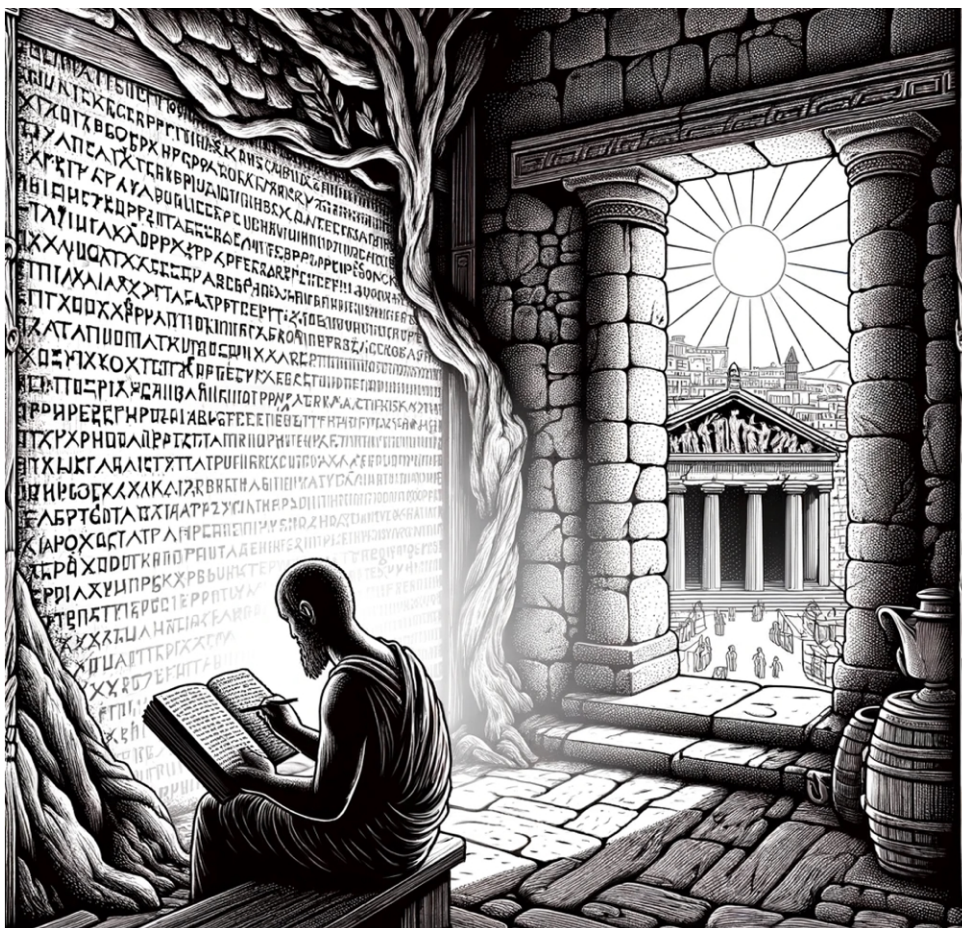


KANTL - colloquium ChatGPT

Op vrijdag 21 juni 2024 vindt in de KANTL een studiedag plaats over Generatieve AI voor het Nederlands en de neerlandistiek.

Het doel van de bijeenkomst is om te reflecteren over de gevolgen van generatieve AI voor het onderwijs, voor de taalkunde, voor het beeld dat we hebben over creatieve taalproductie en over het belang van de constructie van taalmodellen voor het Nederlands.
De ethische aspecten zullen daarbij niet buiten beeld blijven.

Hierbij kunt u het programma terugvinden alsook een korte toelichting bij alle verschillende onderwerpen en een beknopte biografie bij iedere spreker.



PROGRAMMA

12.30-13.00

Onthaal

13.00-13.30

Opening: KANTL-voorzitter Leen van Dijck

MOETEN WE BANG ZIJN VOOR FILOSOFISCHE ZOMBIES?

(Walter Daelemans, Universiteit Antwerpen)

13.30-14.00

DE TEKSTMACHINE EN DE TAALTHEORIE

(Dirk Geeraerts, KU Leuven)

14.00-14.30

BLINKENDE KNOPKE FIJN, DAT NIMMER VAN SCHRIJVEN ZIJT MOE? OVER HET CREATIEVE VERMOGEN VAN GROTE TAALMODELLEN

(Tim Van De Cruys, KU Leuven)

14.30-15.00

Koffiepauze

15.00-15.45

VERANTWOORD INZETTEN VAN GENAI IN ACADEMISCH SCHRIJFONDERWIJS: EEN ILLUSIE?

(Lieve De Wachter, KU Leuven)

15.45-16.15

INTELLIGENTE TECHNOLOGIE: HOE STAAT HET MET NEDERLANDS?

(Frieda Steurs, Instituut voor de Nederlandse Taal)

16.15-16.45

DE GEBRUIKSVORWAARDEN VAN TAAL

(Juliet van Rosendaal, KU Leuven)

16.45-17.30

Publieksdiscussie (Walter Daelemans, moderator)

17.30-18.30

Receptie

TOELICHTING & BIO'S

WALTER DAELEMANS is gewoon hoogleraar computertaalkunde (UAntwerpen), verbonden aan de CLiPS onderzoeksgroep en de UAntwerpen kernfaciliteit voor *text mining* (TEXTUA). Hij doet onderzoek over zelflerende taalmodellen en hun toepassingen in de taaltechnologie en doceert in de internationale MA *Digital Text Analysis*.



Moeten we bang zijn voor filosofische zombies?

Grote taalmodellen zoals ChatGPT hebben duidelijk veel beperkingen, zeker wanneer je ze voor het Nederlands gebruikt, maar ze blijven ook voortdurend positief verrassen. Met een eenvoudig model als kapstok laat ik zien waar die tegenstrijdige waarnemingen vandaan komen en waarom we snel te maken zullen krijgen met niet één maar een heel leger van “filosofische zombies”.

DIRK GEERAERTS is emeritus gewoon hoogleraar in de algemene taalkunde (KU Leuven). Hij is lid van de KANTL.



De tekstmachine en de taaltheorie

In het taalkundige onderzoek gaat sinds het midden van de vorige eeuw veel aandacht naar de menselijke taalverwerking. Moeten we ervan uitgaan dat de kern van ons taalvermogen aangeboren is? Dat is de traditionele Chomskyaanse visie van de moderne taalkunde. Of leren we vooral door patronen te herkennen in het taalaanbod uit onze omgeving? Dat is de recentere, 'taalgebruiksgebaseerde' opvatting. ChatGPT lijkt die tweede benadering gelijk te geven: met alleen bestaand taalgebruik als input bouwt het systeem een verbluffend geavanceerde taalvaardigheid op. Zo eenvoudig is het echter niet: deze lezing zal ingaan op de vraag waarom, en daar enkele conclusies aan verbinden m.b.t. de integratie van ChatGPT in taalkundig onderzoek.

TIM VAN DE CRUYS is hoofddocent computerlinguïstiek aan de faculteit Letteren, KU Leuven. Hij doet onderzoek naar automatische taalverwerking, met een focus op het modelleren van betekenis en creatieve taalgeneratie. Hij is de houder van een ERC Consolidator Grant - TENACITY - waarbinnen onderzoek gedaan wordt naar innovatieve algoritmes voor computationele creativiteit.



Blinkende knopke fijn, dat nimmer van schrijven zijt moe? Over het creatieve vermogen van grote taalmodellen.

Vaak wordt beweerd dat AI-taalmodellen een zeker creatief vermogen tentoonspreiden. Dat staat evenwel in schril contrast met de manier waarop de modellen getraind zijn. De huidige AI-taalmodellen zijn uitsluitend taakgericht: een neurale netwerk wordt getraind op een aanzienlijke hoeveelheid trainingdata, waarbij de parameters van het model worden geoptimaliseerd om de best mogelijke uitvoer te genereren op basis van die trainingdata. Op die manier bootst het netwerk menselijk taalgebruik enkel na en laat het weinig ruimte voor creativiteit.

Een simpele opdracht om een gedicht te schrijven – zonder creatieve invoer van de gebruiker zelf – zal dan ook vaak een saai en clichématig resultaat opleveren. Zelfs al zou er een groot aantal gedichten in de trainingdata vervat zitten, dan nog zal het model de data waarop het getraind is zo goed mogelijk proberen te reproduceren – iets wat niet strookt met ons idee van creativiteit. Om een zekere vorm van creativiteit aan een taalmodel te ontlokken, zullen we paradoxaal genoeg de mogelijkheden van het model moeten inperken, zodat het aangemaand wordt om nieuwe manieren te zoeken om zichzelf uit te drukken.

LIEVE DE WACHTER is gewoon hoogleraar Nederlandse Taalbeheersing en Academisch Nederlands aan de KU Leuven. Ze is lid van de Onderzoeksgroep Taal en Onderwijs. Daarnaast is ze hoofd van het Schrijfcentrum van de KU Leuven. Haar onderzoek focust op academische schrijf- en spreekvaardigheid en op talig testen.



Verantwoord inzetten van GenAI in academisch schrijfonderwijs: een illusie?

In deze sessie komen de mogelijkheden en uitdagingen van het verantwoordelijk inzetten van generatieve AI (zoals ChatGPT) in het hoger schrijfonderwijs aan bod. Vanuit ervaringen aan de KU Leuven, illustreert deze presentatie door middel van concrete voorbeelden hoe GenAI een rol kan spelen bij academisch schrijven. Ook wordt gefocust op de vraag of GenAI ook kansen biedt voor het verder aanscherpen van het taalbeleid in het hoger onderwijs, zeker op het vlak van schrijven. In het tweede gedeelte van deze sessie ligt de focus op mogelijkheden en beperkingen van AI in het vreemdetalenonderwijs.

FRIEDA STEURS is de directeur van het Instituut voor de Nederlandse taal in Leiden en professor emeritus aan de KU Leuven, faculteit Letteren, onderzoeksgroep QLVL. Zij doet onderzoek op het vlak van computerondersteund vertalen, terminologiebeheer en meertalig documentbeheer. Het Instituut voor de Nederlandse Taal staat in voor het aanmaken, archiveren en onderhouden van corpora, lexica, woordenboeken en grammatica's voor het Nederlands, en levert daarmee alle bouwstenen voor de bestudering van de Nederlandse taal.



Intelligente technologie: hoe staat het met Nederlands?

Intelligente technologie maakt een opmars in de maatschappij, in het onderwijs en onderzoek, en in alle sectoren van het bedrijfsleven. De meeste large language models (AI modellen die zijn getraind om menselijke taal te begrijpen, te genereren en te manipuleren) zijn veruit het best getraind op het Engels. Een tool als ChatGPT zal in het Nederlands minder accurate antwoorden geven en gaat daardoor iets slechter presteren dan in het Engels. Er is ook een gemis aan talige nuance.

In het artikel "The new AI assistants may not speak your language" wordt gesteld dat ons werk achterop gaat hinken op Engelstalige landen omdat de nieuwste tools altijd eerst in het Engels worden uitgebracht en niet voor alle talen worden doorontwikkeld.

Tegelijk is het niet wenselijk dat Europa afhankelijk is van de tools van Big Tech uit de VS of China. Soevereine AI is daarom een belangrijk thema. Nederland neemt alvast maatregelen: zo ontvangt TNO een subsidie voor het ontwikkelen van een GPT-NL, samen met het Nederlands Forensisch Instituut en SURF.

Daarnaast is er een Europees project, TrustLLM waarin 6 landen uit dezelfde taalfamilie – Nederland, Duitsland, Denemarken, Noorwegen, IJsland en Zweden – samenwerken voor de

ontwikkeling van een AI-model. Door de nationale taal te overstijgen, krijg je een grotere dataset waar het model op getraind kan worden. En daardoor kan de AI hopelijk ook beter presteren.

Goede taaldata zijn essentieel. Hoe staat het op dat vlak met het Nederlands? En wat zijn de Europese initiatieven ter zake? Wat doet het Instituut voor de Nederlandse Taal op dit vlak?

JULIET VAN ROSENDAAL werkt aan een doctoraatsonderzoek over AI en democratie vanuit de traditie van kritische theorie. Ze is verbonden aan het Hoger Instituut voor Wijsbegeerte en de Chair Ethics & AI aan de KU Leuven en heeft een achtergrond in techniekfilosofie, wetenschapsfilosofie en taaltechnologie.



De gebruiksvoorwaarden van taal

De brede beschikbaarheid van Large Language Models doet ons onze taalpraktijk in een ander licht bezien. Volgens welke normen spreken en schrijven, luisteren en lezen wij? Moet taalgebruik waar zijn, inclusief zijn, transparant zijn? Welke van deze normen worden nageleefd of juist op de proef gesteld door nieuwe taaltechnologieën? In deze presentatie breng ik, refererend aan de AI ethiek, een aantal van deze normen in kaart. Daarnaast werk ik één van de thematieken uit in meer detail, namelijk de vraag of LLMs niet alleen inwerken op taalpraktijk, maar ook op een taal zelf, en of we iets kunnen zeggen over de legitimiteit hiervan.
