

# ¿BURGER PRAAT! EEN CITIZEN SCIENCE PROJECT OVER ARTIFICIËLE INTELLIGENTIE EN GESPROKEN TAAL

## Burgerpraat

Mensen zijn van nature expert in gesproken communicatie. Computers, daarentegen, hebben het moeilijk om gesproken taal te begrijpen en emoties te identificeren. Via Burgerpraat verzamelen we data die zelflerende systemen helpen om beter en **menselijker te communiceren** in het Nederlands.

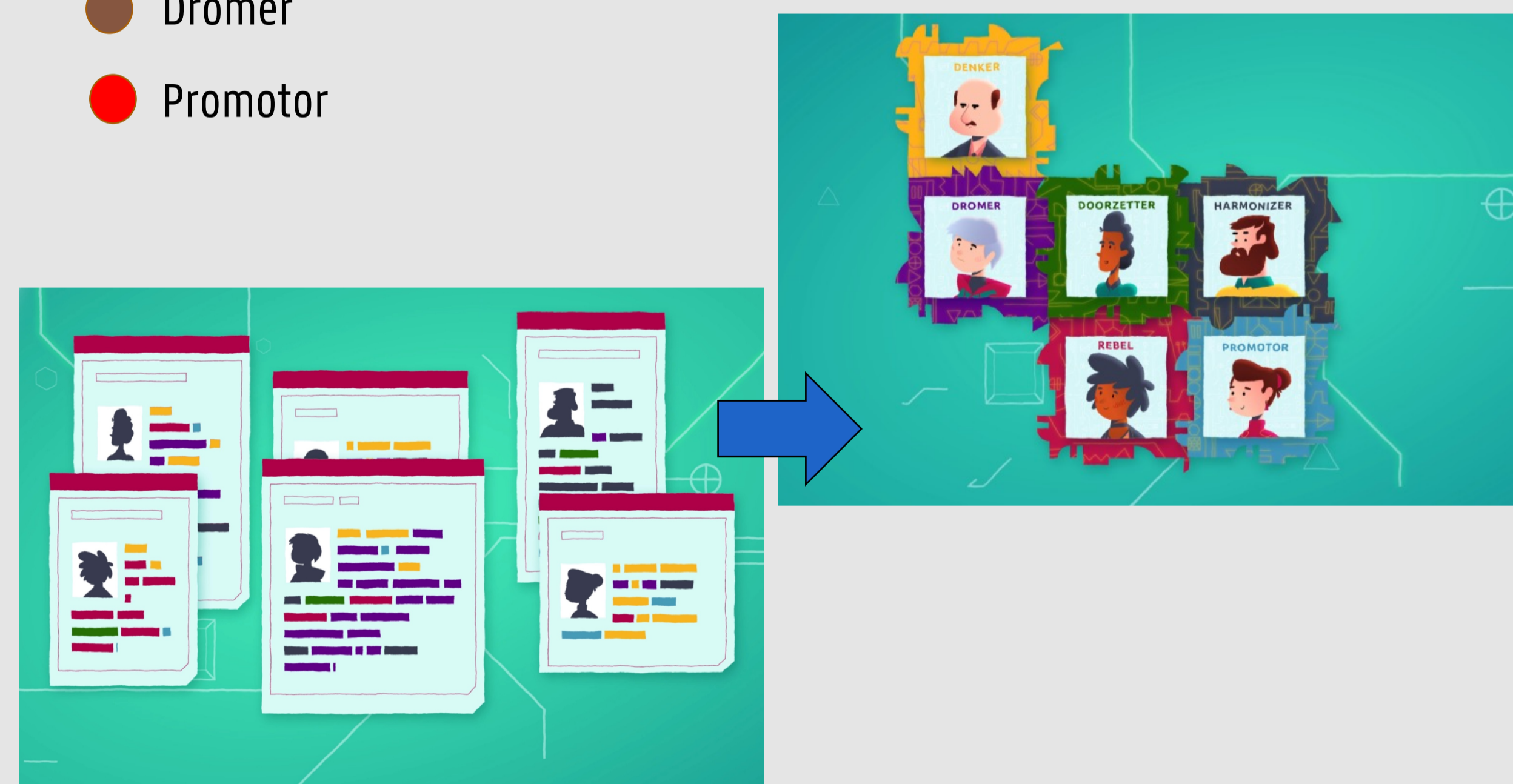
Door samen te werken met vrijwilligers creëren we een **creatieve gemeenschap** van medewerkers, burgerwetenschappers, onderzoekers en professionals die samen hun schouders willen zetten onder dit project op het snijvlak van AI, gesproken taal en gepersonaliseerde communicatie.



## PROCESS COMMUNICATION MODEL

De persoonlijkheidstest is gebaseerd op het *process communication model* ontwikkeld door Taibi Kahler dat zes **persoonlijkheidstypes** onderscheidt:

- Denker
- Harmoniser
- Doorzetter
- Rebel
- Dromer
- Promotor



De deelnemers aan het ¿Burgerpraat! project ontvangen na hun deelname hun persoonlijk communicatieprofiel en krijgen ook toegang tot een tiental infoclips over PCM, AI en taal- en spraaktechnologie..

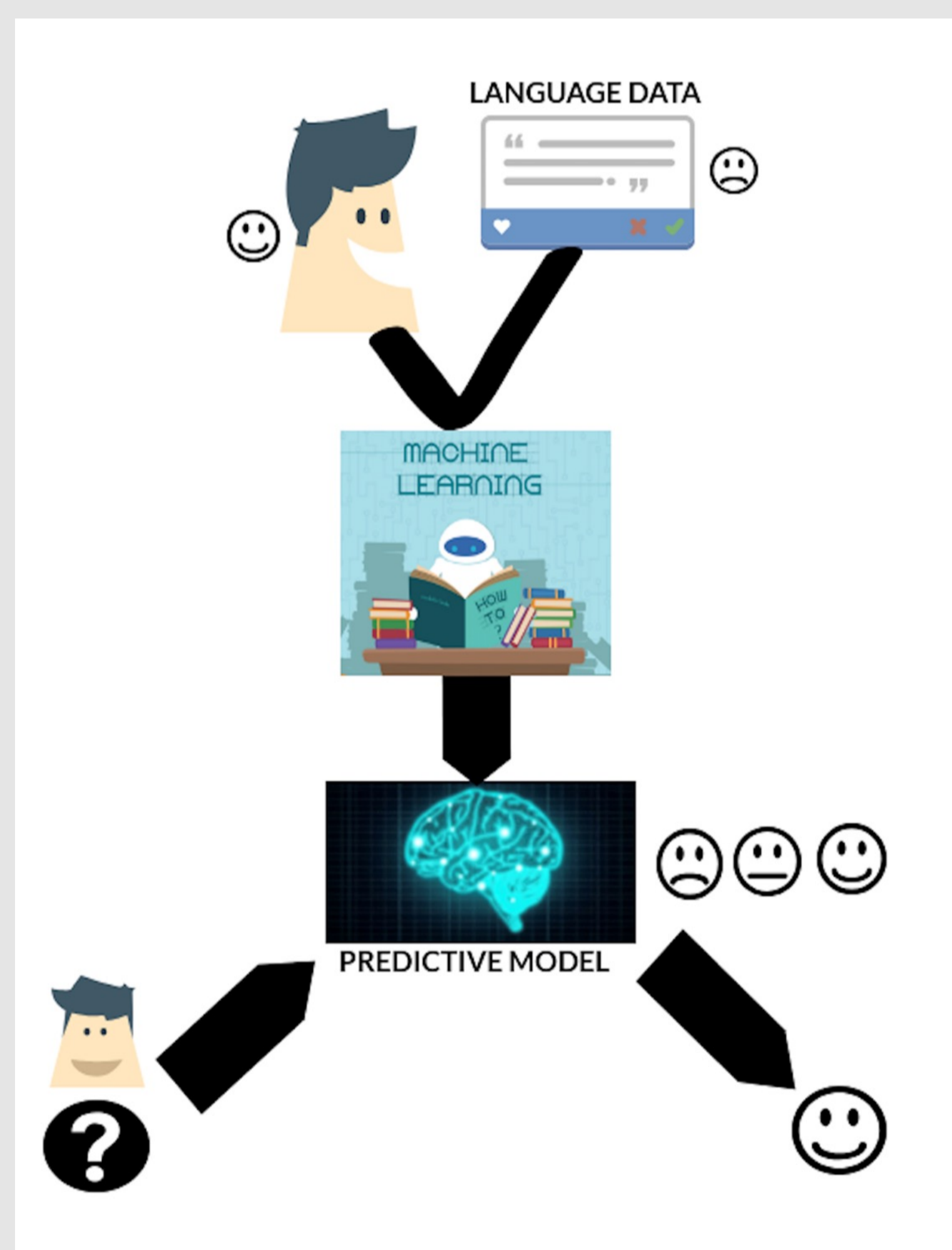
## DOELSTELLING

AI gaat over technieken om computers menselijk gedrag aan te leren, zoals het begrijpen van menselijke taal. Taalbegrip en taalproductie worden vaak gezien als een unieke menselijke vaardigheid, maar met de hulp van AI kunnen computers dit ook tot op zekere hoogte aanleren. Denk maar aan de vooruitgang die er de voorbije jaren geboekt is op het gebied van automatisch vertalen. Computers kunnen echter niet op zichzelf leren zoals een mens, en dus zijn er nog heel wat inspanningen nodig om de algoritmes te voeden met vele voorbeelden van gesproken en geschreven taal waardoor ze steeds meer inzicht krijgen in de complexe nuances van natuurlijke taal.

Vele AI-technieken om natuurlijke taal te verwerken, de zgn. natural language processing (NLP) toepassingen zoals sentimentanalyse of cyberpestdetectie, zijn nu overwegend getraind op databanken van geschreven taal. Maar er is daarnaast een grote nood aan spraakdatabanken die ons moeten toelaten om algoritmes te ontwikkelen om communicatie in gesproken taal beter te begrijpen en er gepaster op te reageren. In ¿Burgerpraat! willen we met de hulp van vele vrijwilligers dergelijke data verzamelen voor het Nederlands.

## METHODE

Om dit taal- en spraaktechnologieonderzoek mogelijk te maken, willen we data van 1000 burgers verzamelen. We zoeken personen vanaf 16 jaar en ouder die moedertaalsprekers zijn van het Nederlands. De dataverzameling bestaat enerzijds uit de opname van een selfie-video van zo'n 10 minuten en anderzijds uit het invullen van een online vragenlijst die dient als persoonlijkheidstest.



De video-opnames worden automatisch getranscribeerd via automatische spraakherkenning en daarna manueel verbeterd door een onderzoeker. Deelnemers kunnen ook hun eigen data bekijken. Na transcriptie wordt de video verwijderd.

De spraakdata, transcriptie en een PCM-profiel (6 cijfers) worden bewaard op een UGent server, waartoe enkel de bevoegde onderzoekers toegang hebben. De data worden via pseudonimisering gekoppeld aan de persoonsgegevens van de deelnemers, zodat we nog tot 5 jaar na deelname aan dit project data van deelnemers kunnen verwijderen. Na 5 jaar worden de persoonsgegevens verwijderd en wordt de bewaring en verdeling van de data overgedragen aan het Instituut voor de Nederlandse Taal (INT) dat als CLARIN B-centrum fungeert voor Vlaanderen.

## Contact

[burgerpraat@ugent.be](mailto:burgerpraat@ugent.be)  
[www.burgerpraat.be](http://www.burgerpraat.be)

