

Onderzoeksverantwoording

De introductie van generatieve AI als ChatGPT en Claude maakt het mogelijk om in korte tijd overtuigende (of overtuigend ogende) teksten te schrijven. Software die AI-gebruik moet detecteren was lange tijd onbetrouwbaar, maar sinds kort is er een zeer goede AI-scanner beschikbaar. Investico en *De Groene Amsterdammer* onderzochten daarmee AI-gebruik bij het schrijven van spreekteksten in de Tweede Kamer – al hanteerden we daarbij strenge normen.

Onderzoeksopzet

We hebben besloten ons te richten op het soort debatteksten waarvan de kans het grootst is dat Kamerleden er AI voor gebruiken: de spreektekst in de eerste inbreng. Die teksten vormen de eerste aanzet van het debat, en bereiden Kamerleden van tevoren voor.

We verzamelden hiervoor alle commissie- en plenaire debatten die dit jaar zijn gevoerd. We keken alleen maar naar de inbreng van de Kamer in de eerste termijn van het debat. In de tweede termijn worden vooral moties ingediend, die we minder interessant vonden. Tweeminutendebatten hebben we om dezelfde reden niet meegenomen. We keken dus ook niet naar de inbreng van het kabinet, omdat Kamerleden de teksten die ze daar uitspreken niet kunnen voorbereiden. Ten slotte sloegen we procedurele aspecten van de in plenaire sessies als de mededelingen en regeling van werkzaamheden over.

Om de Kamerdebatten te analyseren hebben gebruikt gemaakt van opentk.nl, de website waarop Bert Hubert alles wat de Tweede Kamer publiceert doorzoekbaar maakt. Hij deelde ook een database met ons met daarin alle debatten, met per spreker genoteerd of dat moment spreker (dus degene die aan het woord is) of interrumpant was. Met die database konden we geautomatiseerd op een rijtje zetten welke sprekers er allemaal meededen aan elk debat. We zijn Bert Hubert zeer dankbaar voor zijn hulp.

Vervolgens was het zaak om uit al die debatten de eerste inbreng van alle Kamerleden hieruit te filteren. Sprekers worden nogal eens geïnterrupteerd en vervuilen op die manier de data. Met behulp van de sprekersdatabase konden we automatisch de spreekteksten inladen die niet geïnterrupteerd werden, maar de andere spreekteksten moesten we handmatig kopiëren.

Daartoe kopieerden we de bijdrage totdat er een interruptie komt, dat interruptiedebat sloegen we vervolgens over, totdat de spreker weer doorgaat met diens bijdrage, wat vaak (maar niet altijd) wordt gemarkeerd doordat de voorzitter zegt 'U vervolgt uw betoog'. Investico-collega's Jaco Wilschut en Sofyan El Bouchtili staken beiden veel tijd in dit monnikenwerk, we zijn hen daar zeer dankbaar voor.

Op deze manier verzamelden we de spreekteksten van 303 commissie- en plenaire debatten tussen 1 januari en 30 september 2026. In totaal komt dat neer op 2715 debatbijdragen.

Pangram

Om te controleren of deze bijdragen (deels) met AI waren geschreven, gebruikten we AI-scanner Pangram. Pangram komt tot een oordeel door een tekst te vergelijken met een enorme database van menselijk geschreven en AI-gegenereerde teksten. Het is zelf óók een AI-model. Mensen kunnen AI-teksten soms herkennen door overmatig gebruik van typische stijlfiguren als drieslagen, of gedachtestreepjes (meer hierover later), maar Pangram analyseert ook veel zwakkere signalen die door een tekst heen lopen, zoals zinsritme, woordkeuze en zinslengte.

In [deze podcast](#) van de *New York Times* legt Pangram-CEO Max Spero uit dat Pangram vooral kijkt naar een bepaalde voorspelbaarheid door een heel document heen. 'Een AI-model neemt beslissingen op een bepaalde, consistente manier, terwijl mensen een veel bredere, minder 'mode-collapsed' beslisboom hebben.' Meer over de werking van Pangram is te lezen in [deze Q&A](#) met Spero, en op de [website](#) van Pangram zelf.

Pangram is betrouwbaar, blijkt uit onafhankelijk wetenschappelijk onderzoek. [Wetenschappers van de Universiteit van Chicago concluderen](#) bijvoorbeeld dat Pangram geen valspositieven gaf voor de menselijke teksten die ze de detector voorschotelden. Investico sprak een van de auteurs, die aangeeft nog steeds achter de conclusies te staan. Eerder dit jaar vonden ook [onderzoekers van de Universiteit van Brussel](#) dat Pangram geen valspositieven gaf, toen ze de prestaties van verschillende AI-scanners op wetenschappelijke papers onderzochten.

Pangram-CEO Spero benadrukt dat het laag houden van het aantal valspositieven essentieel is: liever detecteert hij sommige AI-teksten niet, dan dat hij een menselijke tekst als AI aanmerkt. Het aantal valspositieven is volgens Pangram zelf ongeveer één op tienduizend.

Hoe langer de tekst is, hoe beter Pangram werkt. Het minimumaantal woorden is 50, maar Kamerdebatten zijn gemiddeld zo'n 800 woorden lang. Daarmee zal het programma nog betrouwbaarder zijn bij spreekteksten in Kamerdebatten dan bijvoorbeeld bij uitingen op sociale media.

Om zelf ook nog een controlestap in te bouwen, verzamelden we ook willekeurig debatten uit de periode tussen 2019 en 2023, toen generatieve AI nog niet bestond, of nog niet geavanceerd genoeg was om dit soort teksten te genereren. 150 van deze debatteksten uit

2019 en 2021 analyseerden we met Pangram, en werden als 100 procent door mensen geschreven beoordeeld.

Alle debatteksten van dit jaar verzamelden we in een spreadsheet van Google Sheets. Vervolgens stuurden we die met een Python-script geautomatiseerd naar de API van Pangram. Pangram stuurt vervolgens de resultaten terug: een percentage van de tekst dat naar het oordeel van de software met AI is geschreven, om welke stukken tekst het gaat, en of het 'hoog', 'medium' of 'laag' vertrouwen had in het oordeel.

Pangram merkt een tekst als 'deels door AI-geschreven' aan als meer dan 10 procent van de tekst AI-gegenereerd is. Wij besloten een strengere grens aan te houden: voor dit onderzoek telt een tekst als 'geschreven met AI' wanneer meer dan de helft AI-gegenereerd is. Bovendien namen we alleen teksten mee waar Pangram een 'hoog' vertrouwen had in het oordeel, al is dit vrijwel altijd het geval bij teksten die meer dan 50 procent met AI zijn geschreven.

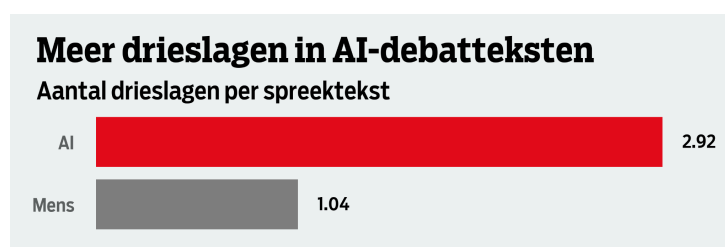
Taalanalyse

We waren ook benieuwd naar hoe de taal verandert met de komst van de chatbots. Duidt het taalgebruik in de teksten die Pangram als AI aanmerkt ook op AI-gebruik? Daartoe analyseerden we deze teksten op twee verschillende manieren.

Allereerst telden we (geautomatiseerd) hoe vaak bepaalde typische AI-stijlfiguren voorkomen. Lees bijvoorbeeld op [deze Wikipedia-pagina](#) meer over tekenen van AI-geschreven teksten. We richtten ons op twee stijlfiguren: de drieslag, en negatieve parallelismen: zinnen van de vorm 'Niet X, maar Y' (en 'niet alleen X, maar ook Y', 'geen X, maar Y', et cetera).

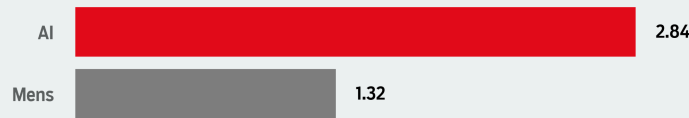
We telden deze stijlfiguren door de teksten te analyseren met zogenoemde reguliere expressies: combinaties van voorwaarden om bepaalde patronen in teksten te vinden. Zeker voor de negatieve parallelismen moesten we die in een aantal stappen perfectioneren, om te zorgen dat de juiste stijlfiguren werden aangemerkt.

Het blijkt dat beide stijlfiguren vaker voorkomen in AI-teksten. Het aantal drieslagen was bijna drie keer zo groot in de AI-bijdragen (1,04 per tekst voor teksten zonder AI, 2,92 per tekst voor teksten die meer dan 50% AI zijn). Het aantal negatieve parallelismen is ruim twee keer zo groot in de AI-bijdragen (1,32 per tekst voor teksten zonder AI, 2,84 per tekst voor teksten die meer dan 50% AI zijn).



Niet X, maar Y

Aantal constructies "Niet X, maar Y" per spreektekst



Daarnaast waren we benieuwd of de AI-teksten andere woorden, of combinaties van woorden bevatten. Daartoe gebruikten we Python-package SpaCy om de teksten om te zetten in lijsten lemma's, zodat verschillende werkwoordsvormen bijvoorbeeld als hetzelfde woord worden geteld. Vervolgens keken we welke woorden relatief meer voorkomen in de AI-teksten en welke relatief meer in de 'menselijke' teksten.

De AI-teksten bevatten significant meer 'verbindingswoorden' als 'daarom', 'juist', 'daadwerkelijk' en 'tegelijkertijd'. Al deze woorden komen gemiddeld meer dan twee keer vaker voor in AI-teksten. De menselijke teksten bevatten juist significant meer 'vulwoorden' als 'heel', 'nou', 'graag', 'eigenlijk', en 'natuurlijk'.

Welke woorden komen vaker voor in AI-teksten?

AI

erkennen **stelsel**
papier concreet
helder **daadwerkelijk**
keuze **daarom**
raken juist digitaal
tegelijkertijd blijven
structureel

Mens

helemaal **benieuwd**
eens **eigenlijk** veel
toch **graag nou** geval
motie **heel** aantal
erg **denken** vorig
jaar **allemaal** collega
natuurlijk bijvoorbeeld

Om zeker te weten dat we niet een ander effect zagen, maar de AI-teksten daadwerkelijk anders waren, hebben we de taalanalyse ook uitgevoerd op de teksten van tussen 2019 en 2023 die we als controle hadden verzameld. Het blijkt dat die teksten veel meer op de 'menselijke' teksten uit 2026 lijken, dan op de AI-teksten: het aantal 'Niet X, maar Y'-constructies is met 1,28 per tekst vrijwel hetzelfde, en het aantal drieslagen is met 0,91 per tekst zelfs nog wat lager. Dit is een extra ondersteuning voor het feit dat de taal in de AI-teksten anders is dan in de menselijke teksten.

De Python-code om de teksten te analyseren schreven we met behulp van het programma Cursor. Dat is een Python-omgeving met daarin een AI-agent, die de code kan schrijven en testen. Alle door Cursor geschreven code is gecontroleerd door onszelf, en door Groene-journalist en datawetenschapper Joris Veerbeek.

Resultaten

Van de 2715 geanalyseerde spreeksten zijn er volgens Pangram 297 meer dan 50% AI-gegeneerd. Dat is ongeveer 11 procent van alle teksten. De verschillen tussen fracties en tussen Kamerleden zijn groot. Zie hieronder voor een overzicht van het AI-gebruik per fractie.

Fractie	Aantal spreekteksten	Aantal AI	Percentage AI
BBB	136	44	32,4%
Groep Markuszower	144	35	24,3%
Lid Keijzer	32	7	21,9%
VVD	298	49	16,4%
PVV	231	35	15,2%
CDA	294	43	14,6%
JA21	245	32	13,1%
DENK	76	7	9,2%
D66	305	23	7,5%
Pro	300	12	4%
FVD	124	4	3,2%
ChristenUnie	124	2	1,6%
50PLUS	65	1	1,5%
SGP	134	2	1,5%
PvdD	59	0	0%
SP	109	0	0%
Volt	34	0	0%

De verschillen tussen Kamerleden onderling zijn misschien wel relevanter dan de verschillen tussen fracties. Waar PVV'ers Rachel van Meetelen en Annette Raijer bijvoorbeeld respectievelijk 10 en 12 keer AI-teksten uitspreken, gaat de AI-scanner nooit af bij PVV-prominenten als Geert Wilders, Dion Graus en Martin Bosma.

Dit zijn de namen van de 47 Tweede Kamerleden die dit jaar minimaal twee debatteksten uitspraken die voor meer dan de helft uit AI-tekst bestonden

De Kamerleden die AI gebruikten voor meer dan de helft van hun bijdragen

1. Esmah Lahlah (Pro)
2. Rachel van Meetelen (PVV)
3. Fatimazhra Belhirsch (D66)
4. Wendy van Eijk (VVD)
5. Harmen Krul (CDA)
6. Dion Huidekooper (D66)

7. Bart Bikkers (VVD)
8. Annette Raijer (PVV)
9. Nicole Moinat (Groep Markuszower)
10. Ani Zalinyan (Pro)
11. René Claassen (Groep Markuszower)
12. Martin de Beer (VVD)

De Kamerleden die AI gebruikten voor meer dan 25 procent van hun bijdragen

13. Chris Jansen (PVV)
14. Emiel van Dijk (PVV)
15. Daniël van den Berg (JA21)
16. Peter de Groot (VVD)
17. Ruud Verkuijlen (VVD)
18. Henk Vermeer (BBB)
19. Ulaş Köse (D66)
20. Ismail el Abassi (Denk)
21. Caroline van der Plas (BBB)
22. Michiel Hoogeveen (JA21)
23. Elles van Ark (CDA)
24. Maes van Lanschot (CDA)
25. Alisha Müller (VVD)
26. Ingrid Coenradie (JA21)
27. Erik van der Maas (VVD)

De Kamerleden die AI gebruikten, maar in minder dan 25 procent van hun bijdragen

28. Femke Wiersma (BBB)
29. Hanneke Steen (CDA)
30. Jantine Zwinkels (CDA)
31. Jan Schoonis (D66)
32. Mona Keijzer (Lid Keijzer)
33. Maikel Boon (PVV)
34. Tijs van den Brink (CDA)
35. Etkin Armut (CDA)
36. Peter van Duijvenvoorde (FvD)
37. Nicole Maes (VVD)
38. Wim Meulenkamp (VVD)
39. Joris Lohman (CDA)
40. Erwin Prickaertz (PVV)
41. Mpanzu Bamenga (D66)
42. Annabel Nanninga (JA21)
43. Shanna Schilder (Groep Markuszower)
44. Tamara ten Hove (Groep Markuszower)
45. Inge van Dijk (CDA)

46. Hidde Heutink (Groep Markuszower)

47. Mirjam Bikker (ChristenUnie)

Wederhoor

Het is belangrijk om op te merken dat Pangram alleen maar analyseert of de tekst uit een chatbot is gerold. Wat gebruikers als input leveren in die chatbot, dat weten we natuurlijk niet. Het zou kunnen dat een Kamerlid vanuit het niets een tekst laat schrijven ('Schrijf een tekst die in 4 minuten past over kernenergie, vanuit een progressief uitpunt'), maar het kan ook dat die gebruiker de tekst bijvoorbeeld laat inkorten ('Dit is een speech van 2000 woorden over kernenergie, kort die in, met behoud van de inhoud, zodat die in 4 minuten past'). Wat we uiteindelijk (vrijwel) zeker weten is dat de Kamerleden de teksten direct vanuit de chatbot kopieerden en uitspraken in het debat.

Daarom vonden we het van groot belang om alle Kamerleden die een AI-tekst uitspraken de gelegenheid te geven om daar uitleg bij te geven. De 12 Kamerleden die meer dan de helft van hun teksten met AI schreven spraken we zelf, of benaderden we meermaals voor uitleg. Daarnaast stuurden we de Kamerfracties een overzicht van alle Kamerleden die AI-teksten voorlezen, met de vraag of ze daar vanuit de fractie, of als Kamerlid zelf op wilden reageren. Zie de website van De Groene en Investico voor een overzicht van de reacties van alle fracties.

Hoewel het aantal valspositieven laag is, is er een kans dat er op ruim 2700 bijdragen er wellicht eentje als AI wordt aangemerkt. Om dat risico weg te nemen, besloten we alleen de namen van Kamerleden te publiceren die minimaal twee AI-teksten uitspraken. De kans dat meer dan één speech onterecht door Pangram als AI wordt aangemerkt is nihil.