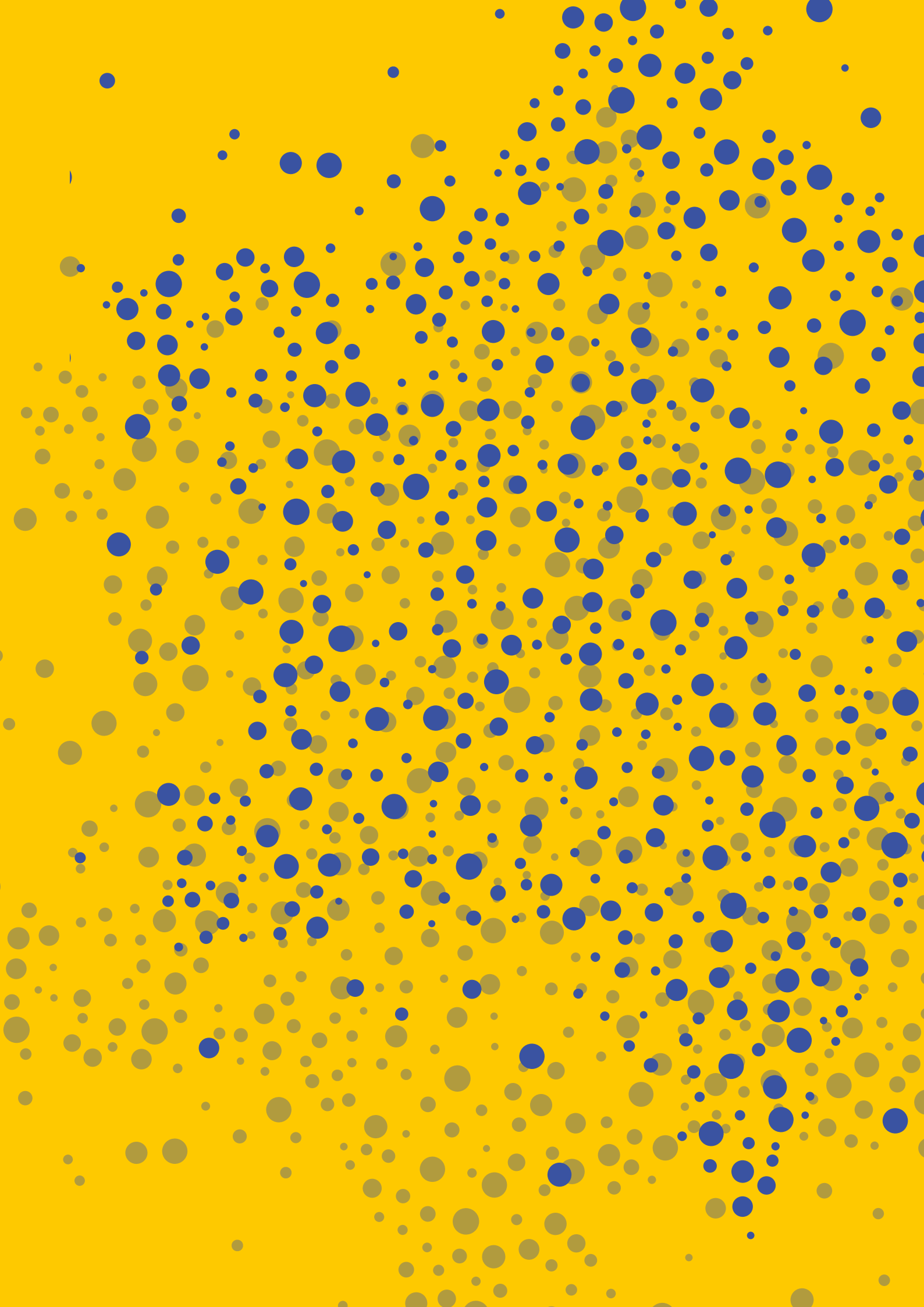


ANTISEMITISME in het Nederlandse SOCIALE MEDIA LANDSCHAP

METHODOLOGISCH

DOCUMENT



Inhoudsopgave

1. Introductie	3
2. Definitie	5
2.1 Definitie	5
2.2 Uitleg bij de IHRA, Nexus White Paper en JDA	5
2.3 Grijze zone rond strafbaarheid van antisemitisme	6
3. Methodologie	7
3.1 Methodologie	7
Dataverzameling en afbakening	7
Platform selectie	7
Technische analyse	7
Onderzoeksdoelstellingen	8
4. Introductie technische aspecten	9
4.1 Dataverzameling	9
Uitleg platformen	10
Kanttekening: Evolutie dataverzamelmethode en implicaties voor longitudinale analyse	11
4.2 Toxiciteits-algoritme	13
4.3 Antisemitisme Classifier	14
Keuze model	14
Trainingsdataset en modelontwikkeling	15
Modelprestaties	17
4.4 Meme database en multimodale modellen	17
5. Kwantitatieve big data analyse: kanttekeningen	21

Een methodologisch document en appendix zijn beschikbaar die uitleg geven bij de gebruikte terminologie, definities, methodologie en technische aspecten. We raden aan deze documenten voor het lezen van de jaarrapporten door te nemen.

1. Introductie

Dit methodologisch document biedt een toelichting op de rapporten over antisemitisme in het Nederlandse sociale media landschap over de periode 2021-2024. Het document behandelt achtereenvolgens de gehanteerde definitie van antisemitisme, de toegepaste methodologie, de technische aspecten van het onderzoek, en een reflectie op de mogelijkheden en beperkingen van big data analyse.

Hoofdstuk 2 expliciteert de gehanteerde definitie van antisemitisme, waarbij bijzondere aandacht wordt besteed aan antisemitisme dat zich uit in de vorm van anti-Israël of anti-zionistische retoriek. Dit hoofdstuk onderzoekt het spanningsveld tussen enerzijds legitieme kritiek op Israël en het zionisme, en anderzijds antisemitische uitingen. Daarnaast wordt een juridisch kader geschetst met betrekking tot de strafbaarheid van antisemitisme in Nederland.

Hoofdstuk 3 beschrijft de gehanteerde methodologie. Achtereenvolgens worden de methoden voor dataverzameling en platformselectie toegelicht: welke data werd verzameld en op basis van welke criteria werd deze geselecteerd, van welke platformen is de data afkomstig en wat motiveerde de keuze voor deze specifieke platformen? Tevens worden in dit hoofdstuk de onderzoeksdoelstellingen geformuleerd.

Hoofdstuk 4 behandelt de technische aspecten van het onderzoek in meer detail. Allereerst worden de toegepaste technieken voor dataverzameling besproken. Vervolgens licht dit hoofdstuk de methodologie toe die is gebruikt om antisemitisme, subcategorieën van antisemitisme en andere vormen van problematische content in online berichten te detecteren. Hierbij worden de ontwikkeling en effectiviteit van Textgain's toxiciteitsalgoritme en de specifieke antisemitisme-classifier uiteengezet. Tot slot biedt dit hoofdstuk inzicht in de technieken die zijn aangewend voor de analyse van memes en andere visuele content¹.

Hoofdstuk 5 reflecteert op de waarde en beperkingen van big data analyse, waarbij zowel de mogelijkheden als de inherente uitdagingen van deze onderzoeksbenadering worden besproken.

¹ Zie het technische rapport voor een meer diepgaande uiteenzetting van de technische aspecten

2. Definitie

Een grondige analyse van antisemitisme vereist allereerst een heldere definitie van het begrip zelf. Dit hoofdstuk behandelt daarom zowel de definiëring van antisemitisme als de vraag in hoeverre antisemitische uitingen strafbaar zijn onder de Nederlandse wetgeving. Definitie en strafbaarheid worden afzonderlijk besproken om conceptuele duidelijkheid te behouden.

2.1 Definitie

De definitie van antisemitisme is gecompliceerd, omdat het zowel expliciete haat als subtiele vooroordelen omvat en de grenzen raakt met vrijheid van meningsuiting en kritiek op Israël en zionisme. Wat sommigen zien als legitieme kritiek, ervaren anderen als antisemitisme, waardoor interpretaties verschillen en een heldere afbakening complex is. Dit bemoeilijkt eveneens het strafbaar stellen van antisemitisme. In dit onderzoek kozen we ervoor de definitie van de International Holocaust Remembrance Alliance (IHRA)² als richtlijn te nemen. Wat betreft het onderscheid tussen enerzijds antisemitisme en anderzijds politieke retoriek met een anti-zionistisch of anti-Israël sentiment, vinden we de voorbeelden van IHRA echter te weinig concreet. Voor dit onderscheid baseren we ons daarom op de *Nexus White Paper*³ en *The Jerusalem Declaration on Antisemitism* (JDA)⁴.

De Nexus White Paper en JDA nemen een genuanceerde positie in door te benadrukken dat kritiek op Israël en/of zionisme op zichzelf niet antisemitisch is, tenzij:

1. Gekende antisemitische tropen worden gebruikt. Denk hierbij o.a. aan antisemitische stereotypen, vooroordelen of samenzweringstheorieën.
2. De Joodse gemeenschap collectief verantwoordelijk wordt gehouden voor het handelen en beleid van Israël of voor daden van Israëlische groeperingen of individuele Israëliërs.

Het huidige Israëlisch-Palestijnse conflict maakt het onderscheid tussen enerzijds geoorloofde kritiek op Israël en/of zionisme en anderzijds antisemitisme zeer actueel. Met name de herkenning van antisemitisme verhoudt als anti-Israël retoriek vormt een belangrijk aandachtspunt. **Binnen dit onderzoek hechten we er daarom belang aan dit onderscheid ten eerste te benadrukken en baseren we onze definitie op dit vlak op zowel de JDA als de Nexus-definitie, zoals hierboven beschreven.**

2.2 Uitleg bij de IHRA, Nexus White Paper en JDA

- De *International Holocaust Remembrance Alliance* (IHRA) biedt een veelgebruikte definitie van antisemitisme. Deze definitie, die op 26 mei 2016 werd aangenomen, is bedoeld als richtlijn om antisemitische incidenten en gedragingen te herkennen en aan te pakken. De definitie luidt als volgt: “Antisemitisme is een bepaalde opvatting over Joden, die kan worden geuit als haat jegens Joden. Retorische en fysieke uitingen van antisemitisme zijn gericht op Joodse of niet-Joodse personen en/of hun eigendommen, op Joodse gemeenschapsinstellingen en religieuze voorzieningen.”
- De *Nexus White Paper* over antisemitisme biedt een verder ingevulde definitie die erop gericht is aan te geven wanneer kritiek op Israël en het zionisme overgaat in antisemitisme. Deze definitie is vooral belangrijk omdat deze het onderscheid probeert te verduidelijken tussen legitieme kritiek en discriminerende retoriek. Volgens het Nexus White Paper houdt antisemitisme in: “een aanhoudende demonisering die Joden niet alleen neerzet als ‘anderen’ (d.w.z. als wezenlijk anders of vreemd), maar ook als onherroepelijk bedreigend en gevaarlijk machtig”⁵.
- De *Jerusalem Declaration on Antisemitism* (JDA) is een verklaring die bedoeld is om een helder en

² <https://holocaustremembrance.com/resources/de-werkdefinitie-van-antisemitisme-van-de-ihra>

³ <https://nexusproject.us/the-nexus-white-paper-understanding-antisemitism-at-its-nexus-with-israel-and-zionism/>

⁴ <https://jerusalemdeclaration.org/>

⁵ “Antisemitism consists of anti-Jewish attitudes, actions or systemic conditions. It includes negative beliefs and feelings about Jews, hostile behavior directed against Jews, and conditions that discriminate against Jews and impede their ability to participate as equals in political, religious, cultural, economic, or social life.”

genueanceerd kader te bieden voor het herkennen en aanpakken van antisemitisme, terwijl het de vrijheid van meningsuiting en legitieme kritiek op Israël beschermt. De JDA definieert antisemitisme als: “Discriminatie, vooroordeel, vijandigheid of geweld tegen Joden of Joodse instellingen om het feit dat ze Joodse zijn”.

2.3 Grijs zone rond strafbaarheid van antisemitisme

In het strafrecht is antisemitisme geen zelfstandig delict. In Nederland kan antisemitisme desalniettemin strafbaar zijn wanneer een uiting of gedraging voldoet aan de criteria van discriminatie in de zin van artikel 429quater⁶ van het Wetboek van Strafrecht, dan wel aan die van groepsbelediging⁷ of het aanzetten tot haat en discriminatie⁸, zoals neergelegd in de artikelen 137c tot en met 137e⁹ van het Wetboek van Strafrecht. Of een concrete uiting of gedraging strafbaar is, dient beoordeeld te worden aan de hand van alle omstandigheden van het geval; er bestaat geen scherp afgebakende grens, en het is uiteindelijk de rechter die hierover beslist.

Sinds 1 oktober 2024 is de ontkenning van genocides, waaronder de Holocaust, zelfstandig strafbaar gesteld als specifieke vorm van groepsbelediging, op grond van artikel 137c lid 2 van het Wetboek van Strafrecht. De Memorie van Toelichting¹⁰ bij het desbetreffende wetsvoorstel verduidelijkt dat hiermee wordt aangesloten bij de bestaande jurisprudentie van

⁶ <https://www.wetboek.org/sr/429quater>

⁷ [https://nl.wikipedia.org/wiki/Groepsbelediging_\(Nederland\)](https://nl.wikipedia.org/wiki/Groepsbelediging_(Nederland))

⁸ <https://www.problemenmetjustitie.nl/onderwerpen/aanzetten-tot-haat-discriminatie-en-gewelddadig-optreden/>

⁹ <http://www.wetboek-online.nl/wet/Wetboek%20van%20Strafrecht.html#1270>

¹⁰ <https://www.rijksoverheid.nl/onderwerpen/nieuw-wetboek-van-strafvordering/documenten/kamerstukken/2026/03/23/tk-memorie-van-toelichting-inzake-eerste-aanvullingswet-nieuw-wetboek-van-strafvordering>

de Hoge Raad. Vanwege de bijzondere historische en geografische verbondenheid van Nederland met de Holocaust neemt de ontkenning, vergoelijking of verregaande bagatellisering van deze genocide een uitzonderingspositie in: anders dan bij andere internationale misdrijven wordt dergelijke uitlatingen geacht op zichzelf een beledigend karakter voor de Joodse gemeenschap te hebben, zonder dat daarvoor bijkomende omstandigheden vereist zijn.

Naast uitingsdelicten kunnen ook andere strafbare feiten, zoals mishandeling of vernieling, een antisemitisch karakter dragen. Op grond van artikel 44bis van het Wetboek van Strafrecht¹¹, zoals wettelijk vastgelegd per 1 juli 2025, kunnen de maximumstraffen voor dergelijke delicten met een derde worden verhoogd indien sprake is van een discriminatoir oogmerk, hetgeen eveneens geldt voor antisemitisch gemotiveerde delicten.

Bij zowel groepsbelediging (artikel 137c Sr) als het aanzetten tot haat of discriminatie (artikel 137d Sr) speelt het recht op vrijheid van meningsuiting een wezenlijke rol, en is de beoordeling van strafbaarheid steeds afhankelijk van de specifieke omstandigheden van het geval. Een nauwkeurig onderscheid tussen strafbare en niet-strafbare vormen van antisemitisme valt buiten de reikwijdte van een kwantitatieve data-analyse zoals het onderhavige onderzoek. De gehanteerde methodologie biedt echter wel inzicht in subcategorieën van antisemitisme die een verhoogde kans op strafbaarheid hebben, waaronder antisemitische bedreigingen met de dood of met geweld en holocaustontkenning (negationisme), waarbij dit laatste delict, gezien de historische context, zowel in Nederland als in breder Europees verband een bijzondere positie inneemt.

¹¹ <https://www.wetboek.org/sr/44bis>

3. Methodologie

Dit hoofdstuk beschrijft de onderzoeks–methodologie en technische aanpak die zijn gehanteerd voor de analyse van antisemitisme in het Nederlandse sociale media landschap. Door gebruik te maken van geavanceerde big data-technieken, Natural Language Processing en artificiële intelligentie hebben we een grootschalige, kwantitatieve analyse uitgevoerd van miljoenen Nederlandse social media berichten. Deze systematische benadering stelde ons in staat om zowel de omvang als de aard van antisemitische uitingen op verschillende platforms in kaart te brengen, evenals de onderliggende patronen en netwerken die bijdragen aan de verspreiding ervan.

3.1 Methodologie

Dataverzameling en afbakening

Voor dit onderzoek hebben we een geautomatiseerde, kwantitatieve big data analyse uitgevoerd van miljoenen Nederlandse social media berichten. De dataverzameling richtte zich uitsluitend op publiek geposte inhoud; berichten uit privégroepen en -chats werden niet verzameld.

Het Nederlandse geografische gebied werd op twee manieren afgebakend: waar mogelijk door gebruikers die hun locatie publiek maakten, en anders door automatische detectie van de Nederlandse taal. Deze benadering betekende dat ook Engelstalige berichten van Nederlandse gebruikers werden meegenomen in de analyse. Andere talen werden buiten beschouwing gelaten. Hierbij houden we rekening met de beperking dat een klein deel van de Nederlandstalige berichten mogelijk afkomstig is van Nederlandstalige gebruikers buiten Nederland.

Platform selectie

Om een representatieve dwarsdoorsnede van het Nederlandse social media landschap te verkrijgen, verzamelden we data van een breed scala aan platformen. Dit omvatte grote Very Large Online platformen (VLOPs¹²) zoals Twitter/X, Facebook en Instagram, maar ook meer alternatieve, zogenoemde ‘fringe’¹³, media zoals 4chan en Telegram, evenals

lokale Nederlandse platformen zoals GeenStijl en Marokko.nl.

Technische analyse

De analyse werd uitgevoerd met behulp van verschillende, veelal inhouse ontwikkelde, technieken en AI-modellen:

- Toxiciteitsdetectie: AI-modellen voor het detecteren en meten van algemene toxiciteit, antisemitisme in het specifiek en subcategorieën van antisemitisme in tekst.
- Netwerkanalyse: Voor het identificeren van verspreidingspatronen en gebruikersnetwerken. Aan de hand van deze techniek kunnen onder andere superverspreiders en echokamers in kaart gebracht worden.
- Topic-extractie: Clustering-algoritmes voor het ontdekken van thematische patronen, zoals terugkerende narratieven, topics en thema’s.
- Beeldanalyse: Multimodale AI-modellen voor de analyse van antisemitische memes en visuele content.
- Geautomatiseerde inhoudssamenvatting: Generatieve LLM’s voor het samenvatten van topics en narratieven.

Onderzoeksdoelstellingen

Deze methodologie stelde ons in staat om 5 kernanalyses uit te voeren: (1) het vaststellen van de graad en ernst van antisemitisme in Nederlandse social media berichten inclusief sub- en aanverwante categorieën van antisemitisme zoals negationisme en antisemitische (doods)bedreigingen, (2) het in kaart brengen van netwerken en verspreidingspatronen, inclusief echokamers en superverspreiders, (3) het identificeren van terugkerende narratieven en thema’s in antisemitische berichten, (4) de analyse van topics in antisemitische memes en visuele content en (5) het meten van de aanwezigheid van antisemitisme in het online discours over een langere periode (tijdslijn) en het analyseren van eventueel aanwezige pieken en dalen.

¹² <https://digital-strategy.ec.europa.eu/en/policies/dsa-vlops>

¹³ <https://www.politico.eu/article/fringe-social-media-telegram-extremism-far-right/>

Uitleg platformen

- **Facebook (Meta, VLOP¹⁸)** is tot dusver sterk gemodereerd. Facebook geeft echter ook de mogelijkheid tot het aanmaken van privé-groepen en chats die minder tot niet gemodereerd zijn. Hiervan kunnen wij echter geen data verzamelen en deze blijven dus onder de radar in dit onderzoek. Facebook kenmerkt zich verder als een platform dat minder en minder aantrekkelijk is voor jongeren.
- **Instagram (Meta, VLOP)** is populairder bij jongeren (gen-z en millennials) en voornamelijk bedoeld om persoonlijke foto's te delen. Het platform is gekend om zijn "influencers"¹⁹, professionele gebruikers die dikwijls gesponsorde foto's delen en producten promoten.
- **Twitter/X (VLOP)** spitst zich toe op korte berichten (microblogs) en voorzag tot 2023 uitgebreide moderatie, waarbij schadelijke inhoud zoals desinformatie of belaging vaak verwijderd werd conform EU-regulering. Eind 2022 werd Twitter overgenomen door Elon Musk en omgedoopt tot X, waarop een terugkeer van o.a. eerder geblokkeerde trollen en extremistische accounts naar het platform volgde en daarbij een stijging in problematische content²⁰. Vanaf begin 2023 werd de moderatie op Twitter ook daadwerkelijk teruggeschoefd met opnieuw een toename in schadelijke inhoud tot gevolg²¹. De data uit deze onderzoeken is dus zowel afkomstig uit het pre-Musk tijdperk (2021) als uit de woelige periode waarin Musk Twitter overnam en moderatie werd beperkt (2022 t/m 2024).
- **TikTok (ByteDance, VLOP)** is een video-gebaseerd sociaal mediaplatform dat enorm populair is bij jongeren. Hoewel het platform enige moderatie toepast, is het gekend om zijn algoritmes dat gebruikers razendsnel kan meeslepen in subculturen en echokamers waarin meer en meer schadelijke content en misinformatie wordt gedeeld (zogenoemde rabbit holes²²). Het platform wordt geëxploiteerd door het Chinese bedrijf ByteDance²³ en heeft te kampen met bezorgdheden rond gegevensprivacy en Chinese staatscontrole²⁴.
- **YouTube (Google, VLOP)** is 's werelds grootste videoplatform en staat bekend om zijn enorme bereik en diversiteit aan content. Het platform heeft uitgebreide algoritmes voor contentmoderatie, maar worstelt nog steeds met het effectief bestrijden van desinformatie en extremistische content. Het aanbevelingsalgoritme van YouTube is bekritiseerd omdat het gebruikers steeds extremere content kan aanbevelen (zogenoemde rabbit holes). Het platform heeft de afgelopen jaren zijn moderatiebeleid aangescherpt, maar blijft een onvrijwillige verspreider van problematische inhoud.
- **Reddit (VLOP)** is een verzameling van talloze subcommunities (subreddits) die elk hun eigen moderatieregels hanteren. Sommige subreddits zijn zeer strikt in hun moderatie, terwijl andere veel vrijheid toestaan. Het platform heeft moeite gehad met het beheersen van haatdragende gemeenschappen en desinformatie (bijvoorbeeld misogyne incel communities²⁵), hoewel het in recente jaren strengere richtlijnen heeft ingevoerd.
- **Telegram (fringe²⁶)** is een berichtendienst (zoals Whatsapp) die zich profileert als een privacy-georiënteerd platform met end-to-end encryptie. Het heeft een reputatie verworven als een vrijhaven voor illegale handelingen (wapen- en drugshandel), groepen met extreme politieke

¹⁸ Very Large Online Platforms: <https://digital-strategy.ec.europa.eu/en/policies/dsa-vlops>

¹⁹ Personen die zeer actief zijn op sociale media en veel fans hebben. Daardoor kunnen ze veel invloed uitoefenen op hun specifieke doelgroep, waaronder het consumeren van aangeprezen goederen of producten of het populariseren van bepaalde denkwijzen of ideologieën en bijbehorende gedragingen.

²⁰ <https://digitalplanet.tufts.edu/musk-monitor-under-musk-hate-speech-is-rising/>

²¹ <https://arxiv.org/html/2304.04129v3>

²² https://en.wikipedia.org/wiki/Down_the_rabbit_hole

²³ <https://nl.wikipedia.org/wiki/ByteDance>

²⁴ <https://www.tijd.be/politiek-economie/europa/algemeen/530-miljoen-euro-boete-voor-tiktok-dat-gebruikersdata-naar-china-stuurt/10605337.html>

²⁵ <https://gnet-research.org/2022/05/23/examining-incel-subculture-on-reddit/>

²⁶ Fringe platformen zijn kleinere en vaak minder bekende, alternatieve platformen met miljoenen, veelal anonieme, gebruikers en zeer beperkte content moderatie. Vaak staan ze bekend om het verspreiden van problematische content, zoals 4chan.

standpunten, complottheorieën en groepen die elders zijn verbannen. De moderatie is reactief²⁷, waardoor het platform vaak wordt gebruikt voor het verspreiden van desinformatie en radicale boodschappen²⁸. Sinds de arrestatie van topman Pavel Durov eind 2024 is de moderatie echter enigszins aangescherpt en werkt het platform meer mee met overheden en politie²⁹. De data uit dit onderzoek valt dus zowel onder de periode van beperkte moderatie (2021 t/m 2023) als de periode van verbeterde moderatie (2024).

- **4chan (fringe)** is een website (imageboard³⁰) waar gebruikers volledig anoniem en zonder moderatie berichten kunnen delen. Het platform ligt aan de basis van tal van complottheorieën (bijvoorbeeld “Pizzagate³¹” en “QAnon³²”), haatcampagnes (bijvoorbeeld “Operation Pridefall³³”) en andere controverses. Zo kondigden verschillende massaschutters er hun aanslag aan en deelden zij zelfs livestreams van aanslagen op het platform³⁴.
- **9gag (fringe)** is een platform dat zich richt op het delen van memes, grappige afbeeldingen en korte video’s. Het heeft een beperkte moderatie en staat bekend om zijn soms controversiële en aanstootgevende humor, die vaak grenst aan cyberpesten en discriminerende en extremistische inhoud.

²⁷ Moderatie volgt enkel na melding van gebruikers en zelfs dan is werkelijk ingrijpen minimaal

²⁸ <https://eenvandaag.avrotros.nl/eenvandaag-vraagt/hoe-veilig-is-telegram-en-andere-vragen-over-de-berichtendienst-beantwoord-152056>

²⁹ https://en.wikipedia.org/wiki/Arrest_and_indictment_of_Pavel_Durov

³⁰ Forum constructie die focust op het posten van afbeeldingen met bijgevoegde tekst: <https://en.wikipedia.org/wiki/Imageboard>

³¹ Complottheorie die viraal ging tijdens de presidentsverkiezingen van 2016 in de Verenigde Staten: <https://nl.wikipedia.org/wiki/Pizzagate>

³² Extreemrechtse complottheorie die de zogenaamde plannen van een overmachtige “deep state” en hun aanhangers beschrijft. De complottheorie werd initieel verspreid door een individu of groep van individuen onder het pseudoniem ‘Q’ (o.a. Op 4chan: <https://nl.wikipedia.org/wiki/QAnon>)

³³ Grotendeels door 4chan georchestreerde anti-LHBTI haatcampagne: <https://www.vice.com/en/article/operation-pridefall-4chan-gay-pride-2020/>

³⁴ <https://www.theguardian.com/us-news/2022/may/18/twitch-discord-4chan-buffalo-shooting-investigation-new-york>

- **GeenStijl (fringe, Nederlands)** is een opinie- en blogplatform dat bekendstaat om zijn provocerende en veelal polariserende toon. Het richt zich vooral op nieuws, politiek en maatschappelijk debat, vaak gebracht met satire en scherpe headlines. Moderatie is beperkt, en de community staat erom bekend dat discussies regelmatig ontsporen richting grove of discriminerende uitingen.
- **Dumpert (fringe, Nederlands)** is een populair platform voor het delen van korte video’s, foto’s en memes, vaak met een humoristische of choquerende insteek. Het is sterk verbonden aan GeenStijl (tot eind 2018 onderdeel van dezelfde uitgever^{35,36}). Hoewel er enige moderatie is, staat de commentsectie bekend als een plek waar grof taalgebruik en kwetsende opmerkingen frequent voorkomen.
- **Marokko.nl (Nederlands)** is een forum dat oorspronkelijk gericht is op jongeren met een Marokkaans-Nederlandse achtergrond. Het platform functioneert als een online gemeenschap waar thema’s zoals actualiteit, cultuur, identiteit, religie, en dagelijks leven worden besproken. Hoewel de insteek voornamelijk gemeenschapsvorming is, heeft het forum door de jaren heen ook kritiek gekregen omdat in sommige discussies discriminerende, seksistische en polariserende uitingen te vinden zijn.
- **blckbx (fringe, Nederlands)** is een relatief nieuw online mediaplatform dat zich profileert als alternatief nieuws- en opiniekanaal. Het richt zich sterk op maatschappelijk en politiek debat en staat bekend om zijn kritische houding tegenover gevestigde media, overheid en instituties. Het platform biedt artikelen, video’s en discussies, en trekt een publiek dat geïnteresseerd is in alternatieve perspectieven, vaak rond thema’s als gezondheidszorg, politiek en vrijheidsrechten. Moderatie is beperkt, waardoor in de reacties ook polariserende en soms complottheoretische standpunten voorkomen. Van blckbox is data geanalyseerd vanaf 2022.

³⁵ https://nl.wikipedia.org/wiki/News_Media

³⁶ https://nl.wikipedia.org/wiki/Mediahuis_Nederland

- **Alternatieve media (WordPress³⁷) sites (fringe, Nederlands)** zijn een reeks van kleinere, Nederlandse blog- en nieuwssites, veelal gepubliceerd met Wordpress, die een alternatieve kijk op de actualiteit bieden. Hierbij kan het gaan om onderbelichte topics en andere perspectieven. Alternatieve media³⁸ staat echter veelal bekend om de verspreiding van mis- en desinformatie en complottheorieën, zowel in nieuwsartikelen en blogposts als in commentsecties. Voorbeelden van dergelijke sites zijn *NineForNews*³⁹ en *Cafe Weltschmerz*⁴⁰. Van deze sites is data geanalyseerd vanaf 2023.
- **Soyzellig** is een Nederlands imageboard dat qua structuur en functionaliteit vergelijkbaar is met 4chan. Het platform werd in 2023 opgericht en hanteert een model van minimale moderatie waarbij gebruikers anoniem kunnen posten. Net als zijn internationale equivalent kenmerkt Soyzellig zich door de verspreiding van extreem problematische en potentieel radicaliserende content, waaronder bedreigingen, complottheorieën en zogenaamde hate memes. Van deze site is data geanalyseerd vanaf 2023.

Kanttekening: Evolutie dataverzamelingsmethodologie en implicaties voor longitudinale analyse

Hoewel de fundamentele dataverzamelingsstrategie – bestaande uit zowel volledig willekeurige steekproeven als keyword-gebaseerde verzameling – consistent is gebleven over de vier onderzochte jaren, heeft de praktische implementatie hiervan een significante methodologische evolutie doorgemaakt die directe consequenties heeft voor de samenstelling en vergelijkbaarheid van de datasets.

In 2021 en 2022 was de dataverzameling grotendeels afhankelijk van het retrospectief ophalen en reconstrueren van historische data. Deze aanpak bracht inherente beperkingen met zich mee: veel platformen bewaren publieke gegevens slechts tijdelijk, oude content wordt regelmatig gemodereerd

en verwijderd, historische datasets zijn niet systematisch gearchiveerd, en de toegang tot deze data is vaak technisch en juridisch beperkt. Deze situatie resulteerde in een ongelijke datacapaciteit per platform. Twitter-gegevens waren relatief goed beschikbaar via publieke API's en externe archieven, waardoor omvangrijke datasets konden worden verzameld. Daarentegen waren platforms zoals Facebook, Instagram, YouTube en TikTok aanzienlijk moeilijker toegankelijk voor historische dataverzameling, resulterend in zeer beperkte datasets van deze platformen. Het gevolg was een sterk Twitter-gedomineerde dataset voor de jaren 2021 en 2022.

Vanaf halverwege 2023 werd een fundamentele transitie ingezet naar een meer gestructureerde en systematische real-time dataverzameling. In plaats van te vertrouwen op historische reconstructie, werd data continu verzameld van meerdere platformen tegelijkertijd. Deze methodologische verschuiving maakte het mogelijk om substantieel meer data te verkrijgen van platformen die voorheen ondervetegenwoordigd waren. Het jaar 2023 fungeerde hierbij als overgangsjaar: de eerste helft van het jaar bestond nog grotendeels uit historische dataverzameling volgens het oude model, terwijl de tweede helft profiteerde van de nieuwe systematische aanpak. Dit verklaart de tussenliggende positie van 2023 wat betreft de platformdistributie.

In 2024 werd voor het eerst een volledig jaar aan data verzameld volgens de verbeterde methodologie. Dit resulteerde niet alleen in een substantieel hoger totaalvolume (38,9 miljoen berichten), maar vooral in een fundamenteel andere platformdistributie. In 2021-2022 domineert Twitter de datasets, terwijl we in 2024 een veel gebalanceerdere platformdistributie zien. Deze verschuiving weerspiegelt geen verandering in gebruikersgedrag of platformpopulariteit, maar is direct toe te schrijven aan de verbeterde datacapaciteit en systematische verzamelingsmethodologie.

Deze methodologische evolutie heeft belangrijke implicaties voor vergelijkingen over de jaren heen (longitudinale analyses). Verschillende sociale mediaplatformen vertonen structureel verschillende niveaus van problematische content als gevolg van variaties in platformcultuur, moderatiebeleid, gebruikersdemografie en communicatievormen.

³⁷ Software om gemakkelijke en snel webcontent te publiceren, zoals blogsites: <https://en.wikipedia.org/wiki/WordPress>

³⁸ https://en.wikipedia.org/wiki/Alternative_media

³⁹ <https://www.ninefornews.nl>

⁴⁰ <https://cafeweltschmerz.nl>

Twitter en fringe platformen zoals 4chan en Telegram kenmerken zich door relatief frequente extreme, haatdragende en antisemitische uitingen, mede door een cultuur van anonimiteit en minimale moderatie. Platforms als Instagram en YouTube daarentegen vertonen structureel lagere percentages van dergelijke problematische content.

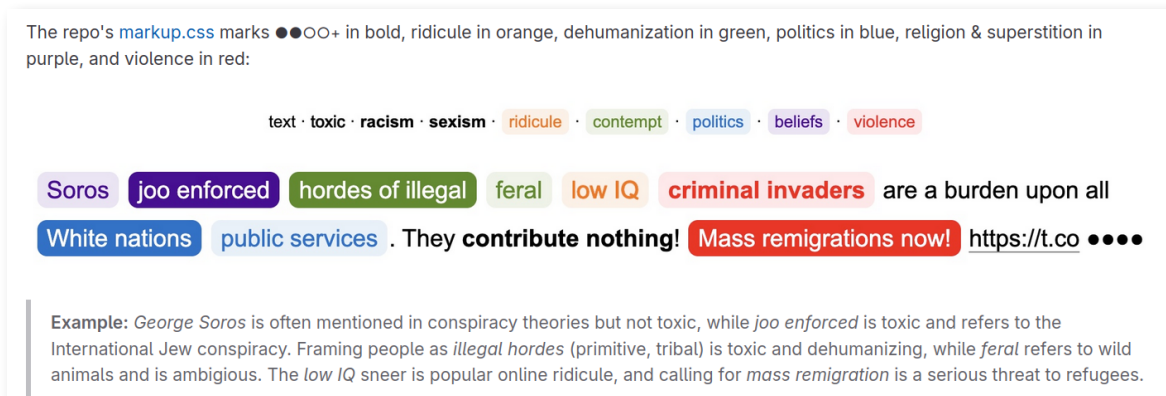
Het directe vergelijken van de prevalentie van antisemitisme tussen de jaren zou daarom tot methodologisch incorrecte conclusies kunnen leiden. Een ogenschijnlijke daling in het percentage antisemitische berichten in 2024 ten opzichte van 2021 zou niet noodzakelijkerwijs een daadwerkelijke afname van antisemitisme weerspiegelen, maar zou grotendeels kunnen worden verklaard door de verhoudingsgewijs veel grotere aanwezigheid van structureel neutralere platforms in de meer recente dataset.

Om deze methodologische valkuil te vermijden en statistisch valide longitudinale vergelijkingen mogelijk te maken, zal voor analyses die zich specifiek richten op trends over de vier jaren heen een harmonisatie van de platfordistributie noodzakelijk zijn. Concreet betekent dit dat de datasets van 2023 en 2024 worden aangepast om een platfordistributie te weerspiegelen die vergelijkbaar is met die van 2021 en 2022, door middel van gewogen sampling^{41,42}. Deze correctie zorgt ervoor dat geobserveerde verschillen

Het is belangrijk te benadrukken dat deze harmonisatie uitsluitend van toepassing is op specifieke longitudinale vergelijkingen tussen de jaren. Voor andere analytische componenten van dit onderzoek, zoals netwerkanalyse, topicmodellering of analyse van memes, blijven de volledige, ongewogen datasets van 2023 en met name 2024 het meest geschikt, aangezien deze analyses juist profiteren van de verbeterde representativiteit en gediversifieerde platfordkking die de nieuwe methodologie biedt.

4.2 Toxiciteits-algoritme

Het toxiciteits-algoritme is een geavanceerd systeem dat problematische (toxische) online inhoud identificeert in meer dan 30 talen, waaronder alle EU-talen aangevuld met Arabisch, Albanees, Macedonisch, Turks, Russisch en Oekraïens. Dit systeem analyseert berichten op een transparante en genuanceerde wijze, waarbij het een **toxiciteitscore** toekent tussen 0.0 en 1.0 – hoe hoger de score, hoe ernstiger de toxiciteit. Bovendien classificeert het algoritme verschillende subcategorieën van toxiciteit, zoals haatdragende uitingen, grof taalgebruik, seksisme, racisme, religieuze intolerantie, politiek extremisme, complottheorieën en bedreigingen. Deze subcategorieën worden nog eens aangevuld met specifieke tags voor antisemitisme, jihadisme, islamofobie, dreigingen met dood en geweld (Threats To Life of zogenoemde TTLs) en homo- en transfobie.



Figuur 2: Tekst met subcategorieën van toxiciteit gemarkeerd door het toxiciteits-algoritme

in prevalentie van antisemitisme daadwerkelijk veranderingen in online discourse weerspiegelen, en niet simpelweg de compositie van de dataset.

De basis voor dit algoritme ligt in **manueel geannoteerde lexicons** die meer dan 250.000 annotaties bevatten, geannoteerd door meer dan 150 menselijke experts. Deze annotaties omvatten triggerende en toxische woorden en woordcombinaties, elk voorzien van toxiciteitscores, categorielabels en contextuele informatie.

⁴¹ <https://plainsignal.com/glossary/data-normalization>

⁴² <https://datascientyst.com/pandas-random-sampling-stratified-and-weighted/>

In het kader van dit onderzoek hebben we het toxiciteits-algoritme op diverse manieren benut. We richtten ons niet alleen op de analyse van antisemitisme zelf, maar onderzochten ook de raakvlakken tussen antisemitisme en andere vormen van toxische content (correlatie en intersectionaliteit⁴³). Het toxiciteits-algoritme stelde ons in staat deze verschillende categorieën en hun onderlinge verbanden systematisch in kaart te brengen. **Bijzondere aandacht ging uit naar het identificeren van bedreigingen met dood en geweld** – zogenaamde TTLs (Threats To Life) – die de juridische grenzen overschrijden⁴⁴.

4.3 Antisemitisme Classifier

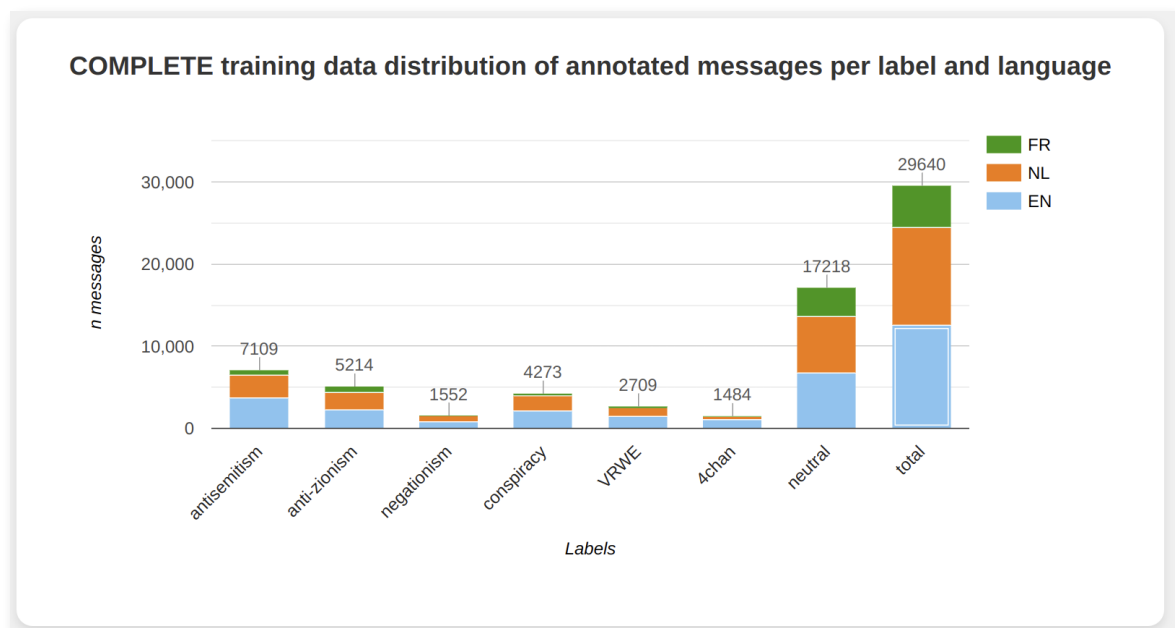
Het bovengenoemde toxiciteits-algoritme werkt op basis van lexicons en identificeert specifieke woorden en woordcombinaties, maar heeft in sommige gevallen moeite met het begrijpen van context. Woorden die op zichzelf toxisch zijn, hoeven niet noodzakelijk te leiden tot toxiciteit binnen de volledige context van een bericht. Voorbeelden hiervan zijn quotes, sarcas-

Daarnaast is het toxiciteits-algoritme niet specifiek ontwikkeld voor de detectie van antisemitisme, waardoor het geen volledig genuanceerd beeld geeft van de verschillende uitingsvormen hiervan. **Om zowel context als de diversiteit aan antisemitische uitingen beter te kunnen analyseren, hebben we daarom een volledig nieuw algoritme ontwikkeld dat specifiek gericht is op het detecteren van online antisemitisme** en dat naast het bestaande toxiciteits-algoritme kan worden ingezet.

Keuze model

Voor de automatische detectie van antisemitische content werd een zogenaamd BERT-model^{45,46,47} getraind. BERT-modellen zijn geavanceerde algoritmes die tekst kunnen begrijpen door de context van woorden te analyseren, in tegenstelling tot traditionele methoden die werken met deterministische regels en woordenlijsten. De keuze voor BERT boven meer recente modellen zoals ChatGPT werd ingegeven door praktische overwegingen: BERT-modellen zijn sneller en

Figuur 3: Distributie van complete trainingsdataset. Let op: teksten kunnen meer dan één label hebben, dus de som van de labels is niet gelijk aan het totaal.



me, humor en songteksten. Zo bevat de uitspraak “hij schold me uit voor ‘leugenachtige Jood’” weliswaar een toxische woordcombinatie, maar is het bericht als geheel niet toxisch omdat het een citaat betreft.

efficiënter bij het verwerken van grote hoeveelheden data, kunnen toxische content analyseren zonder moderatiebeperkingen, bieden meer privacy door

⁴³ <https://en.wikipedia.org/wiki/Intersectionality>

⁴⁴ Voor meer info over het toxiciteits-algoritme zie het technisch rapport en appendix 5 en 6

⁴⁵ <https://arxiv.org/abs/1706.03762>

⁴⁶ <https://huggingface.co/blog/bert-101>

⁴⁷ <https://medium.com/data-science/deconstructing-bert-part-2-visualizing-the-inner-workings-of-attention-60a16d86b5c1>

lokale verwerking, en zijn transparanter in hun besluitvorming.

Trainingsdataset en modelontwikkeling

Voor het trainen van het model werd een zorgvuldig samengestelde dataset van 29.640 berichten gebruikt, afkomstig uit historische data van de in paragraaf 4.1 genoemde sociale mediaplatformen tussen 2022 en 2024⁴⁸. Deze dataset omvatte berichten in het Nederlands (11.898), Engels (12.565) en Frans (5.177)⁴⁹.

Alle berichten werden handmatig geannoteerd via een tweeledig systeem:

1. Annotatie van de binaire labels ‘antisemitisme’ of ‘neutraal’.
2. In het geval van ‘antisemitisme’ werden eventuele sub- of aanverwante categorieën geannoteerd (zie figuur 4).

Om de data nog fijnmaziger te kunnen categoriseren, labelt het model **6 antisemitische sub- en aanverwante categorieën**, getoond in figuur 4.

De labels “Negationisme” en “TTL” stellen ons in staat om uitingen te identificeren die mogelijk strafbaar zijn. De labels “VRWE” en “4chan” geven meer inzicht in **ideologie-specifiek jargon**; aanhangers van ideologieën waarin antisemitisme structureel prevaleert. Aan de hand van het “anti-zionisme” label kunnen we legitieme kritiek op Israël en/of zionisme onderscheiden van **antisemitisme verhuld als anti-zionisme/Israël**⁵⁰ retoriek.

⁴⁸In de rapporten analyseren we data uit de jaren 2021 t/m 2024. Het AI model dat antisemitisme in tekstuele data detecteert (hier besproken) is getraind op data uit 2022 t/m 2024. Op basis van (voorbeeld) data uit 2022 t/m 2024 is het model ook in staat antisemitisme in data uit voorafgaande jaren te detecteren, aangezien deze data gelijkaardig is. Wel dient een dergelijk model regelmatig “hertraind” te worden op basis van nieuwe data om up to date te blijven.

⁴⁹Engels werd bijgevoegd, omdat online Engels vaak de voertaal is (zeker op platformen zoals 4chan en 9gag). Frans werd bijgevoegd, zodat het algoritme ook kan worden ingezet in de Belgische context. In de toekomst zouden we het taalbereik graag nog uitbreiden.

⁵⁰Zie hoofdstuk 2 voor een uitgebreide definitie van antisemitisme en het onderscheid met legitieme kritiek op Israël en zionisme

Sublabel	Omschrijving	Voorbeelden
Conspiracy	Antisemitische samenzweringstheorieën en gerelateerde terminologie.	'Joodse lobby', 'Joodse kinderbloeddrinkers', 'Soros elite'.
Negationisme	Ontkenning, ridiculisering, verheerlijking of bagatellisering van de Holocaust.	'de holocaust is een mythe', '6 gorillion', 'tijd voor een tweede holocaust'.
TTL (Threat to Life)	Bedreigingen met of oproepen tot geweld en/of dood.	'Total kike death', 'de enige goede jood is een dode jood'.
Anti-zionisme (en anti-Israël)	Uitingen die specifiek de staat Israël en/of het zionisme bekritisieren. In dit onderzoek maken we geen onderscheid tussen anti-zionistische en anti-Israël retoriek. Belangrijk is dat dergelijke uitingen in eerste instantie niet als antisemitisch worden geclassificeerd. Alleen wanneer een bericht zowel dit label als het 'antisemitisme' label krijgt, wordt het bericht als antisemitisch gerekend. Met behulp van dit label willen we inzicht krijgen in de mate van kritiek op Israël of zionisme die eveneens antisemitisch is.	'Israhel', 'zio-lobby', 'mossad psyop'.
VRWE (Gewelddadig Rechts-Extremisme)	Gewelddadig rechtsextremistisch jargon ⁵¹ . Veelal typisch neonazistische jargon, TTLs en nazistische termen uit WOII.	'alle joden aan het gas', 'lampenkappen maken van de joden', 'blut und bodem'
4chan	Jargon kenmerkend voor gebruikers van 4chan, vergelijkbare alternatieve fringe-platformen en de zogenaamde 'chan-cultuur' die berucht is voor het verspreiden van antisemitisme en andere toxische content ⁵² .	'lolocaust', 'goyslop', 'holohoax'

Figuur 4: Antisemitische sublabels gebruikt voor data annotatie

⁵¹ Voor meer info over VRWE zie appendix 4.

⁵² <https://gnet-research.org/2024/07/26/extremist-chan-culture/>

Modelprestaties

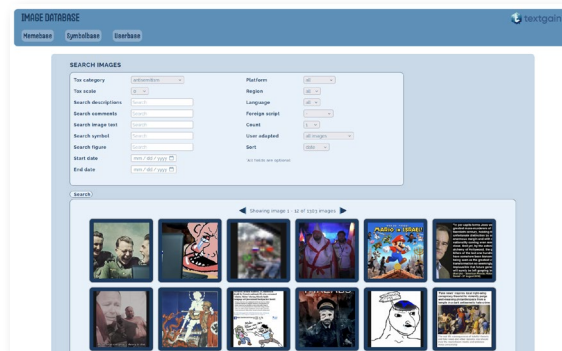
Het getrainde model toonde uitstekende prestaties in onafhankelijke tests op een aparte test set, met accuratie-scores van 0.95-0.99, precisie-scores van 0.90-0.95, recall-scores van 0.80-0.97 en f1-scores van 0.85-0.94⁵³ voor alle (sub)labels. Het model is in staat om:

- Contextueel begrip te tonen.
- Onderscheid te maken tussen neutrale teksten over de Joodse gemeenschap en antisemitische content.
- Verschillende vormen van antisemitisme te herkennen.
- Obscure antisemitische termen, jargon en slang te herkennen.
- Antisemitische en niet-antisemitische vormen van anti-Israël/anti-zionistische retoriek te onderscheiden.

De sterke prestaties gelden voor alle drie de talen waarin het model werd getraind (NL-EN-FR)⁵⁴.

4.4 Meme database en multimodale modellen

Antisemitisme wordt online niet alleen verspreid via tekst, maar ook via afbeeldingen, met name memes. Om dit fenomeen beter in kaart te brengen, hebben we daarom naast tekstueel materiaal ook honderden memes uit onze dataset geanalyseerd. Hiervoor ontwikkelden we een gebruiksvriendelijk dashboard waarmee gebruikers gemakkelijk door de meme-collectie kunnen navigeren.



Figuur 7: Screenshot van het dashboard om de meme database gemakkelijk te doorzoeken

Het in paragraaf 4.2 beschreven AI-model is enkel in staat antisemitisme in tekst te herkennen. Voor het identificeren en analyseren van antisemitische content in afbeeldingen hebben we andere gespecialiseerde technieken toegepast. Het doorzoeken en analyseren van grote beeldcollecties vereist namelijk andere methoden dan tekstanalyse. Hieronder lichten we de belangrijkste technieken toe die we hebben geïmplementeerd in onze meme-database aan de hand van de multimodale Mistral⁵⁵ (Pixtral) en Gemini⁵⁶ (Gemini Vision) modellen.

```
runs.summary["evaluation_report5"]
```

	class	precision	recall	f1	acc	...	mmc	roc_auc	hamming	...	pos_docs	neg_docs	correct_pos	correct_neg
1	antisemitism	0.94	0.86	0.9	0.95	...	0.87	0.92	0.05	...	1028	3296	883	3239
2	anti-zionism	0.94	0.86	0.9	0.97	...	0.88	0.92	0.03	...	770	3554	664	3510
3	negationism	0.92	0.89	0.91	0.99	...	0.9	0.94	0.01	...	233	4091	208	4074
4	conspiracy	0.92	0.85	0.88	0.97	...	0.86	0.92	0.03	...	632	3692	536	3644
5	VRWE	0.9	0.8	0.85	0.97	...	0.83	0.9	0.03	...	403	3921	322	3886
6	4chan	0.95	0.86	0.9	0.99	...	0.9	0.93	0.01	...	215	4109	184	4100
-	neutral	0.91	0.97	0.94	0.93	...	0.85	0.92	0.07	...	2493	1831	2423	1585

Figuur 5: Testresultaten van de gehele test set

⁵³ <https://medium.com/data-science/accuracy-precision-recall-or-f1-331fb37c5cb9>

⁵⁴ Voor meer info bij het antisemitisme algoritme zie het uitgebreide technische rapport

⁵⁵ <https://mistral.ai/news/pixtral-12b>

⁵⁶ <https://cloud.google.com/vision?hl=en>

Figuur 6:
Voorbeelden
van hoe het
uiteindelijk
antisemitisme
model teksten
labelt

```

6 miljoen joden is overdreven, de holocaust leugen
{'antisemitism': 1.0, 'anti-zionism': 0.0, 'negationism': 1.0, 'conspiracy': 1.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 0.0}
-----
we herdenken de 6 miljoen slachtoffers van de holocaust. Dat dit nooit meer mag gebeuren
{'antisemitism': 0.0, 'anti-zionism': 0.0, 'negationism': 0.0, 'conspiracy': 0.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 1.0}
-----
er wonen veel joden in Amsterdam
{'antisemitism': 0.0, 'anti-zionism': 0.0, 'negationism': 0.0, 'conspiracy': 0.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 1.0}
-----
de holocaust was een verschrikkelijke gebeurtenis
{'antisemitism': 0.0, 'anti-zionism': 0.0, 'negationism': 0.0, 'conspiracy': 0.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 1.0}
-----
de holocaust is nooit gebeurd
{'antisemitism': 1.0, 'anti-zionism': 0.0, 'negationism': 1.0, 'conspiracy': 1.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 0.0}
-----
6 million jews didn't die
{'antisemitism': 1.0, 'anti-zionism': 0.0, 'negationism': 1.0, 'conspiracy': 1.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 0.0}
-----
we commemorate the 6 million jews who died during WWII
{'antisemitism': 0.0, 'anti-zionism': 0.0, 'negationism': 0.0, 'conspiracy': 0.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 1.0}
-----
lots of jews live in New York
{'antisemitism': 0.0, 'anti-zionism': 0.0, 'negationism': 0.0, 'conspiracy': 0.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 1.0}
-----
Jewish pig
{'antisemitism': 1.0, 'anti-zionism': 0.0, 'negationism': 0.0, 'conspiracy': 0.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 0.0}
-----
#freepalestine
{'antisemitism': 0.0, 'anti-zionism': 1.0, 'negationism': 0.0, 'conspiracy': 0.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 0.0}
-----
de Joodse monsters plegen genocide in palestina
{'antisemitism': 1.0, 'anti-zionism': 1.0, 'negationism': 0.0, 'conspiracy': 0.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 0.0}
-----
verboden voor joden
{'antisemitism': 1.0, 'anti-zionism': 0.0, 'negationism': 0.0, 'conspiracy': 0.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 0.0}
-----
antisemitisme is terug van weggeweest, "verboden voor joden" is blijkbaar het nieuwe normaal
{'antisemitism': 0.0, 'anti-zionism': 0.0, 'negationism': 0.0, 'conspiracy': 0.0, 'VRWE': 0.0, '4chan': 0.0, 'neutraal': 1.0}
-----

```



Figuur 8: Screenshot van hoe Pixtral een toxiciteitslabel en beschrijving aan een antisemitische meme toevoegt. Uit de beschrijving blijkt dat Pixtral ook in staat is de tekst en symboliek in de afbeelding te begrijpen.

- **Image classification and tagging** zijn technieken waarbij AI-systemen automatisch afbeeldingen analyseren om te bepalen wat erop staat en hier labels (tags) aan toe te kennen. Bijvoorbeeld 'antisemitisme'.
- **Object recognition** is een AI-technologie die specifieke voorwerpen of objecten in afbeeldingen of video's kan identificeren en lokaliseren. Bijvoorbeeld antisemitische of nazistische figuren en symbolen.



Figuur 9: Screenshot van similarity search voor antisemitische memes met de 'Happy Merchant' trope

- **Automatic image captioning** is een AI-technologie die automatisch tekstbeschrijvingen genereert voor afbeeldingen.
- **OCR (Optical Character Recognition)** is een technologie die tekst in afbeeldingen of gescande documenten herkent en extraheert naar digitale tekst die een computer kan bewerken en doorzoeken. Een afbeelding op zich hoeft immers niet antisemitisch te zijn, maar de bijgaande tekst of de combinatie van tekst en beeld is dat misschien wel⁵⁷.

⁵⁷ Hiervoor gebruikten we eveneens de Python Library pytesseract: <https://pypi.org/project/pytesseract/>

- **Similarity search & clustering-algoritmes**
groeperen visueel vergelijkbare afbeeldingen automatisch samen op basis van gemeenschappelijke kenmerken zoals kleur, vorm, inhoud of stijl.
- Dit maakt het mogelijk om grote verzamelingen afbeeldingen te organiseren in thematische clusters zonder vooraf gedefinieerde categorieën. Bijvoorbeeld 'samenzweringstheorieën', 'nazi propaganda', 'TTLs'⁵⁸.

⁵⁸ Hiervoor gebruikten we eveneens de Blip-2 modellen:
https://huggingface.co/docs/transformers/model_doc/blip-2

5. Kwantitatieve big data analyse: kanttekeningen

Bij het interpreteren van de resultaten van dit onderzoek is het belangrijk enkele **inherente beperkingen** van kwantitatieve big data analyse te begrijpen. Deze beperkingen doen geen afbreuk aan de robuustheid of betrouwbaarheid van onze methodologie, maar bieden context voor een genuanceerde interpretatie van de bevindingen.

Ten eerste is het bij de analyse van sociale media **onmogelijk om werkelijk *alle* relevante content te analyseren**. Onder andere omdat berichten voortdurend worden aangepast of verwijderd en sommige discussies plaatsvinden in privé groepen. Daarnaast vindt onze dataverzameling plaats via steekproeven op gezette momenten. Dit betekent dat onze dataset, hoewel omvangrijk, altijd een deel vertegenwoordigt van het grotere geheel.

Vergelijkende analyses tussen verschillende sociale media platformen bieden waardevolle inzichten, maar moeten worden gezien als benaderingen. Ieder platform heeft unieke eigenschappen, verschillende gebruikersdemografieën en varieert in toegankelijkheid voor onderzoekers. Zo kunnen we bijvoorbeeld meer data en meer gedetailleerde data verzamelen van platform A dan van platform B, wat directe één-op-één vergelijkingen bemoeilijkt. Vergelijkende analyses tussen platformen moeten dus enigszins worden genuanceerd.

Bij het classificeren van content maken we gebruik van geavanceerde algoritmes en AI-systemen. **Hoewel deze nooit 100% nauwkeurig zijn**, vormen ze een essentieel instrument voor het analyseren van grote hoeveelheden data. Een beperkt aantal valse positieven (data die onterecht wordt geclassificeerd als problematisch) en valse negatieven (problematische data die wordt gemist) is onvermijdelijk. Door zorgvuldige validatie, regelmatige verfijning van onze modellen en het hanteren van statistische correcties, minimaliseren we deze afwijkingen en zorgen we voor betrouwbare resultaten op grote schaal. Iets wat met handmatige analyse van dergelijke grote datasets nooit mogelijk zou zijn.

Ondanks deze beperkingen biedt kwantitatieve big data analyse onmisbare inzichten die met traditionele onderzoeksmethoden niet te verkrijgen zijn. Het stelt ons in staat om patronen te identificeren, trends over tijd te volgen en verbanden te leggen die anders onzichtbaar blijven.

Colofon

Titel

Antisemitisme in het Nederlandse sociale media landschap:
Methodologisch document

Publicatiedatum

Mei 2026

Opdrachtgever

Centrum Informatie en Documentatie Israël (CIDI) *in casu* Stichting Informatie en Meldpunt Antisemitisme (meldantisemitisme.nl)

Uitvoering en onderzoek

Textgain

Financiële ondersteuning

De onderzoeken zijn mogelijk gemaakt door financiële ondersteuning van de Nationaal Coördinator Antisemitismebestrijding (NCAB).

Methodologie

Dit rapport is gebaseerd op de analyse van publiek toegankelijke online bronnen, waaronder sociale mediaplatformen, discussiefora en andere openbare digitale omgevingen. Voor de detectie en analyse van online antisemitische uitingen werd een combinatie ingezet van AI- en NLP-technieken, meertalige lexicons, contextuele analyse en menselijke validatie, gericht op het in kaart brengen van trends, narratieven, verspreidingsdynamieken en taalgebruik binnen online discussies.

Juridische en ethische disclaimer

De analyses in dit rapport zijn gebaseerd op geautomatiseerde detectie en interpretatie van publieke online content. Hoewel maximale zorgvuldigheid is betracht, kunnen classificaties en interpretaties onderhevig zijn aan contextuele beperkingen. Het rapport heeft als doel maatschappelijke trends en risico's inzichtelijk te maken en vormt geen juridische kwalificatie van individuele uitingen of personen. Persoonsgegevens worden waar mogelijk geaggregeerd, gepseudonimiseerd of geanonimiseerd verwerkt conform de AVG/GDPR en relevante Europese regelgeving.

Vormgeving

Luna 3

Contact

cidi.nl of meldantisemitisme.nl
textgain.com

Copyright

© 2026 Textgain.
Alle rechten voorbehouden.

Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming, tenzij wettelijk anders is toegestaan.



Online antisemitisme: hoe staat het ervoor?

Sociale media zijn uitgegroeid tot een belangrijke plek waar antisemitische denkbeelden, complottheorieën en haat zich verspreiden. Om een volledig beeld te krijgen van antisemitisme in Nederland liet CIDI, samen met het Meldpunt Antisemitisme en de NCAB, onderzoek uitvoeren door Textgain naar het Nederlandse online landschap.

Deze rapportages brengen de belangrijkste trends, narratieven en ontwikkelingen op Nederlandse sociale media van 2021 tot 2024 in kaart.

Het onderzoek vormt onderdeel van een monitor die de ontwikkeling van online antisemitisme door de jaren heen inzichtelijk moet maken.

Alleen door online en offline ontwikkelingen structureel te volgen, kunnen we antisemitisme beter begrijpen, herkennen en bestrijden.

CIDI meld.
antisemitisme

 **textgain**


Nationaal Coördinator
Antisemitismebestrijding
Ministerie van Justitie en Veiligheid