



# Is er een copywrAlter in de zaal?

## Scriptorijconferentie over copy, communicatie en AI



Kan een robot in de nabije toekomst een zinvolle tekst produceren? Dat was de insteek voor de Scriptorijconferentie AI AI HOERA?! op 21 november in Gent. Bij het copybureau Scriptorij lag de vraag al een tijdje op tafel. Gedreven door nieuwsgierigheid en met de ervaring van een vijftienjarig bestaan besloten we om enkele experts aan het woord te laten. En om zelf een experiment met een robot uit te voeren. Voor dat laatste klopten we aan bij Maarten Callaert en Anna Krogager, *solution engineers* bij ML6. Professor computerlinguïstiek (UA) Walter Daelemans gaf vanuit zijn expertise tekst en uitleg.

Op dit moment zijn er heel wat AI-toepassingen op de markt die helpen bij het schrijven van een tekst. Denk aan tal van *company name generators*, Grammarly, Persado, EmailCopyDyno ... Een interessante toepassing om bij stil te staan is *AI-Writer*. Die tool stelt puur op basis van een headline die je ingeeft, een artikel voor. Het resultaat is op het eerste gezicht vrij goed. Je twijfelt zelf bij aanvang even of er hier wel een robot in het verhaal is. Hoe dat komt? Omdat elke zin nog altijd door een mens geschreven is. De machine speurt het internet af naar artikels waar je onderwerp in voorkomt.

De tekst die hij voorstelt, is dus een samenraapsel van artikels die ooit al eens geschreven zijn. Er is dus best nog wat werk aan de winkel, want de tekst slaat de plank compleet mis op het vlak van opbouw en doel. Wel brengt de tekst je in contact met links naar blogartikels die je anders nooit zou vinden. Maar *out of the blue* een tekst schrijven, kan de machine dat ook?

### TALK TO TRANSFORMER

Het systeem *Talk to Transformer* maakt die belofte waar. De toepassing specialiseert zich in het schrijven van nieuwe content. Het systeem is getraind om het volgende woord in een tekst te voorspellen. Hoe werkt het model precies en maakt het zijn belofte waar? We leggen de vraag in handen van ML6. Die Gentse start-up is gespecialiseerd in onderzoek naar AI en zelflerende systemen. Daarnaast werkt het bedrijf heel wat toepassingen op maat van bedrijven uit. Anna Krogager, gespecialiseerd in *Natural Language Processing*, is een van de zestig *machine learning & data engineering* experts bij ML6 en zij zal ons bij het experiment begeleiden.

*Talk to Transformer* is van OpenAI, een bedrijf dat gespecialiseerd is in artificiële intelligentie, dat in 2015 opgericht werd door onder meer Brockman, Elon Musk en Google Brain-wetenschapper Sutskever. *Talk to Transformer* draait op het GTP-2-model (*Generative Pre-*

*trained Transformer*). Het bevat maar liefst 1,5 biljoen parameters en is getraind op een dataset (= woorden) van 800 miljoen websites. Bijzonder is dat dit wiskundige model – dus geen taalkundig model! – heeft ontdekt hoe taal gestructureerd is. Door in een enorme dataset van woorden en karakters patronen te herkennen, probeert het model het volgende woord in een zin te voorspellen. Vertaald naar ons taalkundige brein: een machine weet – na het verwerken van een massale set van data – dat een adjectief voor een substantief komt en gevolgd wordt door een werkwoord, dat op zijn beurt weer vervoegd wordt. Dat zijn patronen die het model herkent.

## THE PROOF OF THE PUDDING IS IN THE EATING

Hoe werkt die techniek precies? Hoever staat die? Dat wilden wij weten. Als opwarming voor de conferentie deden we de test. We wilden de machine niet zomaar willekeurige tekst laten genereren. We kozen voor speeches. Een speech is een specifiek genre, met een aanspreking, een intro ..., kortom: een duidelijke structuur. Maar een speech heeft ook zekere *tone of voice*, een doelstelling. Bovendien beschikken we bij Scriptorij na jaren speechschrijven voor het Agentschap voor Natuur en Bos over wel meer dan zestig speeches. *Gefundenes Fressen* voor die machine dus.

Het experiment leggen we hieronder stap voor stap uit.

### - Voorbereiding

Alle speeches, in totaal 51 104 woorden, hebben we in een document gebundeld. Ondertussen downloadde Anna Krogager het model GPT-2-simple, een kleiner model, met iets minder data en parameters: maar 700 miljoen in plaats van die anderhalf biljoen. Het eigenlijke model is namelijk te zwaar voor een gewone computer. Anna paste de code nog wat aan, waarna het model vervolgens die 51 104 woorden te verwerken kreeg. Daardoor werd het model verder getraind en verfijnd op basis van de input en de uiteindelijke opdracht.

### - Eerste poging

Na de invoering en codering was de machine klaar voor de opdracht en die luidde schrijf een nieuwe speech. Het eerste resultaat sloeg echt nergens op. We halen er eentje uit, lees maar even mee:

Beste genodigden, ik kijk uit naar wat je verfijnen. Een voorbeeld van een omgeving die deze kraadt de voorbode van een groens van de een battistaa. De battistaa een battistaaal, maar je goed bij.

De voornaamste oorzaak van het schabouwelijke resultaat: de machine is getraind in het Engels en kreeg een te beperkte dataset om mee aan de slag te gaan. Als je het experiment correct wilt uitvoeren, heb je een dataset nodig van ongeveer een half miljoen woorden. Nu waren het er maar 51 000. Toch vonden we dat we de machine nog een tweede kans moesten geven. En dus deden we drie aanpassingen.

- 1 De integrale dataset haalden we door de vertaalssoftware van DeepL (ook een AI-toepassing) van het Nederlands naar het Engels.
- 2 We gaven de machine al iets meer context mee. Zo kreeg ze de eerste paragraaf cadeau die ze vervolgens zelf verder kon aanvullen. Hoe meer context je de machine geeft, hoe beter ze er – in principe – in kan slagen om een zinvol volgende woord te voorspellen.
- 3 De context die ze kreeg, was een speech die ze nog niet verwerkt had in de oorspronkelijke dataset van speeches.

De uitdaging? Brei een vervolg aan de intro. Zo geschiedde ...

### - Tweede poging

De korte intro die de machine kreeg en waarop ze verder moest bouwen, ging over de aankondiging van een spreker, de werking van het brein, de amygdala en vooral: het belang van wandelen. Daarna kwam de robot aan de beurt. Hij mocht een vervolg verzinnen. Benieuwd? Lees even mee:

Want onderzoek heeft ons laten zien dat de natuur ontspannend en inspirerend is. En ook: rijk aan neurochemische stoffen. In ieder geval is het altijd een goed idee om een neurochemicus in het kabinet te hebben.

De zinsopbouw is logisch. En grappig bovendien. Maar zelfs een dronken tooghanger kan een samenhangender verhaal vertellen.





Maarten Callaert, solution engineer bij ML6, en UA-professor Computerlinguïstiek Walter Daelemans

- Derde keer, goede keer?

We probeerden het nog eens. Dezelfde intro, maar andere output:

**Amfetamine, zo lijkt het, geeft ons het gevoel dat we leven. Maar wat is precies de beloning voor het feit dat we in het ritme zitten? Of anders gezegd: wat is de beloning voor het feit dat we ons in een ritme bevinden? Niets in het leven is gratis. Als je vooruit blijft gaan, kom je dicht bij God.**

Het zou niet misstaan op een kalender van de Druivelaar. Grammatica is dit correct. Inhoudelijk slaat het de plank behoorlijk mis. En zo kan de robot met een druk op de knop honderden 'speeches' schrijven op basis van diezelfde intro en het getrainde datamodel.

## INTELLIGENTIE ONTBREEKT

We zijn er dus nog lang niet. Maar toegegeven: alleen het feit dat een robot volop het domein van taal verkent en probeert te begrijpen, is fascinerend (en misschien ook wel beangstigend).

Los daarvan blijft de vraag: zullen we er ooit raken? Tijd om er een academicus bij te halen die van deze materie zijn onderzoeksdomein heeft gemaakt. Professor Walter Daelemans is verbonden aan CLiPS (UA), een onderzoeksgroep die zich op psycholinguïstiek, sociolinguïstiek en computertaalkunde toelegt.

"Het grootste probleem met de diepe neurale netwerken is dat het systeem geen betekenis vat", legt hij uit. "Het enige wat het model doet, is zinnen produceren. Op zich is dat fantastisch, want op het vlak van grammatica scoort dit model beter dan een doorsnee eerstejaarsstudent", aldus Daelemans. Maar de zinnen staan autonoom en verhouden zich semantisch niet tot elkaar. Van intelligentie is geen sprake (niet in het model, wel bij studenten). Laat staan dat het systeem humor kan aanbrengen of een originele wending.

## PAPEGAAIENWERK

Wat het systeem wel kan, is bepaalde patronen ontdekken. In het experiment met GPT-2-simple ontdekte het model dat er een bepaalde structuur in de speeches zat, bijvoorbeeld dat elke tekst met een aanspreking begon. Als een goedgetrainde papegaai neemt het systeem dat over. Maar waar die aanspreking voor staat, dat weet het niet. Laat staan dat het erachter komt hoe je een speech opbouwt.

Want, legt Daelemans uit, een neuraal netwerk kan geen stijl in een tekst aanbrengen. Bij beelden kan het dat wel. Je kunt het systeem bijvoorbeeld de opdracht geven om een beeld te reproduceren in de stijl van Van Gogh, Kandinsky of Picasso. Maar taal kun je vooralsnog niet opdelen in verschillende lagen (inhoud, verhaal, stijl enzovoort) die je dan afzonderlijk bewerkt.

## CONCLUSIE: ER IS NOG WERK AAN DE WINKEL

Walter Daelemans geeft toe dat dit in grote lijnen een negatief verhaal is. Maar toch blijft hij hoopvol. Met de toegenomen rekenkracht heeft de techniek al heel wat vooruitgang geboekt. Vertalingen, en het omzetten van tekst naar spraak en dergelijke, lukt al behoorlijk goed. Maar als het over semantiek en betekenis gaat, blijft de techniek ter plaatse trappelen. De neurale netwerken begrijpen niet echt iets en ze hebben ook niets te vertellen. En dat is toch wel een voorwaarde voor tekstproductie. Vooralsnog goed nieuws voor de tekstschrijvers onder ons.

Lees nog meer over het experiment en AI op [www.scriptorij.be](http://www.scriptorij.be) en bekijk het filmpje van onze Scriptorijconferentie.

## AANBEVOLEN BRONNEN EN LITERATUUR

[www.grammarly.com](http://www.grammarly.com)

[www.persado.com](http://www.persado.com)

[www.home.copydyno.com](http://www.home.copydyno.com)

[www.ai-writer.com](http://www.ai-writer.com)

[www.talktotransformer.com](http://www.talktotransformer.com)

[www.uantwerpen.be/en/research-groups/clips/](http://www.uantwerpen.be/en/research-groups/clips/)

[www.ml6.eu](http://www.ml6.eu)

En lees ook dit eens: <https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker>.



## Petra Broeders & Bavo Van Landeghem

Petra Broeders is copywriter bij Scriptorij. Oprichter Bavo Van Landeghem legt de focus op tone of voice en verbal branding. [petra@scriptorij.be](mailto:petra@scriptorij.be) / [bavo@scriptorij.be](mailto:bavo@scriptorij.be)