt.

Dienste für die Moderation von Inhalten

Das Fehlen großer annotierter Korpora führt zu einem Mangel an Diensten zur automatisierten Erkennung von antisemitischen Inhalten. Fortschritte wurden jedoch bei der Entwicklung von Webservices für die Erkennung anderer sprachlicher Phänomene erzielt, die mit Antisemitismus in Verbindung stehen, wie bspw. Hassrede und toxische Sprache. Ein prominentes Beispiel ist die *Perspective API*, ein kostenloser Dienst, der von *Jigsaw* und dem Counter Abuse Technology-Team von *Google* entwickelt wurde und weithin für die Moderation von Inhalten sowie für die Forschung eingesetzt wird – bspw. für Analysen von Moderationsmaßnahmen auf *Reddit* (Horta Ribeiro et al. 2021), für Untersuchungen politischer Online-Communities auf *Reddit* (Rajadesingan/Resnick/Budak 2020) und *Telegram* (Hoseini et al. 2021) sowie zur Identifizierung antisemitischer und islamfeindlicher Texte auf *4chan* (González-Pizarro/Zannettou 2022).

Der Dienst ermöglicht die Erkennung von missbräuchlichen Inhalten, indem er numerische Werte (zwischen 0 und 1) für verschiedene Attribute wie ‚toxicity' (Toxizität), ‚insult' (Beleidigung) oder ‚threat' (Bedrohung) vergibt. Diese Werte werden von Machine Learning-Modellen berechnet [10] die mithilfe von *crowd-labelled* Daten trainiert wurden, also Daten, die durch eine große Anzahl von Personen über Crowdsourcing-Plattformen annotiert werden. Die zugrundeliegende Strategie besteht darin, große Mengen von (divers) annotierten Daten zu erstellen. Dafür werden einfache Definitionen verwendet, die auch von Nicht-Expert*innen

9 – Transformer und neuronale Netze im Allgemeinen sind in Schichten organisiert, die aus Recheneinheiten bestehen, die typischerweise Neuronen genannt werden und die Eingänge und Ausgänge des Modells verbinden. Es gibt verschiedene Arten von Schichten, die spezifische Berechnungsanforderungen erfüllen, und die Architektur eines neuronalen Netzes kann aus einer unterschiedlichen Anzahl von Schichten bestehen. So bestehen beispielsweise die weit verbreiteten BERT-Sprachmodelle aus so genannten Transformatorblöcken, die in weitere Schichten wie self-attention und normalize-Schichten zerlegt werden können. Die Basisvariante von BERT besteht aus 12 Transformer-Blöcken, während BERTlarge doppelt so viele hat. Eine Architektur mit mehr Schichten ist im Allgemeinen komplexer (und ‚tiefer') und besteht aus mehr Parametern, die während des Trainings gelernt werden müssen, so dass in der Regel mehr Daten für das Training benötigt werden.

10 – Weitere Einzelheiten über die auf Transformern basierende Architektur sind bspw. unter https://arxiv.org/pdf/2202.11176.pdf zu finden.

verstanden und angewendet werden können. So sollen beispielsweise Inhalte als toxisch eingestuft werden, wenn sie als „unhöflich, respektlos oder unangemessen [...] gelten, was dazu führen kann, dass Menschen eine Diskussion verlassen" (Google 2022, Thain/Dixon/Wulczyn 2017Übersetzung der Autor*innen). Um der Subjektivität und Ungenauigkeit der Definition entgegenzuwirken, werden die Texte von mehreren Personen gelabelt und ihre Bewertungen vor dem Training der Modelle zusammengefasst.

Theoretisch könnte die *Perspective API* einen leicht zugänglichen Ansatz zur Erkennung bestimmter Formen antisemitischer Äußerungen bieten. Aktuelle Arbeiten zur deutschsprachigen Kommunikation auf *Telegram* und *Twitter* weisen jedoch auf Einschränkungen bei der Verwendung des Dienstes für diese Aufgabe hin – nämlich eine Übersensibilität gegenüber bestimmten identitätsbezogenen Schlüsselwörtern wie „Jude" oder „Israel", wodurch der Dienst dazu neigt, Texte fälschlicherweise als antisemitisch einzustufen, nur weil sie Jüdischsein ansprechen bzw. Israel erwähnen (Mihaljević/Steffen 2022). Darüber hinaus hat sich gezeigt, dass der Dienst subtilere oder verschlüsselte Formen des Antisemitismus nur unzureichend erkennt und sie oft nicht als toxisch einstuft (ebd.).

Um ein umfassenderes Bild zu erhalten, haben wir die *Perspective API* auf einen Teil der mehrsprachigen Daten des Projekts „Decoding Antisemitism" angewendet, die aus etwa 3.500 manuell als antisemitisch[11] gelabelten und etwa 53.500 als nicht antisemitisch gelabelten Kommentaren bestehen und insgesamt 57.021 Datensätze ergeben. Wir haben die Bewertungen für die Attribute ‚identity attack'[12] (Angriff auf die Identität), ‚toxicity' (Toxizität) und ‚severe toxicity' (schwere Toxizität) ausgewertet. Wir betrachteten dabei insbesondere, wie viele Kommentare, die von den menschlichen Annotator*innen als antisemitisch eingestuft wurden, auch von dem Dienst mit einem Wert über 0,5 bewertet wurden, und untersuchen, ob bestimmte Schlüsselwörter die Performance beeinflussen.

Die Verteilungen der Werte für die drei Attribute unterscheiden sich signifikant zwischen antisemitischen und nicht-antisemitischen Kommentaren, wie in Abbildung 4 dargestellt ist, wobei die Werte für antisemitische Kommentare höher sind. Allerdings wurden 75 % der antisemitischen Kommentare in Bezug auf ‚toxicity' oder ‚severe toxicity' mit weniger als 0,5 bewertet, was ein typischer Schwellenwert für die Zuordnung von Kommentaren zu einer der beiden Gruppen ist. **Das bedeutet, dass ein großer Teil der antisemitischen Inhalte, der Bewertung durch *Perspective API* folgend, nicht als toxisch eingestuft werden würde.** Wenn man bedenkt, dass einige Forschungsstudien einen Schwellenwert von 0,8 wählen, würde dies eine noch größere Anzahl von sogenannten *false negatives* bedeuten. Die Werte für die Gruppe der antisemitischen Kommentare sind für das Attribut ‚identity attack' am höchsten. Aber selbst hier liegen 75 % der antisemitischen Kommentare unter 0,8 und wären somit von den oben genannten Forschungsdesigns übersehen worden.

11 – Texte, die als kontextueller Antisemitismus gelabelt sind, wurden aus dem Datensatz ausgeschlossen, da der Dienst nur für den Text selbst Punkte vorhersagt und nicht in der Lage ist, außerhalb des Textes verortete Informationen zu berücksichtigen.

12 – Negative oder hasserfüllte Kommentare, die sich gegen jemanden aufgrund seiner Identität richten: https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages?language=en_US (s. Google 2022).

Abbildung 4: Verteilungen der Werte für ‚identity attack', ‚toxicity' und ‚severe toxicity' getrennt nach antisemitisch und nicht-antisemitisch eingestuften Kommentaren. Die horizontalen Linien der Kästen markieren das untere Viertel (25 %), den Median (50 %) und das obere Viertel (75 %) der untersuchten Werte.

Die höheren Werte für ‚identity attack' sind nicht überraschend, da Antisemitismus eine identitätsbezogene Form des Hasses ist, die Vorurteile und Diskriminierung gegen jüdische Menschen aufgrund ihrer wahrgenommenen Identität als Gruppe beinhaltet. Allerdings könnten die hohen Werte für dieses Attribut auch darauf hindeuten, dass der Dienst übermäßig sensibel auf bestimmte identitätsbezogene Schlüsselwörter wie „Jude/jüdisch" oder „Israel" reagiert.

Diese falsch-positive Verzerrung, d. h. die Tendenz des Systems, den Grad der Toxizität zu überschätzen, wenn ‚Minderheiten' erwähnt werden, unabhängig von ihnen gegenüber geäußerter Haltung, wurde bereits von den Entwicklern der API thematisiert (Dixon et al. 2018) und durch andere Forschungsarbeiten bestätigt (Hutchinson et al. 2020, Röttger et al. 2021).

Um die mögliche Auswirkung von identitätsbezogenen Schlüsselwörtern auf die Bewertungen von ‚identity attack' zu untersuchen, haben wir alle Texte markiert, die Variationen der Schlüsselwörter „Jude/Jüdisch" und „Israel" enthalten. Abbildung 5 veranschaulicht die Werteverteilung, wenn wir diese zusätzlichen Informationen berücksichtigen: Kommentare, die identitätsbezogene Schlüsselwörter (s. Punkte in orange) enthalten, weisen tendenziell höhere Werte für ‚identity attack' auf. Das Ergebnis deutet darauf hin, dass Texte mit Verweisen auf Jüdinnen*Juden, Jüdischsein oder Israel, auch wenn sie keinen Antisemitismus beinhalten, wahrscheinlich als ‚identity attack' gewertet werden. Obwohl das alleinige Vorkommen der entsprechenden Schlüsselwörter nicht für einen hohen ‚identity attack' Wert[13] verantwortlich ist (siehe bspw. die erste Spalte in Tabelle 1), zeigt sich dennoch eine hohe positive Korrelation. Konkret ist der Medianwert für ‚identity attack' bei Kommentaren, die als nicht antisemitisch gekennzeichnet sind, um 0,43 höher, wenn der Text eines der identitätsbezogenen Schlüsselwörter enthält. Bei antisemitischen Texten ist der Unterschied mit 0,15 weniger stark ausgeprägt. Ähnliche Effekte können auch für die beiden anderen *Perspective API*-Attribute beobachtet werden.

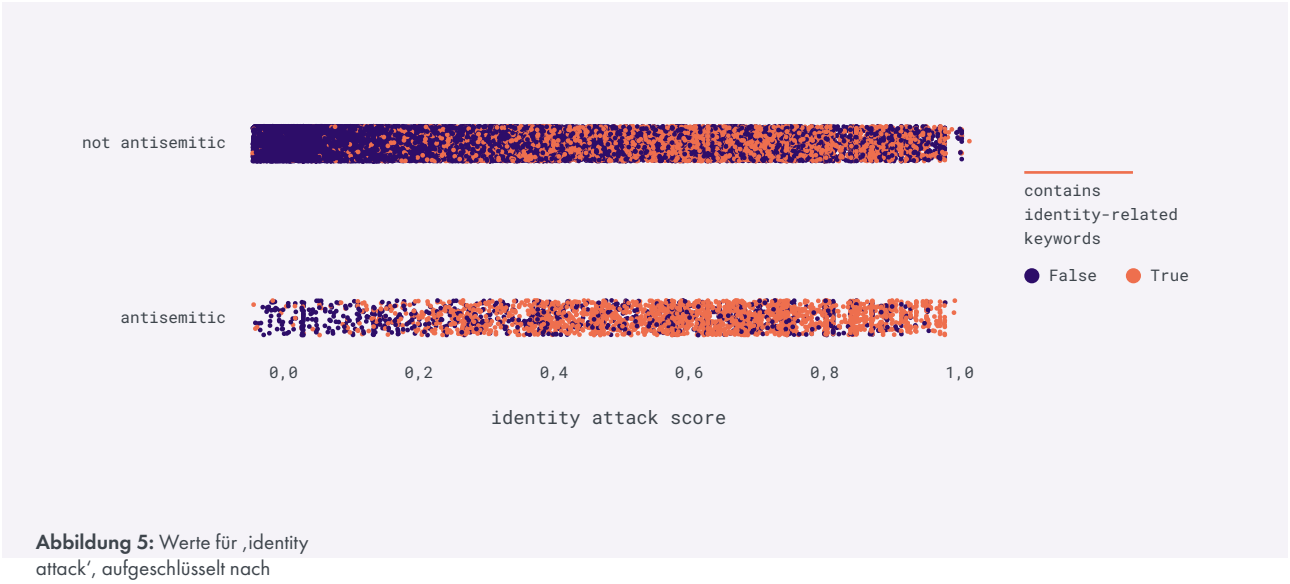

Abbildung 5: Werte für ‚identity attack', aufgeschlüsselt nach Antisemitismus-Kennzeichnung und Vorhandensein von identitätsbezogenen Schlüsselwörtern.

13 – Dies ist auch nicht zu erwarten, da die *Perspective API*-Modelle weit mehr als nur die Häufigkeit bestimmter Wörter verwenden.

	Medianwerte für ‚identity attack'		
	Texte ohne identitätsbezogene Schlüsselwörter	Texte mit identitäts-bezogenen Schlüsselwörtern	Unterschied
Als nicht antisemitisch gelabelte Texte	0,152659 (N=45761)	0,585239 (N=7769)	**+0,43258**
Als antisemitisch gelabelte Texte	0,492150 (N=969)	0,642324 (N=2522)	**+0,150174**
Unterschied	**+0,339491**	**+0,057085**	

Tabelle 1: Medianwerte für ‚identity attack', aufgeschlüsselt nach Antisemitismus-Kennzeichnung und Vorhandensein von identitätsbezogenen Schlüsselwörtern. Die Gruppengrößen sind in Klammern angegeben.

Die Analyse belegt keinen kausalen Zusammenhang zwischen dem Vorkommen von Schlüsselwörtern, die sich entweder auf das Jüdischsein oder den Staat Israel beziehen, und höheren Werten für ‚identity attack'. Es wäre zum Beispiel denkbar, dass Texte, die Israel oder Jüdischsein thematisieren, toxisch oder beleidigend sein können, ohne Antisemitismus aufweisen zu müssen. Möglicherweise handelt es sich also nicht um eine stichwortbezogene falsch-positive Verzerrung, sondern um andere Textaspekte, die einen hohen Wert hervorrufen. Frühere Studien haben jedoch gezeigt, dass die Zugabe dieser Schlüsselwörter die Werte der Texte signifikant erhöht (Mihaljević/Steffen 2022), was eine stichwortbezogene Verzerrung wahrscheinlich macht. Die Ergebnisse müssen noch eingehender erforscht werden. Dazu werden wir in unserer zukünftigen Forschung die API-Bewertungen auf der Ebene der einzelnen Textabschnitte untersuchen, um festzustellen, welche Segmente eines Textes höhere Werte auslösen.

Bisherige Experimente zeigen, dass die *Perspective API* für die Moderation von Inhalten im Hinblick auf antisemitische Äußerungen und für die Erforschung des Online-Antisemitismus eher begrenzt ist. Die signifikante positive Korrelation von identitätsbezogenen Schlüsselwörtern mit höheren Werten deutet auf ein gesteigertes Risiko hin, einen Text fälschlicherweise als antisemitisch einzustufen, während das Vorkommen antisemitischer Codes die Identifizierung solcher Inhalte als toxisch antisemitisch erheblich erschweren kann. Letzteres bietet Nutzer*innen die Möglichkeit, sprachliche Codes, Emojis oder Ironie und Sarkasmus strategisch einzusetzen, um stichwortbasierte automatische Erkennungsmethoden zu umgehen. Vermutlich ist der Labeling-Ansatz von *Perspective API* grundsätzlich nicht für die Erkennung antisemitischer Inhalte geeignet, da es selbst für Expert*innen schwierig ist, kurze Texte, wie sie für die Online- und Social Media-Kommunikation typisch sind, dahingehend zu annotieren. Daher bleibt die automatisierte Erkennung antisemitischer Äußerungen weiterhin eine Herausforderung und erfordert eine sorgfältige Modellierung auf der Grundlage von qualitativ hochwertig gelabelten Daten.

Erste Experimente mit Transfer-Lernen und Transformer-Architekturen

Vorausgehende Arbeiten im Decoding Antisemitism-Projekt untersuchten das Training einer logistischen Regression (ein klassischer Ansatz aus der Statistik) auf der Grundlage von *Bag of words*-Textrepräsentationen, um antisemitische von nicht-antisemitischen Texten in Englischer Sprache zu unterscheiden (Ascone et al. 2022). Ein großer Vorteil dieser einfachen Modelle ist ihre höhere Modelltransparenz, mit der man leicht herausfinden kann, welche Wörter vorwiegend zur Entscheidung für die jeweilige Klasse beitragen. Zur Klassifikation als antisemitisch hat das Wort ‚apartheid' (Gewicht=13,72) am meisten beigetragen, gefolgt von ‚genocide' (Gewicht=9,24). Unter den 20 häufigsten Wörtern für die Vorhersage, dass ein Text antisemitisch ist, befinden sich auch die Codes ‚israhell'[14] und ‚satanyahu' sowie Lexeme, die im Zusammenhang mit Gewalt stehen (wie etwa ‚murder'), das Wort ‚lobby', das auf Verschwörungserzählungen hinweist, oder das Wort ‚devil', ein dämonisierendes rhetorisches Element. Während diese Wörter für die vorhandenen Korpora plausibel sind, weisen sie auch auf potenzielle Probleme in Bezug auf die Anwendung des Modells hin: So würde beispielsweise ein Satz, der ‚apartheid' definiert, mit sehr hoher Wahrscheinlichkeit als antisemitisch vorhergesagt werden. Außerdem ist eine Verzerrung für das Wort ‚Israel' zu erkennen, durch welche Sätze, die dieses Wort enthalten, mit höherer Wahrscheinlichkeit als antisemitisch eingestuft werden.[15] Dennoch hat das beschriebene Modell, das auf einem Teil des aktuellen englischsprachigen Korpus trainiert wurde, erste vielversprechende Ergebnisse[16] erzielt, die als Ausgangspunkt für weitere Entwicklungen dienen können.

Unser Ansatz besteht demgegenüber aus einem Fine-Tuning von Transformer-basierten Sprachmodellen für eine Klassifikationssaufgabe, wie zuvor beschrieben. Zudem haben wir einige Anpassungen bei der Zuordnung von Kommentaren zu Klassen vorgenommen: Im Projekt Decoding Antisemitism wird unterschieden zwischen Kommentaren, deren antisemitischer Charakter ohne weitere Informationen erkannt werden kann, und solchen, die ‚kontextuell antisemitisch' sind, d. h. es ist zusätzlicher Kontext wie der Inhalt hinter einer verlinkten URL, Informationen aus früheren Kommentaren oder das Weltwissen der Leser*innen erforderlich, um antisemitische Inhalte zu erkennen. Der Kommentar „Ich glaube, dir wurde gesagt, dass du das tun sollst" kann beispielsweise nicht vollständig interpretiert werden, ohne die Mehrdeutigkeit dessen zu klären, worauf sich „das" und „du" beziehen. Ein maschinelles Lernmodell benötigt diese Informationen ebenfalls, um korrekte Schlüsse zu ziehen, aber die Bereitstellung dieser Informationen stellt in einem praktischen Anwendungsszenario eine Herausforderung dar. Während menschliche Kommentator*innen (oder Moderator*innen) solche Mehrdeutigkeiten in der Regel vollständig auflösen können – nämlich, dass einer/einem Nutzer*in unterstellt wird, aufgrund eines imaginierten jüdischen Einflusses auf eine bestimmte Weise zu handeln –, stellt dies beim Versuch der automatisierten Erkennung eine große Herausforderung dar. Daher nutzen wir nur Texte, die auch ohne zusätzliche Kontextinformationen als antisemitisch erkennbar sind, um potenzielle Mehrdeutigkeiten zu vermeiden.

14 – Tatsächlich ist die Summe der Gewichte der beiden Varianten ‚israhell' und ‚israhel' höher als die von ‚apartheid'.

15 – Sogar völlig harmlose Sätze wie ‚Israel ist ein Land reich an Kultur und Geschichte, und seine pulsierenden Städte sind voller Leben und Energie'.

16 – Im vorherigen Bericht wurde ein F1-Ergebnis von 0,75 angegeben. Wir haben das Modell jedoch repliziert und mit verschiedenen Splits des aktuellen Korpusstandes kreuzvalidiert und dabei eine durchschnittliche F1-Punktzahl von 0,63 ermittelt. Dies deutet darauf hin, dass das Modell recht instabil ist, was angesichts der geringen Datenmenge in der positiven Klasse nicht überraschend ist.

Dies hat jedoch den Nachteil, dass unser ohnehin schon unausgewogener Datensatz, bei dem etwa 85 % der Kommentare als nicht antisemitisch (negative Klasse oder Klasse 0) gekennzeichnet sind, noch stärker verzerrt wird, da so nur noch etwa 10 % der Texte als antisemitisch (positive Klasse oder Klasse 1) gewertet werden.

Es gibt verschiedene Ansätze, um mit stark unausgewogenen Daten während des Trainings umzugehen, wie z. B. das zufallsbasierte Reduzieren der Mehrheitsklasse, die Vergrößerung der Minderheitenklasse (bspw. durch kleine Variationen bestehender Texte) oder die stärkere Bestrafung von Fehlern des Modells für Datenpunkte aus der Minderheitenklasse. Um den Einfluss zusätzlicher Aspekte zu untersuchen, berücksichtigen wir auch die Wahl des vortrainierten Sprachmodells,[17] Standard-Hyperparameter für die Optimierung von Transformermodellen (bspw. die Lernrate und *weight decay*) und datenbezogene Einstellungen (bspw. Behandlung besonders kurzer Texte und Entfernung von Emojis). Diese Hyperparameter bestimmen die Gesamtfähigkeit eines maschinellen Lernmodells, so dass Kombinationen verschiedener Werte evaluiert werden, um den optimalen Wert zu finden. Da der Raum der Hyperparameter jedoch tendenziell sehr groß ist, muss ein Gleichgewicht zwischen ‚Exploration' und ‚Exploitation' gefunden werden, um eine effiziente Abstimmung der Hyperparameter zu ermöglichen. Um dies zu erreichen, verwenden wir die Bayes'sche Optimierung, bei der ein probabilistisches Modell die Leistung verschiedener Hyperparameter-Konfigurationen vorhersagt. Auf diese Weise können wir die besten Parameter nutzen und gleichzeitig neue Optionen erforschen, um die Identifikation möglichst optimaler Parameter sicherzustellen.

Wir haben 80 % der Daten zum Training verwendet (16.539 Datensätze in Klasse 0 und 1.936 in Klasse 1), 10 % zur Validierung, bei welcher die besten Hyperparameter identifiziert werden, und 10 % zum Testen des Modells, welches auf den Validierungsdaten die geringsten Fehler erzielt hat. Die beschriebenen Experimente ergeben ein Modell mit einem F1-Score von 0,7 für die positive Klasse und 0,97 für die negative Klasse. Weitere Metriken sind in Tabelle 2 dargestellt:

	Precision	Recall	F1-Score	Number of records	Accuracy
Class 1 (AS)	0,75 (0,73)	0,65	0,7 (0,69)	225 (249)	0,94
Class 0 (nicht-AS)	0,96	0,97	0,97 (0,96)	2084 (2061)	

Tabelle 2: Bewertung des leistungsstärksten Modells anhand von Test- und Validierungsdaten, mit Werten für die Validierungsdaten in Klammern, sofern abweichend.

16 – Wir verwenden BERT-base und RoBERTa-base.

Diese Werte können wie folgt interpretiert werden: 96 % aller Kommentare, die das Modell als nicht antisemitisch vorhersagt, wurden von den menschlichen Annotator*innen tatsächlich als solche gekennzeichnet (Precision class 0), und das Modell findet 97 % der Kommentare in dieser Kategorie (Recall class 0). Andererseits wurden von den Texten, die als antisemitisch vorhergesagt wurden, 75 % als solche gekennzeichnet, während das Modell 65 % der Texte finden konnte, die von den Annotator*innen als antisemitisch gekennzeichnet wurden. Um dies zu verdeutlichen: Wenn ein*e Moderator*in dieses Modell auf 1.000 Kommentare anwendet, von denen 100 als antisemitisch angenommen werden, würde das Modell 65 der 100 antisemitischen Inhalte finden und 35 davon übersehen. Aus der Perspektive, den Kommentarbereich frei von antisemitischen Äußerungen halten zu wollen, könnte man dies als eine niedrige Rate ansehen. Die Anzahl der Fehlalarme wäre jedoch mit 22 niedrig, so dass der manuelle Aufwand relativ gering wäre. Dieses Beispiel verdeutlicht den Zielkonflikt zwischen zwei Arten von Fehlern: Während man die Auffindbarkeit von Klasse 1 erhöhen möchte, wäre es auch wünschenswert, die Anzahl der Fehlalarme niedrig zu halten.

Aus der Anwendungsperspektive muss also entschieden werden, welche Art von Fehlern (falsch-positiv vs. falsch-negativ) vorrangig behandelt werden sollten, und bspw. welcher Recall für Klasse 1 mindestens erreicht werden muss und welche Präzision dafür akzeptiert werden sollte. Um dies zu veranschaulichen, nehmen wir an, dass wir einen Recall von mindestens 0,8 erreichen wollen, während die Präzision so hoch wie möglich sein soll. Eine einfache Möglichkeit wäre, die Wahrscheinlichkeitsschwelle für die Zuordnung einer Vorhersage zu einer Klasse anzupassen. Die Klassifikatoren, die wir trainieren, sind probabilistisch, d. h. sie erzeugen für jeden Text Wahrscheinlichkeiten für die Zugehörigkeit zu jeder der beiden Klassen. Standardmäßig ist der Schwellenwert für die binäre Klassifikation auf 0,5 eingestellt, d. h. die Klasse mit der höheren Wahrscheinlichkeit gewinnt. Der Schwellenwert kann jedoch geändert werden, um den Wert einer gewünschten Metrik zu erhöhen. Wir verwenden die Validierungsmenge, um herauszufinden, welcher Schwellenwert einen Recall von mindestens 0,8 bei gleichzeitiger Maximierung der Präzision erfüllt. Mit einem sehr niedrigen Schwellenwert von 0,06 erreichen wir in den Validierungsdaten einen Recall von 0,81 und eine Präzision von 0,52. Die Werte für die Testdaten sind in Tabelle 3 aufgeführt. Ihr ist zu entnehmen, dass wir fast 80 % aller antisemitischen Texte erfassen würden, dafür allerdings fast jeder zweite Alarm ein Fehlalarm wäre.

	Precision	Recall	F1-Score	Accuracy
Class 1 (AS)	0,51	0,79	0,62	0,90
Class 0 (nicht-AS)	0,98	0,92	0,95	

Tabelle 3: Bewertung des leistungsstärksten Klassifizierungsmodells nach Verschiebung des unteren Schwellenwerts für die positive Klasse von 0,5 auf 0,06.

Wie geht es weiter?

Wie aus den vorangegangenen Ausführungen hervorgeht, gibt es bei der Erstellung von Klassifikationsmodellen noch Raum für Verbesserungen, mit dem Ziel, zuverlässige Modelle mit höheren Präzisions- und Recall-Werten zu erstellen. Unser nächster Schritt im Projekt wird eine detaillierte Bewertung der Performanz des aktuellen Modells sein. Dazu gehört eine qualitative Untersuchung von Texten, bei denen das Modell die größten Fehler macht, sowie statistische Auswertungen, wie etwa die Korrelation zwischen Fehlern und spezifischen Formulierungen, um herauszufinden, ob das Modell beispielsweise bestimmte Arten von antisemitischen Inhalten besser erkennt als andere (z. B. bestimmte Stereotype oder offen hasserfüllte Sprache). Diese Erkenntnisse werden dem gesamten Projektteam dabei helfen, bessere Modelle zu entwickeln und ggf. das Annotationsschema entsprechend anzupassen.

Da die Menge der Beispiele immer einer der wichtigsten Faktoren für das Training leistungsfähiger Modelle ist, werden wir verschiedene Strategien zur Erweiterung der Trainingsdaten erforschen – bspw. durch das Hinzuziehen von annotierten Texten in anderen Sprachen und die Verwendung mehrsprachiger Modelle oder die Anwendung automatisierter Übersetzungen.

Darüber hinaus werden wir die Domänenanpassung testen, indem wir einen Teil der Diskurse für das Training reservieren, andere zum Testen verwenden und dieses Verfahren mehrfach wiederholen.

Da ein erheblicher Teil der Nachrichten mit antisemitischem Inhalt zusätzlichen Kontext außerhalb des Textes erfordert, wollen wir außerdem an Konzepten für den (praktischen) Umgang mit solchen Daten arbeiten. Langfristig halten wir es für wichtig, über die Nutzung von Klassifikationsmodellen in der Praxis – also über den akademischen Gebrauch hinausgehend – nachzudenken, wie bspw. den konkreten Einsatz bei der Moderation von Nachrichten-Webseiten und Social Media-Plattformen.

Derzeit ist es (noch) schwierig, sich gut funktionierende Modelle für die Erkennung antisemitischer Äußerungen ohne manuelle Annotation der Daten vorzustellen. Da die Labeling-Arbeit zeitaufwändig ist und Fachwissen erfordert, sind wir der Meinung, dass langfristige Konzepte für skalierbare Strategien erforderlich sind. Dabei sollte auch die Möglichkeit der Zusammenarbeit von verwandten Forschungs- und Aktivismus-Projekten sowie die Annotation durch Personen mit weniger Fachwissen in Betracht gezogen werden.

Literaturverzeichnis

Aluru, Sai Saketh/Mathew, Binny/Saha, Punyajoy/Mukherjee, Animesh, 2020. Deep Learning Models for Multilingual Hate Speech Detection https://doi.org/10.48550/arXiv.2004.06465.

Appandurai, Arjun, 1990. Disjuncture and Difference in the Global Cultural Economy. In: Theorie, Kultur & Gesellschaft, Bd. 7, Nr. 2/3.

Ascone, Laura/Becker, Matthias J./Bolton, Matthew/Chapelan, Alexis/Krasni, Jan/Placzynta, Karolina/Scheiber, Marcus/Troschke, Hagen/Vincent, Chloé, 2022. Decoding Antisemitism: Eine KI-gestützte Studie zu Hate Speech und Bildmaterial im Internet. Diskursreport 3. Berlin: Technische Universität Berlin. Zentrum für Antisemitismusforschung. https://doi.org/10.14279/depositonce-14976.

Ascone, Laura/Becker, Matthias J./Bolton, Matthew/Chapelan, Alexis/Krasni, Jan/Placzynta, Karolina/Scheiber, Marcus/Troschke, Hagen/Vincent, Chloé, 2022. Decoding Antisemitism: An AI-Driven Study on Hate Speech and Imagery Online. Diskursreport 4. Technische Universität Berlin. Zentrum für Antisemitismusforschung. https://doi.org/10.14279/DEPOSITONCE-16292.

Basile, Valerio/Bosco, Cristina/Fersini, Elisabetta/Nozza, Debora/Patti, Viviana/Rangel Pardo, Francisco Manuel/Rosso, Paolo/Sanguinetti, Manuela, 2019. SemEval-2019 Task 5: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter. In: Proceedings of the 13th International Workshop on Semantic Evaluation. Minneapolis, Minnesota, USA: Association for Computational Linguistics, 54–63 https://doi.org/10.18653/v1/S19-2007.

Becker, Matthias J., 2018. Analogien der „Vergangenheitsbewältigung". Antiisraelische Projektionen in Leserkommentaren der Zeit und des Guardian. Baden-Baden: Nomos (englische Übersetzung von 2021: Antisemitism in Reader Comments: Analogies for Reckoning with the Past. London: Palgrave Macmillan).

Becker, Matthias J./Ascone, Laura/Bolton, Matthew/Chapelan, Alexis/Krasni, Jan/Placzynta, Karolina/Scheiber, Marcus/Troschke, Hagen/Vincent, Chloé, 2021. Decoding Antisemitism: Eine KI-gestützte Studie zu Hate Speech und Bildsprache im Internet. Diskursreport 2. Berlin: Technische Universität Berlin. Zentrum für Antisemitismusforschung.

Chandra, Mohit/Pailla, Dheeraj/Bhatia, Himanshu/Sanchawala, Aadilmehdi/Gupta, Manish/Shrivastava, Manish/Kumaraguru, Ponnurangam, 2021. „Subverting the Jewtocracy": Online Antisemitism Detection Using Multimodal Deep Learning. http://arxiv.org/abs/2104.05947.

Dixon, Lucas/Li, John/Sorensen, Jeffrey/Thain, Nithum/Vasserman, Lucy, 2018. Measuring and Mitigating Unintended Bias in Text Classification. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. New Orleans LA USA: ACM, 67–73. https://doi.org/10.1145/3278721.3278729.

Elroy, OrAbraham Yosipof, 2022. Analysis of COVID-19 5G Conspiracy Theory Tweets Using SentenceBERT Embedding. In: Artificial Neural Networks and Machine Learning - ICANN 2022: 31st International Conference on Artificial Neural Networks, Bristol, UK, September 6–9, 2022, Proceedings, Part II. Berlin, Heidelberg: Springer, 186–196. https://doi.org/10.1007/978-3-031-15931-2_16.

Falkenberg, Mark/Baronchelli, Andrea, 2023. Wie können wir die Rolle der sozialen Medien bei der Verbreitung von Klimafehlinformationen besser verstehen? Grantham Research Institute on Climate Change and the Environment. Januar 2023. https://www.lse.ac.uk/granthaminstitute/news/how-can-we-better-understand-the-role-of-social-media-in-spreading-climate-misinformation (last accessed on 5 February 2023).

Friendberg, Brian, Wired, July 31st, 2020. The Dark Virality of a Hollywood Blood-Harvesting Conspiracy. https://www.wired.com/story/opinion-the-dark-virality-of-a-hollywood-blood-harvesting-conspiracy (letzter Zugriff am 5. Februar 2023).

González-Pizarro, Felipe/Zannettou, Savvas, 2022. Verstehen und Erkennen von hasserfüllten Inhalten durch kontrastives Lernen. http://arxiv.org/abs/2201.08387.

Horta Ribeiro, Manoel/Jhaver, Shagun/Zannettou, Savvas/Blackburn, Jeremy/Stringhini, Gianluca/De Cristofaro, Emiliano/West, Robert, 2021. „Do Platform Migrations Compromise Content Moderation? Evidence from r/The_Donald and r/Incels. In: Proceedings of the ACM on Human-Computer Interaction 5 (CSCW2), 1–24. https://doi.org/10.1145/3476057.

Hoseini, Mohamad/Melo, Philipe/Benevenuto, Fabricio/Feldmann, Anja/Zannettou, Savvas, 2021. On the Globalization of the QAnon Conspiracy Theory Through Telegram. http://arxiv.org/abs/2105.13020.

Hutchinson, Ben/Prabhakaran, Vinodkumar/Denton, Emily/Webster, Kellie/Zhong, Yu/Denuyl, Stephen, 2020. Social Biases in NLP Models as Barriers for Persons with Disabilities. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 5491–5501. https://doi.org/10.18653/v1/2020.acl-main.487.

Jikeli, Günther/Awasthi, Deepika/Axelrod, David/Miehling, Daniel/Wagh, Pauravi/Joeng, Weejoeng, 2021. Detecting Anti-Jewish Messages on Social Media. Building an Annotated Corpus That Can Serve as A Preliminary Gold Standard. In: Workshop Proceedings of the 15th International AAAI Conference on Web and Social Media. US: ICWSM. https://doi.org/10.36190/2021.14.

Jikeli, Günther/Cavar, Damir/Jeong, Weejeong/Miehling, Daniel/Wagh, Pauravi/Pak, Denizhan, 2022. Toward an AI Definition of Antisemitism? In: Hübscher, Monika/von Mering, Sabine (Hg.). Antisemitism on Social Media. London: Routledge, 193–212.

Jikeli, Günther/Cavar, Damir/Miehling, Daniel, 2019. Annotating Antisemitic Online Content. Towards an Applicable Definition of Antisemitism. https://doi.org/10.5967/3r3m-na89.

Mandl, Thomas/Modha, Sandip/Shahi, Gautam Kishore/Madhu, Hiren/Satapara, Shrey/Majumder, Prasenjit/Schaefer, Johannes, et al., 2021. Overview of the HASOC Subtrack at FIRE 2021: Hate Speech and Offensive Content Identification in English and Indo-Aryan Languages. https://doi.org/10.48550/arXiv.2112.09301.

Mathew, Binny/Saha, Punyajoy/Yimam, Seid Muhie/Biemann, Chris/Goyal, Pawan/Mukherjee, Animesh, 2022. HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, Nr. 17, 14867–14875. https://doi.org/10.48550/arXiv.2012.10289.

Meta, 2022. Community Standards Enforcement | Transparency Center. https://transparency.fb.com/data/community-standards-enforcement (letzter Zugriff am 14. Februar 2023).

Mihaljević, Helena/Steffen, Elisabeth, 2022. How Toxic Is Antisemitism? Möglichkeiten und Grenzen eines automatisierten Toxizitätsscorings für antisemitische Online-Inhalte. In: Proceedings of the 2nd Workshop on Computational Linguistics for Political Text Analysis (CPSS-2022), 1–12. Potsdam, Deutschland.

Moffitt, J. D./King, Catherine/Carley, Kathleen M., 2021. Die Jagd auf Verschwörungstheorien während der COVID-19-Pandemie. In: Social Media + Society, Vol. 7, Nr. 3. https://doi.org/10.1177/20563051211043212.

Phillips, Samantha C./Ng, Lynnette Hui Xian/Carley, Kathleen M., 2022. Hoaxes and Hidden Agendas: A Twitter Conspiracy Theory Dataset: Data Paper. In: Companion Proceedings of the Web Conference 2022. WWW '22. New York, NY, USA: Association for Computing Machinery, 876–880. https://doi.org/10.1145/3487553.3524665.

Pogorelov, Konstantin/Schroder, Daniel Thilo/Burchard, Luk/Moe, Johannes/Brenner, Stefan/Filkukova, Petra/Langguth, Johannes, 2020. FakeNews: Corona-Virus und 5G-Verschwörungsthema bei MediaEval 2020. In: Working Notes Proceedings of the MediaEval 2020 Workshop. http://ceur-ws.org/Vol-2882/paper64.pdf.

Poletto, Fabio/Basile, Valerio/Sanguinetti, Manuela/Bosco, Cristina/Patti, Viviana, 2021. Ressourcen und Benchmark-Korpora für die Erkennung von Hassrede: A Systematic Review. In: Language Resources and Evaluation, Vol. 55, Nr. 2, 477–523. https://doi.org/10.1007/s10579-020-09502-8.

Proust, Serge/Michalon, Jérôme/Maurin, Marine/Noûs, Camille, 2020. Dieudonné: Antisemitismus, moralische Panik und abweichende Gemeinschaft, Deviance and Society, Vol. 44, Nr. 3.

Rajadesingan, Ashwin/Resnick, Paul/Budak, Ceren, 2020. Quick, Community-Specific Learning: How Distinctive Toxicity Norms Are Maintained in Political Subreddits. In: Proceedings of the International AAAI Conference on Web and Social Media, 14, 557–568. https://ojs.aaai.org/index.php/ICWSM/article/view/7323.

Röttger, Paul/Vidgen, Bertram/Nguyen, Dong/Waseem, Zeerak/Margetts, Helen/Pierrehumbert, Janet B., 2021. HateCheck: Functional Tests for Hate Speech Detection Models. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 41–58. https://doi.org/10.18653/v1/2021.acl-long.4.

Solomon, Daniel, K. La Revue, January 5th, 2023. K. La Revue. Kanye and the new Far West of American Antisemitism. https://k-larevue.com/en/kanye-and-the-new-far-west-of-american-antisemitism/ (letzter Zugriff am 14. Februar 2023).

Steffen, Elisabeth/Mihaljević, Helena/Pustet, Milena/Bischoff, Nyco/Varela, María do Mar Castro/Bayramoğlu, Yener/Oghalai, Bahar, 2022. Codes, Patterns and Shapes of Contemporary Online Antisemitism and Conspiracy Narratives -- an Annotation Guide and Labeled German-Language Dataset in the Context of COVID-19. http://arxiv.org/abs/2210.07934.

Vaswani, Ashish/Shazeer, Noam/Parmar, Niki/Uszkoreit, Jakob/Jones, Llion/ Gomez,Aidan N./Kaiser, Łukasz/Polosukhin/Illia, 2017. Attention is all you need. In: Advances in Neural Information Processing Systems 30.

Wiegand, Michael/Siegel, Melanie/Ruppenhofer, Josef, 2018. Überblick über die GermEval 2018 Shared Task zur Identification von Schimpfwörtern. In: Proceedings of GermEval 2018, 14th Conference on Natural Language Processing (KONVENS 2018). https://epub.oeaw.ac.at/0xc1aa5576_0x003a10d2.pdf (letzter Zugriff am 23. Februar 2023).

Wilson, Jason, Southern Poverty Law Center, December 7th, 2022. Kanye's Antisemitic Hate Speech Platformed by Enablers in Tech, Media, Politics. https://www.splcenter.org/hatewatch/2022/12/07/kanyes-antisemitic-hate-speech-platformed-enablers-tech-media-politics (letzter Zugriff am 14. Februar 2023).

Wodak, Ruth, 2020. The Politics of Fear: The Shameless Normalization of Far-Right Discourse. London: Sage Publications.

Zampieri, Marcos/Malmasi, Shervin/Nakov, Preslav/Rosenthal, Sara/Farra, Noura/Kumar, Ritesh, 2019. SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffensEval). In: Proceedings of the 13th International Workshop on Semantic Evaluation. Minneapolis, Minnesota, USA: Association for Computational Linguistics, 75–86. https://doi.org/10.18653/v1/S19-2010.

Zampieri, Marcos/Nakov, Preslav/Rosenthal, Sara/Atanasova, Pepa/Karadzhov, Georgi/Mubarak, Hamdy/Derczynski, Leon/Pitenis, Zeses/Çöltekin, Çağrı, 2020. SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffensEval 2020). http://arxiv.org/abs/2006.07235.

Quellenverzeichnis

Die antisemitischen Äußerungen von Kanye West im Herbst 2022

Vereinigtes Königreich

BBC-FB[20221025] BBC News, 25. Oktober 2022, „Adidas cuts ties with Kanye West," https://www.facebook.com/bbcnews/posts/pfbid0Fwjyxqnrw-Xz4j5kbGTh2n9TKkE3fwpezrmSxhRDS41adSW-pBT5KyrmezfsTXGwyfl.

DAILY-FB[20221019] Daily Mail, 10. Oktober 2022, „Howard Stern compares Kanye West to HITLER following troubled rapper's stream of anti-Semitic remarks," https://www.facebook.com/DailyMail/posts/pfbid-02wx4bNiso7yGDEkVww2PpPGUAHkSrQ34uchJRs-9Qf3eavjJZzQ347YUBx46QhPp7cl.

DAILY-TW[20221029] Daily Mail, 29. Oktober 2022, „Kanye doubles down on anti-Semitic claim Jews control the media by sharing SPREADSHEET 'filled with names of Jewish execs,'" https://twitter.com/MailOnline/status/1586452232844754945.

GUARD-FB[20221025] The Guardian, 25. Oktober 2022, „Kanye West reportedly no longer a billionaire as companies cut ties," https://www.facebook.com/theguardian/posts/pfbid0bxs2WNVbXKzGuaqMDQRTmBCEq-9VouFdR3MCR7TkNTsmgejj7BpC95F161qCh4Xxol.

GUARD-FB[20221026] The Guardian, 26. Oktober 2022, „Kanye West's wax figure removed from public view," https://www.facebook.com/theguardian/posts/pfbid0GzMRz62gLzuZmjqVPPLHQ3B5MtvUFdFf-QvvTnUKCLMfPqYnZhEqdBqWqFxu6mGsal.

INDEP-FB[20221009] The Independent, 9. Oktober 2022, „Calls for Kanye West to be banned from Twitter," https://www.facebook.com/TheIndependentOnline/posts/pfbid02wUyh1E4b3cMdKD9Q941c89hef5fxt-jBwa9Hsip7oWckAA5pWn3TQdvhsxhSn9trrl.

INDEP-FB[20221027] The Independent, 27. Oktober 2022, „Opinion: Kanye West's comeuppance is too little, too late," https://www.facebook.com/TheIndependentOnline/posts/pfbid026XadeqpgthwfwhZzyEecfpd7S-STpY8fTm7nHZHdGh33KjYi1hUz4kCcxFhmp9HgKl.

INDEP-FB[20221029] The Independent, 29. Oktober 2022, „Former Kanye fan burns Yeezy shoe collection after rapper's antisemitic remarks," https://www.facebook.com/TheIndependentOnline/posts/pfbid0gas9LxFdanSgD-w1F2M4jvZX2zysM5yZxX4K8SbJwrkXH2iwdeBj7zZ-hTdQw6U1iWl.

METRO-FB[20221025] Metro, 25. Oktober 2022, „Kanye West 'no longer a billionaire' after being dropped by Adidas," https://www.facebook.com/MetroUK/posts/pfbid029SS15zCBVt2z2JUPmU1UT8rFFiGpGqxuqi-Paj7z3fFa5Gh1bS3ddZoBJGaenUiTRl.

MIRROR-FB[20221021] Daily Mirror, 21. Oktober 2022, „Kanye West dumped by major fashion chain after string of major outbursts," https://www.facebook.com/MirrorCeleb/posts/pfbid0UfZkC4nF19Bae7nfvYBPJgS95j9fAnp32FAzX2FRtB9nJhZcNdeEL2cqWSr9BYbul.

MIRROR-FB[20221103] Daily Mirror, 3. November 2022, „Kanye West announces 'verbal fast' which will see him avoid speaking for 30 days," https://www.facebook.com/MirrorCeleb/posts/pfbid0gdvW4dPKaZSC-zfSHo5HCoPXcmDExArmkjvNYJ8BRKyqQyYhuWrS-JeW9nugJP5WrMl.

SUN-FB[20221020] The Sun, 20. Oktober 2022, „Kanye West storms out of Piers Morgan TalkTV show after furious row," https://www.facebook.com/thesun/posts/pfbid02w8mv4CiUymG8uev6CeuHXFp6eDPwq2n-SfromKVtZpVhxYbJ4AJQ5WWgCez5xiL8Vl.

TIMES-FB[20221026] The Times, 26. Oktober 2022, „Kanye West's school Donda Academy has closed," https://www.facebook.com/timesandsundaytimes/posts/pfbid0yWt9XLq5V1JuWyp7TeHAq9BKRqwCHnfhtdJLyV-D3Y2WuoJs2bpATGKDtq7whUqgdl.

VICE-FB[20221019] Vice UK, 19. Oktober 2022, „Why I've Finally Given Up on Kanye West," https://www.facebook.com/viceuk/posts/pfbid035riKXHngJmQzrn6uXMNH3fA5ejyv5Xbajy13x-72gwtEgmwqh1eS3mMnGh6QRqvUsl.

Frankreich

BFMTV-FB[20221027] BFMTV, 27. Oktober 2022, „La statue de cire de Kanye West retirée du musée Madame Tussauds à Londres," https://www.facebook.com/BFMTV/posts/pfbid0pgh35wsyzC7ekaQxe1iKEr7ntp7nU3SAX23ep9QsWFC4soqbFy4e4R5igi94fgJGl.

BFMTV-FB[20221030] BFMTV, 30. Oktober 2022, „Kanye West s'excuse après ses propos sur George Floyd et ses remarques antisémites," https://www.facebook.com/BFMTV/posts/pfbid0Rsq4cGR2V2HYmNVUnE2hvDBNh7fAYnwXtAaS82RUCKkD2g6rL2yTC4k9MtiVo45bl.

FRENC-TW[20221026] FrenchRapUS, 26. Oktober 2022, " La statue de cire de Kanye West vient d'être retiré au musée Madame Tussauds ! [...]," https://twitter.com/FrenchRapUS/status/1585341880438784001.

LCI.F-FB[20221026] LCI, 26. Oktober 2022, „Propos antisémites : et si c'était (aussi) la fin de la carrière musicale de Kanye West ?," https://www.facebook.com/LCI/posts/pfbid0tL1iFLvzDchxRts2re5RnNimFfe6t5ub8ia5N8FoJYfYiKgEUMW253iEDk9N84MFl.

LEFIG-FB[20221025] Le Figaro, 25. Oktober 2022, „Adidas rompt son partenariat avec Kanye West après des remarques antisémites," https://www.facebook.com/lefigaro/posts/pfbid0TF7f6AiJeA48jvRaoxNLZx2kTxwQUBE3rh43r7UBZKZgpH5HjxGtPmmX6CA7ca95l.

LEPOI-FB[20221025] Le Point, 25. Oktober 2022, „Propos antisémites – Adidas met fin à sa collaboration avec Kanye West," https://www.facebook.com/lepoint.fr/photos/a.389816860702/10158755608000703.

LESIN-FB[20221027] Les Inrockuptibles, 27. Oktober 2022, „Kanye West – Un drame américain," https://www.facebook.com/lesinrockuptibles/posts/pfbid028FjMUGx8SffdLYys261dFrgoMHdyxizoCBAWMwLMWw89d3chWa5dkgwCCDQ5QTFkl.

Die israelischen Wahlen im Dezember 2022 (Deutsche Medien)

ARTED-YT[20221229] ARTE, 29. Dezember 2022, „Israel: Demokratie in der Sackgasse?," https://www.youtube.com/watch?v=x9ok02QXKgk.

SPIEG-TW[20221113] Spiegel, 13. November 2022, „Wenn in Israel eine rechtsradikale Regierung an die Macht kommt, droht eine neue Welle des Antisemitismus – gegen Juden in Europa und Deutschland," https://twitter.com/SPIEGiegel/status/1591850652262862849.

SPIEG[20221113] Spiegel, 13. November 2022, „Sieg der Rechtsparteien in Israel: Eine neue Welle des Judenhasses," https://www.spiegel.de/ausland/israel-eine-neue-rechte-regierung-wird-eine-welle-des-antisemitismus-ausloesen-a-e2388318-3c2b-46c4-a0f6-c7a253164b17.

WELT[20221102] Welt, 2. November 2022, „Netanjahus autoritäre Züge sind für Israels Demokratie gefährlich Rechtsruck bei Israel-Wahl – Netanjahu laut Prognosen vor Comeback," https://www.welt.de/politik/deutschland/article241918157/Wahl-in-Israel-Netanjahus-autoritaere-Zuege-sind-fuer-die-israelische-Demokratie-gefaehrlich.html.

ZDF-YT[20221229] ZDF, 29. Dezember 2022, „Rechts und religiös – Israels neue Regierung | auslandsjournal," https://www.youtube.com/watch?v=mgPY8PapgUc.

ZEIT[20221101] Zeit, 1. November 2022, „Eine gefährliche Wahl," https://www.zeit.de/politik/ausland/2022-11/israel-parlamentswahl-benjamin-netanjahu-rechtsextremismus-faq.

ZEIT[20221216] Zeit, 16. Dezember 2022, „Warum die Ultrarechten in Israel so stark sind," https://www.zeit.de/politik/ausland/2022-11/parlamentswahl-israel-benjamin-netanjahu-extremismus.

ZEIT[20221217] Zeit, 17. Dezember 2022, „Sehr rechts und sehr religiös," https://www.zeit.de/politik/ausland/2022-11/israel-regierungsbildung-koalitionsverhandlung-benjamin-netanjahu.

ZEIT-IG[20221102] Zeit, 2. November 2022, „Parlamentswahl. Warum sind die Ultrarechten in Israel so stark?," https://www.instagram.com/p/CkdMUM4q2Rp/.

Antisemitische Vorfälle bei der FIFA-Fußball-Weltmeisterschaft 2022 in den britischen Medien

TW-AMRO[20221206] Twitter, 6. Dezember 2022, „Morocco celebrates their victory by raising the Palestinian flag 🇵🇸 [...],“ https://twitter.com/_amroali/status/1600193400598253568.

TW-CARTER[20221204] Twitter, 4. Dezember 2022, „England football fan chants 'FREE PALESTINE' in Israel TV interview following win over Senegal,“ https://twitter.com/Bob_cart124/status/1599534417755897856.

TW-CARTER[20221210] Twitter, 10. Dezember 2022, „Just imagine if Morocco wins the World Cup in Qatar! [...],“ https://twitter.com/Bob_cart124/status/1601643552630415360.

TW-HARAWI[20221127] Twitter, 27. November 2022, “🧵1/4 Israeli journalists at the Qatar World Cup are being shunned left, right & center by fans, particularly from the Middle East but also beyond,“ https://twitter.com/yarahawari/status/1596879671169527808.

TW-JASKOLL[20221126] Twitter, 26. November 2022, „People enjoying FIFA in Qatar, a wild abuser of human rights, oppressor of homosexuals & women, master of slave labor, funder & harborer of terrorists [...],“ https://twitter.com/skjask/status/1596596107357790208.

TW-JENNINE[20221126] Twitter, 26. November 2022, „more people rejecting the isr@eli news channels. small acts of boycott by the people ❤️,“ https://twitter.com/jennineak/status/1596625234752598016.

TW-KOMUGISHA[20221201] Twitter, 1. Dezember 2022, „Morocco 🇲🇦 hoist the Palestine 🇵🇸 flag after their 2–1 victory over Canada in solidarity with Palestine,“ https://twitter.com/UsherKomugisha/status/1598364139801419776.

TW-MILSTEIN[20221127] Twitter, 27. November 2022, “#Qatar's disastrous #WorldCup is so xenophobic that some fans are now attacking Arab journalists accusing them of being Israelis [...],“ https://twitter.com/AdamMilstein/status/1596740211060273152.

TW-OGDEN[20221206] Twitter, 26. November 2022, „Morocco celebrate their win against Spain with a Palestinian flag [...],“ https://twitter.com/MarkOgden_/status/1600189047812390917.

TW-SAKIB[20221202] Twitter, 2. Dezember 2022, „This is killing me,“ https://twitter.com/mertesakib/status/1598698752574996480.