

2023Z10342

Vragen van het lid **Van Raan** (PvdD) aan de Staatssecretaris van Binnenlandse Zaken en Koninkrijksrelaties en de Minister voor Rechtsbescherming over *het aan banden leggen van ChatGPT vanwege privacyzorgen* (ingezonden 8 juni 2023).

Vraag 1

Kent u het artikel «Dat zijn toch gewoon ál onze artikelen?»¹

Vraag 2

Kunt u bevestigen dat ChatGPT zijn taalmodel traint op basis van miljoenen onrechtmatig verkregen documenten van Docplayer.nl?

Vraag 3

Klopt het dat die documenten vol staan met onder andere BSN-nummers, persoonlijke belastingaangiftes, curricula vitae en andere persoonsgegevens?

Vraag 4

Deelt u de stelling uit het artikel: «De persoonlijke informatie van docplayer.nl was op de website zélf al in strijd met de wet, laat staan wanneer die ook nog eens in chatbots wordt verwerkt»? Zo nee, waarom niet?

Vraag 5

Klopt het dat ChatGPT hiermee onrechtmatig en onwettig persoonsgegevens verwerkt?

Vraag 6

Hoe oordeelt u daarover? Wat gaat u er aan doen?

¹ <https://www.groene.nl/artikel/dat-zijn-toch-gewoon-al-onze-artikelen>; De Groene Amsterdammer d.d. 7 juni 2023.

Vraag 7

Klopt het dat ChatGPT voor zijn taalmodel ongeveer net zoveel geleerd heeft van de neonazistische website Stormfront als van de website van RTL Nieuws?²

Vraag 8

Hoe oordeelt u daarover? Wat gaat u er aan doen?

Vraag 9

Klopt het dat ChatGPT gebruik heeft gemaakt van onder andere 594.000 NRC artikelen en 162.000 Volkskrant artikelen?

Vraag 10

Deelt u de mening dat ChatGPT hiermee inbreuk maakt op het auteursrecht?

Vraag 11

Hoe oordeelt u daarover? Wat gaat u er aan doen?

Vraag 12

Klopt het dat het «kwaliteitsfilter» is gebaseerd op drie bronnen waarvan er twee (Wikipedia en Reddit) een zeer sterke oververtegenwoordiging van mannelijke input kennen?³

Vraag 13

Hoe oordeelt u daarover? Wat gaat u er aan doen?

Vraag 14

Deelt u de mening dat de Autoriteit Persoonsgegevens zo snel mogelijk met een spoedoordeel moet komen over ChatGPT en het blokkeren ervan geen taboe is? Kunt u de Autoriteit Persoonsgegevens daartoe aansporen?

Vraag 15

Kunt u deze vragenset zo snel als mogelijk beantwoorden?

² «In de top-tweehonderd van meest geciteerde websites vonden we Wikipedia en ongeveer elke grote Nederlandse krant, en ook de neonazistische complotwebsite Stormfront. Die laatste staat maar één plek lager in de bronnenlijst dan RTL Nieuws. Van beide websites leert AI dus ongeveer evenveel.»

³ 90% van de teksten op Wikipedia is geschreven door mannen, 74% van de Reddit gebruikers is man.