

# Technische fiche PROBE

28 augustus 2024

**Entiteit**

**Domein Informatie**

**Contactpersoon**

Leen Van Campe, Frederik D'haemer, Jens Deprez

---

## *Inleiding*

Het doel van PROBE was om de besluitvorming dichterbij de Gentenaar/burger te brengen. Concreet kan artificiële intelligentie die complexe data toegankelijker maken voor de Gentenaar én de stadsmedewerker.

Dit hebben we verwezenlijkt door in te zetten op een NLP (Natural Language Processing) oplossing. Aan de hand van deze oplossing is het mogelijk om besluiten geautomatiseerd te koppelen aan op voorhand gedefinieerde thema's. Verder wordt er binnen het project gekoppeld met een front-end applicatie die ervoor zorgt dat stadsmedewerkers op een eenvoudige en elegante manier de nodige aanpassingen kunnen doen aan geannoteerde besluiten. Deze applicatie maakt het mogelijk om de AI-annotaties te corrigeren, hierdoor kan diezelfde AI de correcties meenemen in zijn volgende iteratie (supervised learning).

De resultaten van deze geannoteerde besluiten kunnen als handige stukjes informatie, ofwel widgets, aan verschillende pagina's van de website van de Stad Gent worden toegevoegd. Hierdoor zijn ze gemakkelijk terug te vinden voor de bezoekers. In deze widgets ziet de burger een overzicht van de besluiten, die door AI zijn geanalyseerd en ingedeeld op thema. Op deze manier is de cirkel rond.

Meer informatie over dit project kan hier terug gevonden worden:

<https://stad.gent/nl/over-gent-stadsbestuur/stadsbestuur/wat-doet-het-bestuur/gent-internationaal/gesubsidieerde-projecten/probe-proactieve-openbaarheid-van-bestuur>

## Componenten

In dit document vatten we kort samen uit welke technologieën en applicaties PROBE is opgebouwd, we linken in de mate van het mogelijke ook door naar de gerelateerde repositories.

PROBE is opgebouwd uit volgende componenten:

### 1. Harvester:

Mus-semtech stack voor het harvesten en verwerken van besluiten uit externe bronnen.

LINK: <https://github.com/lblod/app-lblod-harvester>

### 2. Apache Airflow

Open-source platform dat wordt gebruikt om workflows programmeerbaar te maken, te plannen en te monitoren.

LINK: <https://airflow.apache.org/>

Hier worden volgende DAGs voor gebruikt, dit zijn specifieke workflows van een verzameling taken die zijn georganiseerd op een manier die hun relaties en afhankelijkheden duidelijk definieert.

De DAGS die binnen PROBE gebruikt worden zijn hier gedefinieerd:

<https://github.com/semantic-ai/app-probe-airflow-dags>

Binnen deze DAGs worden volgende modellen van <https://huggingface.co> gebruikt:

*"paraphrase-multilingual-mpnet-base-v2", "intfloat/multilingual-e5-small", "thenlper/gte-large", "multi-qa-mpnet-base-dot-v1" "MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7", "facebook/bart-large-mnli", "mjjwong/multilingual-e5-base-xnli-anli"*

### 3. MLFLOW

Open-source platform om de Machine Learning cycle te managen.

LINK: <https://github.com/mlflow/mlflow>

### 4. BeAT: Besluiten Annotatie Tool

De user interface die gebruikt wordt om manueel besluiten te annoteren.

LINK: <https://github.com/semantic-ai/frontend-probe-beat>

## Extra documentatie kan hier gevonden worden:

Documentatie + codebase overview:

<https://semantic-ai.github.io/probe-sparql-mono-service/>

LINK naar Docker Container Images:

<https://hub.docker.com/u/stadgent>

## Algemene architectuur van het project:

