

Gepast



Ongepast

Data

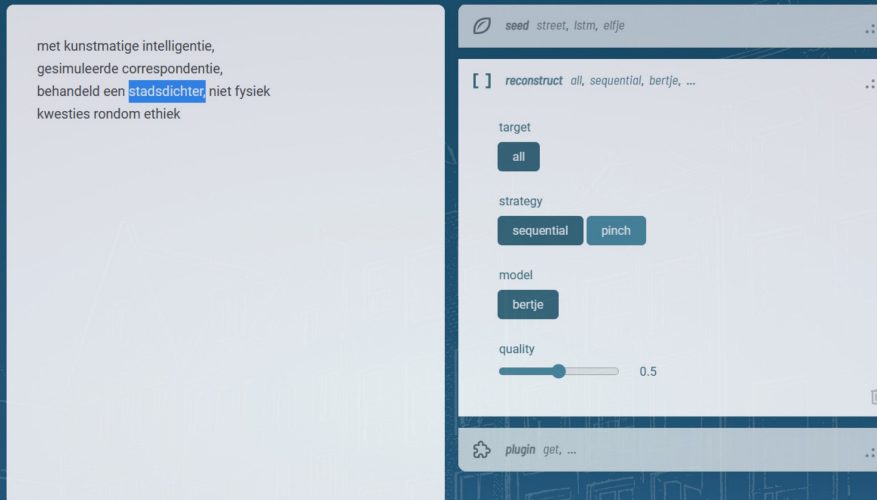
Het vinden van een geschikte dataset bleek al direct een probleem. Om ons model *eerlijk* en *niet-discriminerend* te houden wilde wij idealiter een dataset representatief van de verschillende maatschappelijke achtergronden binnen Amsterdam. Hiertoe hebben wij gebruik gemaakt van een dataset bestaand uit straatpoezie, maar na deze te analyseren werd het al snel duidelijk dat wij enkele standaarden moesten stellen vanuit **beginselethiek**; de vraag: wat is gepast? Hiertoe hebben wij de data eerst gefilterd aan de hand van een woordenlijst van woorden die over het *algemeen* worden beschouwd als scheldwoorden.

Human in the loop

AI snel beseftte wij ons dat het niet reëel is om alle aanstootgevende uitvoer te voorkomen. Hierom hebben altijd de mogelijkheid bewaard het gegenereerde gedicht met de hand bij te stellen en het model de juiste richting op te helpen.

Responsible IT

In opdracht van de Hogeschool van Amsterdam kregen wij de taak om een zogenaamde *Digitale Stadsdichter* te maken als cadeau aan Gemeente Amsterdam. Met deze **stakeholders** in gedachten begonnen wij onze onderlinge discussie in de context van **gevolgenethiek**. Dit bleek voor ons het meest succesvolle uitgangspunt, gezien wij zelfs als een relatief nieuw team het al direct eens konden zijn over praktische zaken zoals het voorkomen van conflict met de AVG-wet. Dit liep vervolgens soepel over in een *ethische* discussie door dieper te graven in de maatschappelijke consequenties die wij hiermee indirect zouden behandelen en de **waarden** hierachter.



met kunstmatige intelligentie,
gesimuleerde correspondentie,
behandelt een stadsdichter, niet fysiek
kwesties rondom ethiek

niet betreffend
met een voor en

Technologie

Zero-shot learning vormt de basis van ons project, waar de uitdaging is om een gedicht te genereren dat qua stijl overeenkomt met een enkel voorbeeld. Deze voorbeelden kunnen van de gebruiker komen, of vanuit een LSTM-model die de gehele dataset generaliseert. Het gegeven voorbeeld wordt vervolgens vernietigd en hersteld met de masking pipelines van voorgetrainde transformers met voldoende encoder lagen zoals BERTje en ROBERTus. Dit proces staat een hoge maat aan *transparantie* en *uitlegbaarheid* toe, doordat deze hele pipeline in beeld kan worden gebracht voor de eindgebruiker; hiertoe hebben wij een systeem gebouwd om dit zo toegankelijk mogelijk te maken. Met *privacy* in gedachten onthoudt de server hiervoor enkel de invoer zo lang nodig om deze te verwerken, waarna de geschiedenis client-side wordt opgeslagen.

Doelgroep

Voor het systeem waar de pipeline kan worden bijgesteld hadden wij een publiek in gedachten met een technische achtergrond, maar dit bleek ongeschikt voor een mogelijke publieke installatie waarbij dergelijke locaties voornamelijk kunst- en dichtliehebbers aan zullen trekken. Dit leidde tot een ondersteunend systeem met een statische pipeline en een lijke verf.

Hierbij zoekt het systeem aan de hand van de voorgenoemde masking pipelines naar woorden met veel mogelijk alternatieven, waarna de eindgebruiker enkel een keuze hoeft te maken tussen deze alternatieven. Geïnspireerd door het Schip van Theseus, hopen wij voor dit publiek **deugdethiek** aan te pakken door *goed* en *fout* in de context te zetten van inspiratie en plagiaat waarbij zij voor hun neus het initiële gedicht zien evolueren aan de hand van een algoritme en hun keuzes.