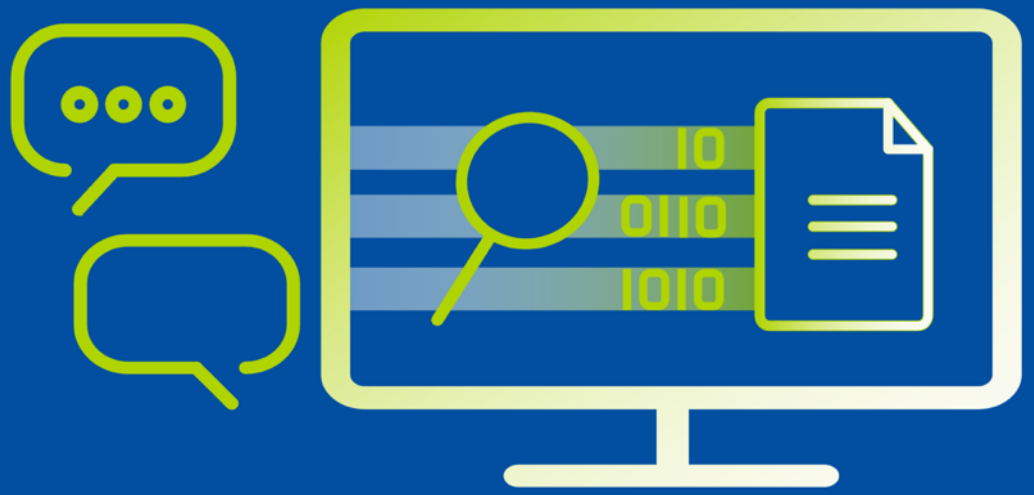


ITU Open Source Programme Office

ITU Open Source Programme Office (OSPO) is committed to harnessing the power of open-source technologies to empower nations, organizations, and individuals to leverage cutting-edge digital solutions for the sustainable development.

GovStack

The GovStack Global Inititive

Open Source Ecosystem Enabler

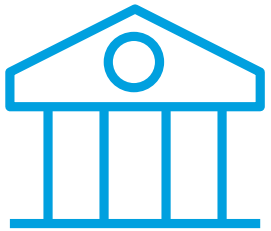
ITU-D Network & Digital Infrastructure

ITU OSPO funded by: TIDES A Force for Social Good

Challenge "AIntuition"

Retrieval Augmented Generation (RAG) for Public Services and Administration Tasks

RATIONALE



Public sector

plays critical role in advancing the Sustainable Development Goals (SDGs) as most SGD targets critically depend on the work of public institutions.

How can we harness AI technology to enhance the efficiency of the public sector and deliver better services to people?

Open-Source AI for Public Services

Use-Cases and Risks

Unacceptable risk | High risk | Limited risk | Minimal risk

Intelligence, law enforcement & defence
- Mass surveillance
- Predictive policing
- Realtime crime mapping
- Crowd management
- Roadway intelligence

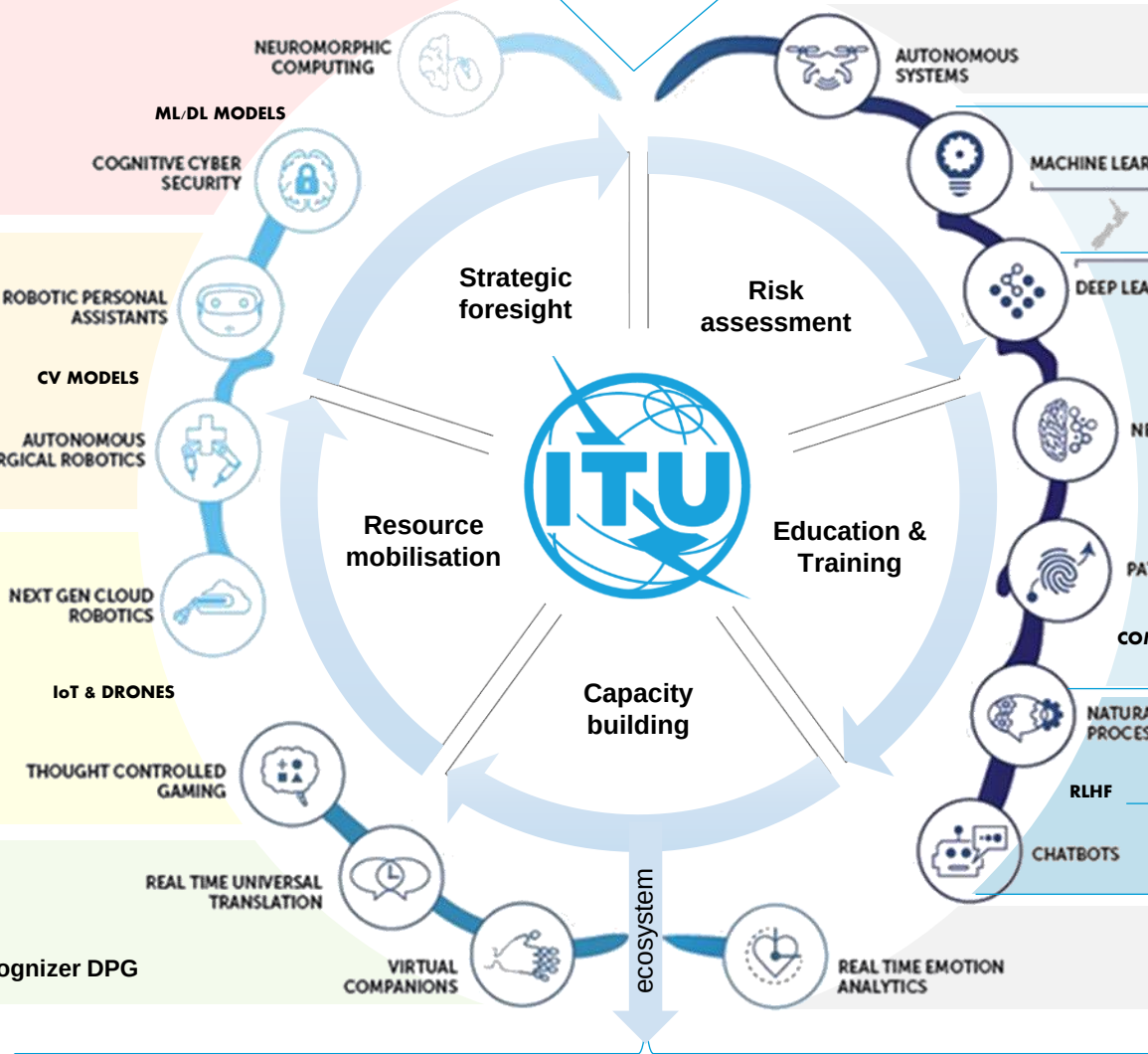
Healthcare & humanitarian response
- Child malnutrition detection / **Meron DPG**
- Dengue cases and deaths correlation with real-time climate data / **AEDES DPG**
- Predictive AAC platform for people with speech impairment / **OTTAA DPG**
- Humanitarian aid based on poverty status / **Cider DPG**

Public administration
- Automated back office and RPA

Agriculture
- Precision agriculture / **AI Agro DPG**

Climate & Energy
- Energy demand and response / **CityLearn DPG**

Public administration and education
- Citizen services chatbots / **Burokratt**
- Abusive and cyberbullying behaviours / **Kindly DPG**
- Real-time feedback on pronunciation / **Phoneme Recognizer DPG**



Open-Source Technologies

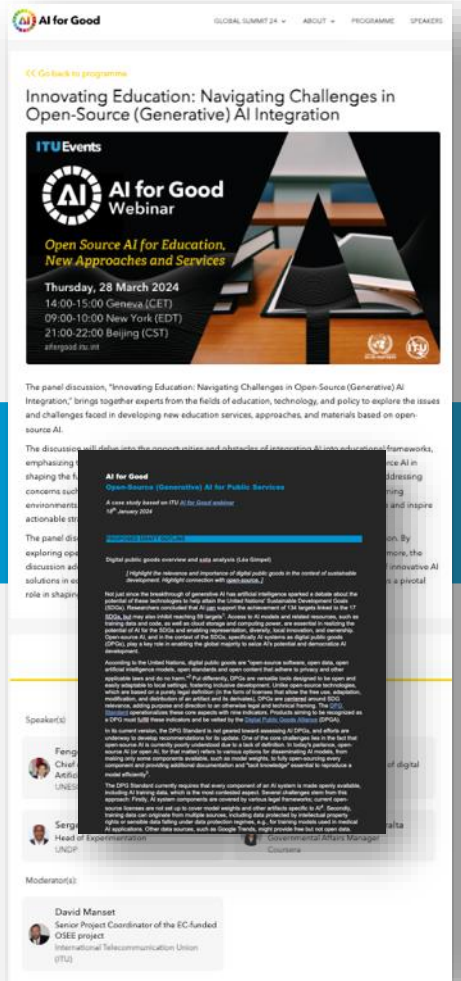
Open data for AI models development | Open-source AI technologies (OST) | Responsible AI | Explainable AI (XAI)

Stable Diffusion, mindsdb, Ivy.ai, tflearn/tflearn, mxnet, theano, fast.ai, TensorFlow, PyTorch, opennn, Keras, Detectron2, Hugging Face, OpenCV, Open Assistant, Meta AI / LLaMa, GPT Engineer, stability.ai, StackLLaMA, opendilab / awesome-RLHF

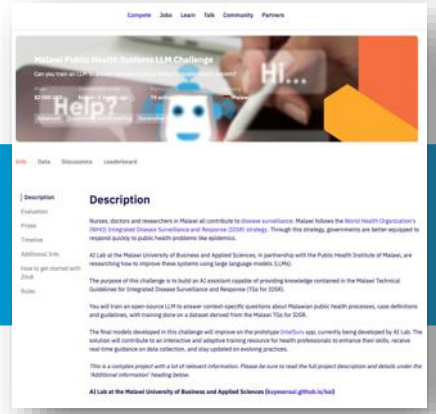
Digital Public Goods Alliance, UNICC, UNDP, European Commission, APACHE Software Foundation, eGov Foundation, Linux Foundation, open source initiative, ...

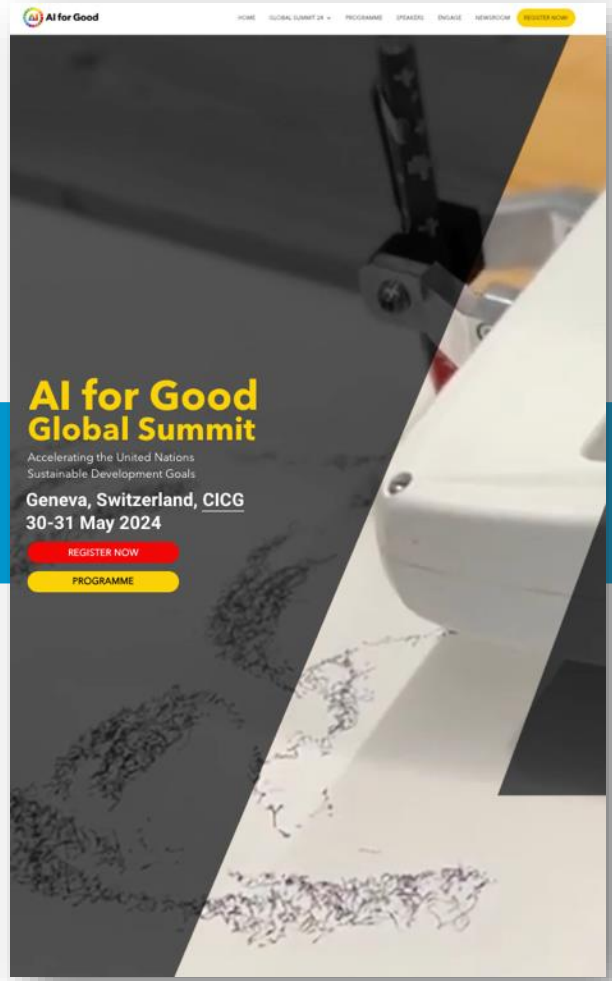
The hitchhikers guide to the open-source AI galaxy…

Focussing on the AI track…



Create a community





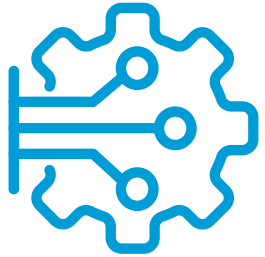
Prototype Implementation for Public Services

WHY OPEN-SOURCE ?

Security, privacy,
and reliability

Performance
and quality

Cost and
administration



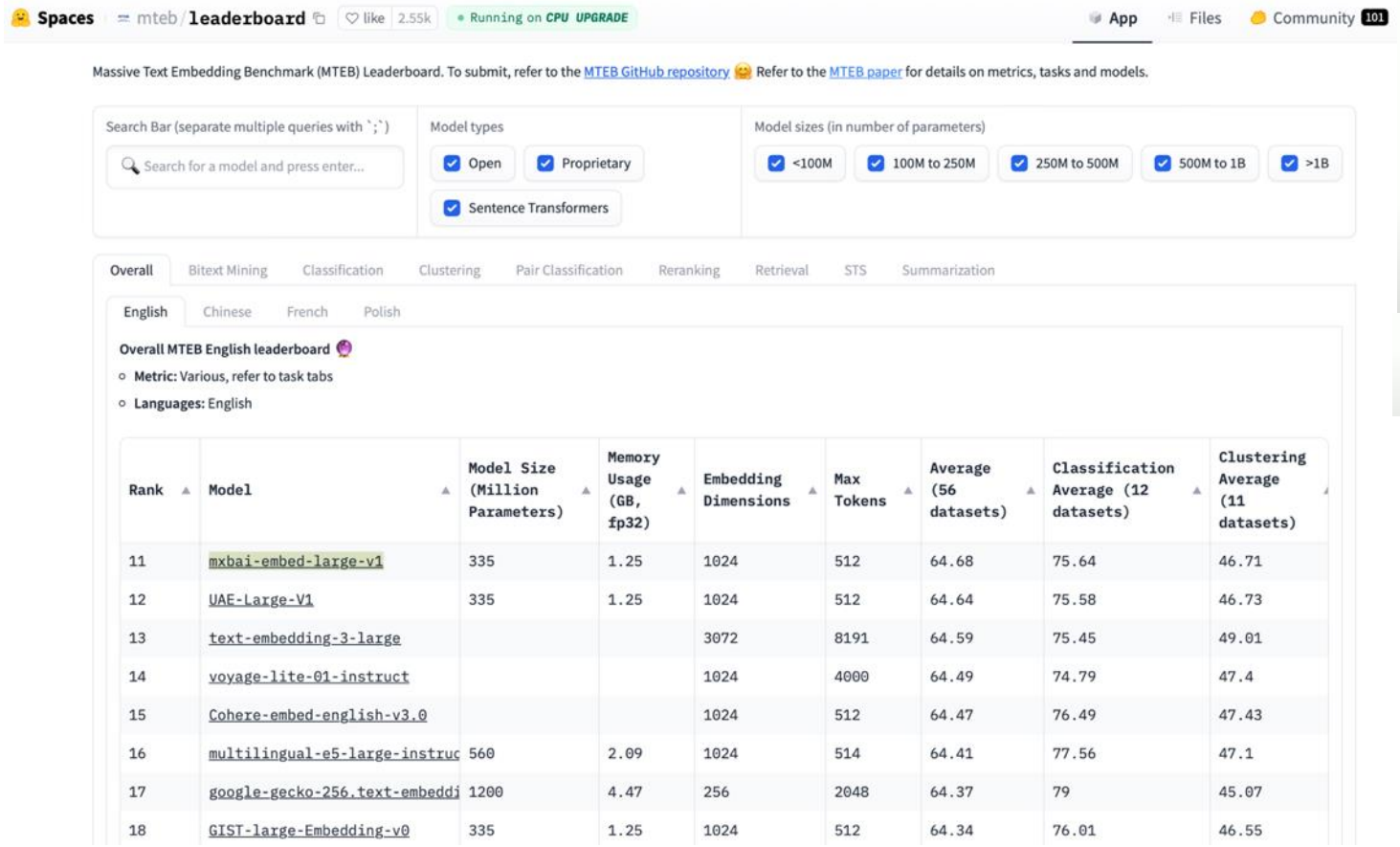
Flexibility and
fine-tuning

Collaboration
and exchange

Ethics, standards,
and transparency

WHY OPEN-SOURCE ? (OPEN-SOURCE VS PROPRIETARY RANKING)

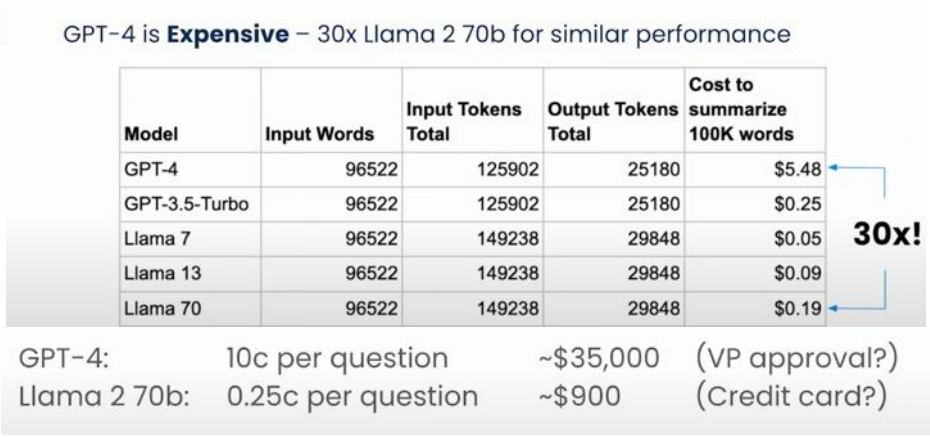
Performance



Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)	Classification Average (12 datasets)	Clustering Average (11 datasets)
11	mxbai-embed-large-v1	335	1.25	1024	512	64.68	75.64	46.71
12	UAE-Large-V1	335	1.25	1024	512	64.64	75.58	46.73
13	text-embedding-3-large			3072	8191	64.59	75.45	49.01
14	voyage-lite-01-instruct			1024	4000	64.49	74.79	47.4
15	Cohere-embed-english-v3.0			1024	512	64.47	76.49	47.43
16	multilingual-e5-large-instruc	560	2.09	1024	514	64.41	77.56	47.1
17	google-gecko-256.text-embeddi	1200	4.47	256	2048	64.37	79	45.07
18	GIST-large-Embedding-v0	335	1.25	1024	512	64.34	76.01	46.55

Source: HuggingFace MTEB leader board: https://huggingface.co/spaces/mteb/leaderboard

Cost and administration



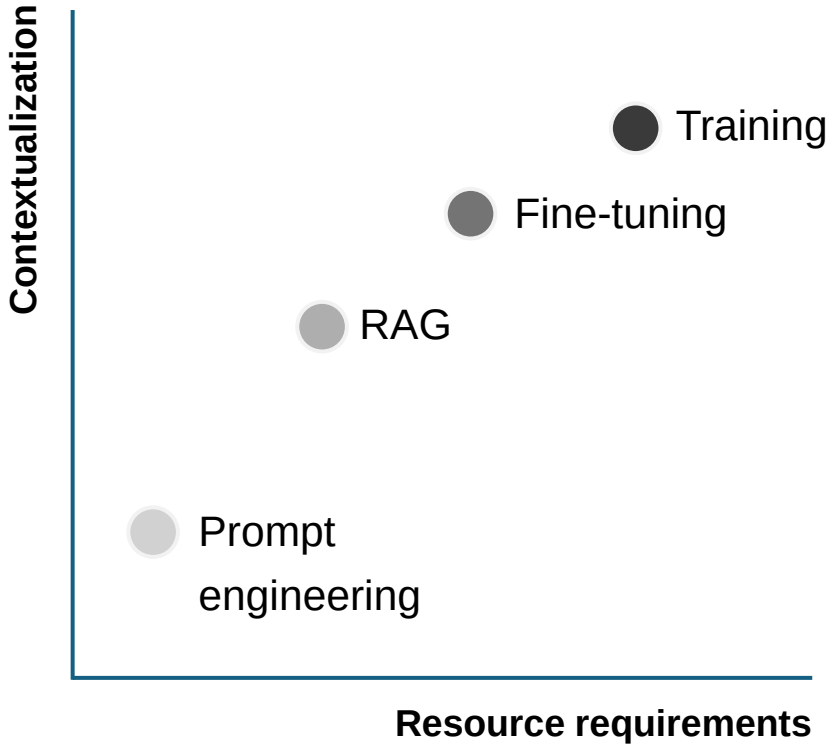
Model	Input Words	Input Tokens Total	Output Tokens Total	Cost to summarize 100K words
GPT-4	96522	125902	25180	$5.48
GPT-3.5-Turbo	96522	125902	25180	$0.25
Llama 7	96522	149238	29848	$0.05
Llama 13	96522	149238	29848	$0.09
Llama 70	96522	149238	29848	$0.19

Source: Anyscale at Ray Summit 23:

https://www.youtube.com/watch?v=Ri_LJ_qOTPM

FINE TUNING VS RAG?

Fine-Tuning vs RAG



Aspect	RAG	Supervised Finetune
Dynamic data	✓	✗
Static data	✓	✗
Internal Data	✓	✗
Reduce Hallucinations	✓	✓
Transparency of Generation	✓	✗
Fine Tune Smaller Model	✗	✓
Brand Voice in Generation	✗	✓

Source: Galileo RAG Vs Fine-Tuning Vs Both: A Guide For Optimizing LLM Performance

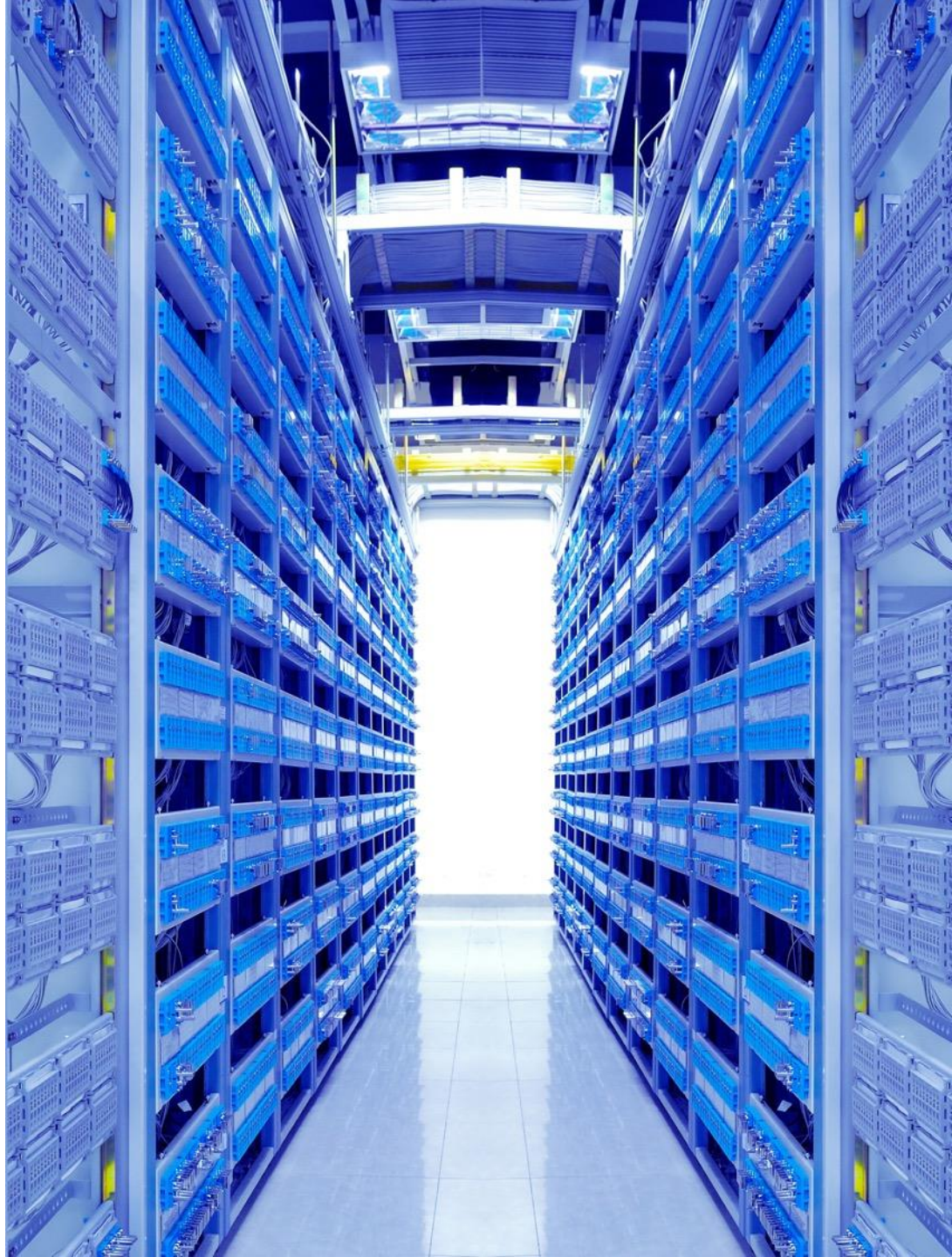
HACKATHON: “AIntuition”

Retrieval Augmented Generation (RAG) for Public Services and Administration Tasks

Engage community

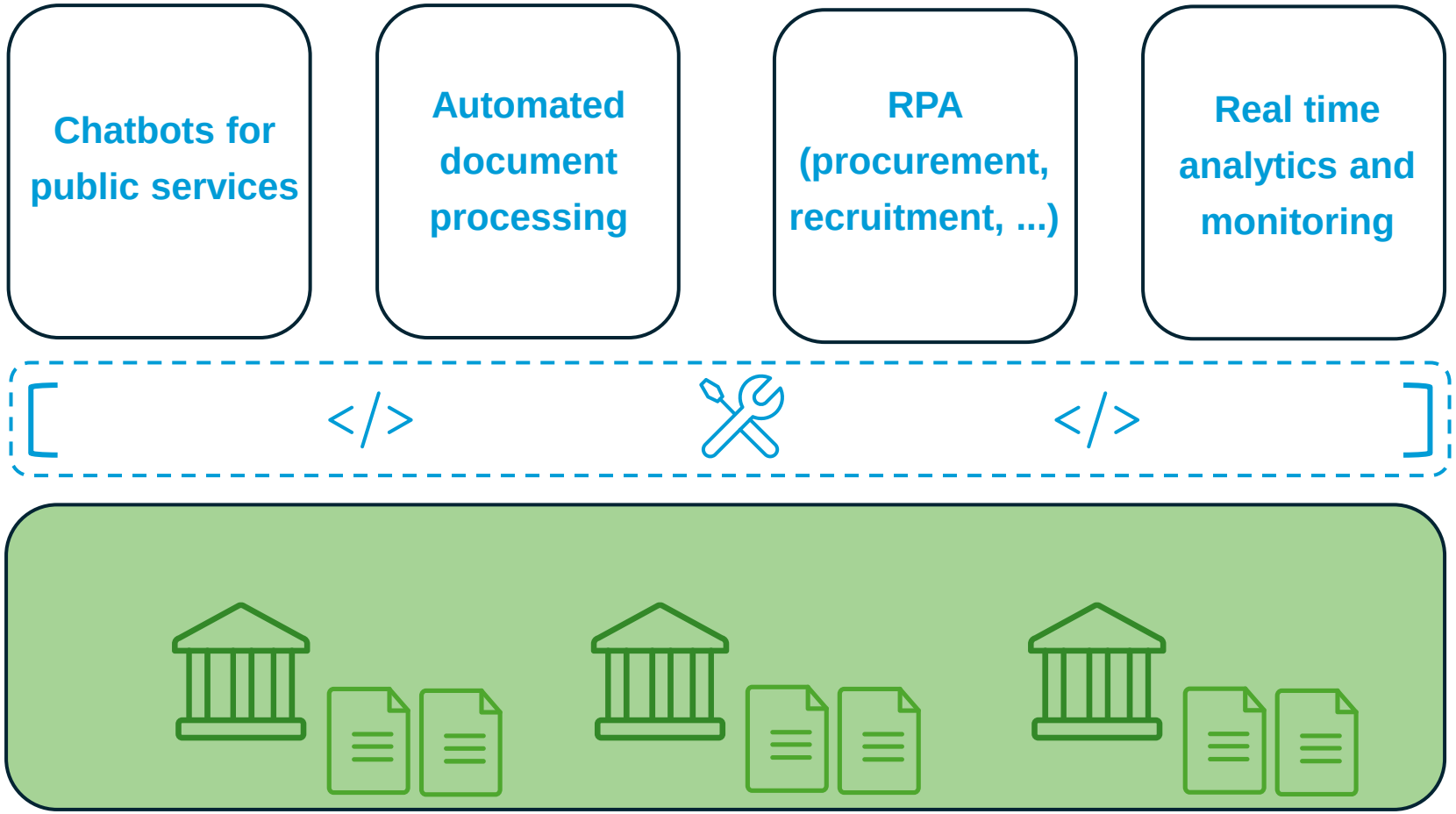
Promote open-source uptake

Promote public sector use-case



WHY RETRIEVAL AUGMENTED GENERATION ?

Public sector LLM use-cases

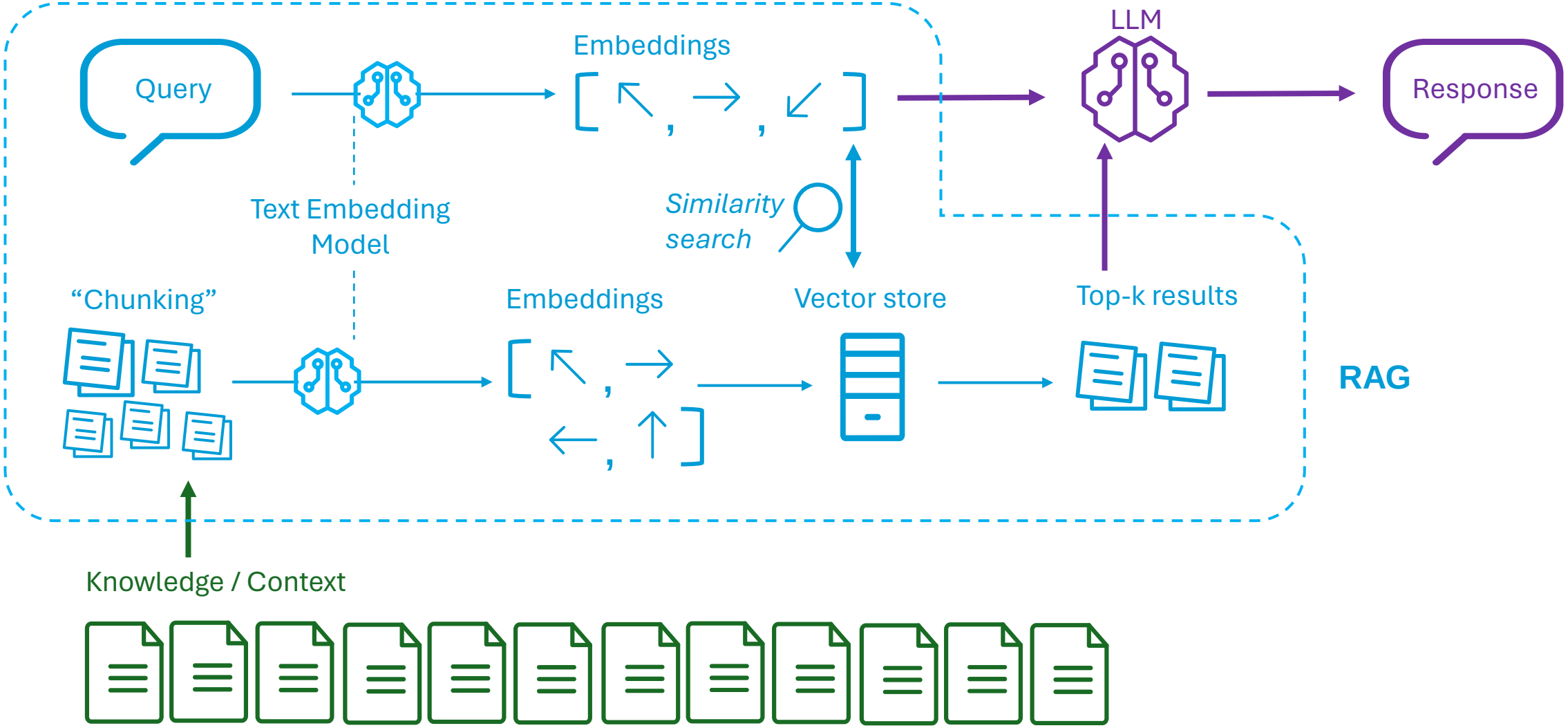


Opportunity to improve efficiency and deliver better services using data

RAG

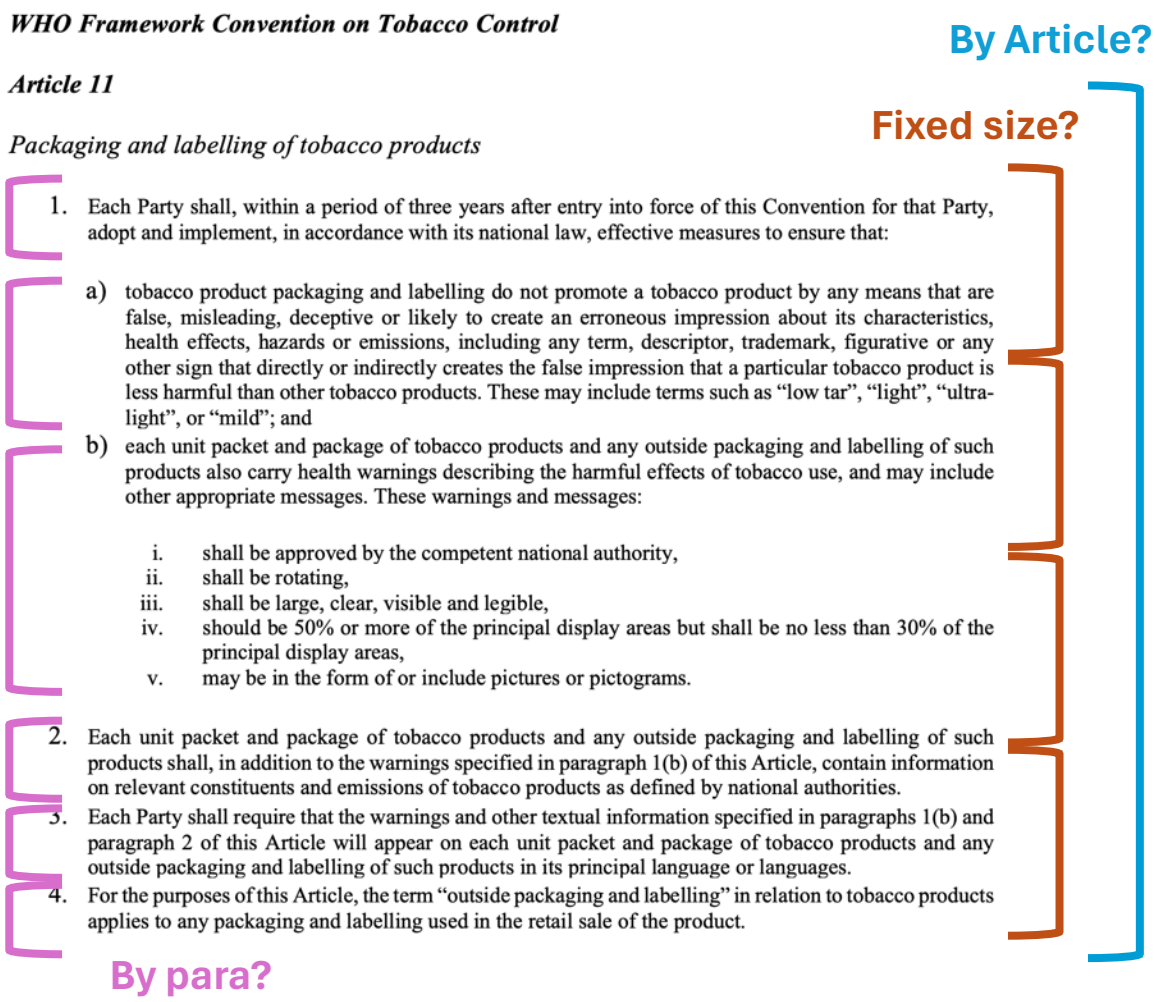
Public institutions are typically the main owners of data in countries

WHAT IS THE CHALLENGE ABOUT ?



WHAT IS THE CHALLENGE ABOUT ?

“Chunking”: what is the right approach?



WHO Framework Convention on Tobacco Control

Article 11

Packaging and labelling of tobacco products

1. Each Party shall, within a period of three years after entry into force of this Convention for that Party, adopt and implement, in accordance with its national law, effective measures to ensure that:

   a) tobacco product packaging and labelling do not promote a tobacco product by any means that are false, misleading, deceptive or likely to create an erroneous impression about its characteristics, health effects, hazards or emissions, including any term, descriptor, trademark, figurative or any other sign that directly or indirectly creates the false impression that a particular tobacco product is less harmful than other tobacco products. These may include terms such as “low tar”, “light”, “ultra-light”, or “mild”; and

   b) each unit packet and package of tobacco products and any outside packaging and labelling of such products also carry health warnings describing the harmful effects of tobacco use, and may include other appropriate messages. These warnings and messages:

      i. shall be approved by the competent national authority,
      ii. shall be rotating,
      iii. shall be large, clear, visible and legible,
      iv. should be 50% or more of the principal display areas but shall be no less than 30% of the principal display areas,
      v. may be in the form of or include pictures or pictograms.

2. Each unit packet and package of tobacco products and any outside packaging and labelling of such products shall, in addition to the warnings specified in paragraph 1(b) of this Article, contain information on relevant constituents and emissions of tobacco products as defined by national authorities.
3. Each Party shall require that the warnings and other textual information specified in paragraphs 1(b) and paragraph 2 of this Article will appear on each unit packet and package of tobacco products and any outside packaging and labelling of such products in its principal language or languages.
4. For the purposes of this Article, the term “outside packaging and labelling” in relation to tobacco products applies to any packaging and labelling used in the retail sale of the product.

Fine-tune embedding models?

```
Query term: deliverables

Compare term:  project
Cosine Similarity nomic-embed-text:  0.28449301649656
Cosine Similarity mxbai-embed-large:  0.6816990047633592

Compare term:  report
Cosine Similarity nomic-embed-text:  0.3169439197432541
Cosine Similarity mxbai-embed-large:  0.5981404731324258

Compare term:  goods
Cosine Similarity nomic-embed-text:  0.3978022001712965
Cosine Similarity mxbai-embed-large:  0.612249724637433

Compare term:  services
Cosine Similarity nomic-embed-text:  0.37418290999418924
Cosine Similarity mxbai-embed-large:  0.6503712207152618

Compare term:  promise
Cosine Similarity nomic-embed-text:  0.47427282960418016
Cosine Similarity mxbai-embed-large:  0.5496129986876679


Diplaying top 3 for each model:

nomic-embed-text:
[('promise', 0.47427282960418016), ('goods', 0.3978022001712965), ('servi

mxbai-embed-large:
[('project', 0.6816990047633592), ('services', 0.6503712207152618), ('goo
```

WHAT IS THE CHALLENGE ABOUT ?

Query pre-processing?

Query: “How do I get an approval for a publication?”

Relevant chunk: “To obtain an approval for a publication follow the procedure described in the Guidelines for planning future publications and events.”

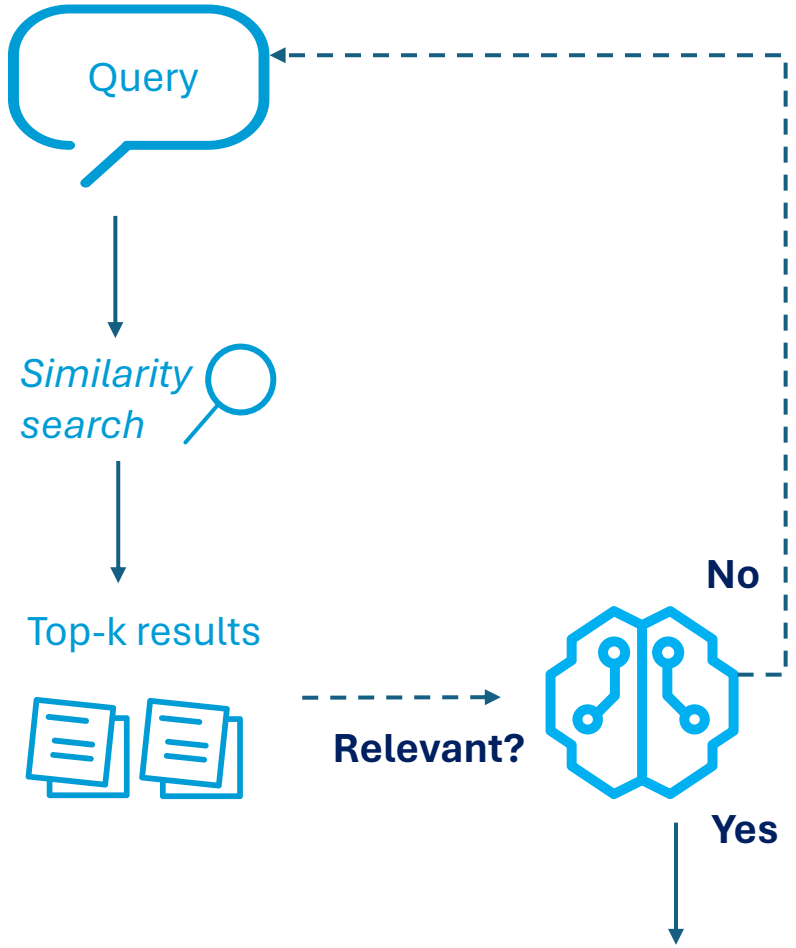
Query: “How do I get an approval for a publication?”

Sub-Query: “How does publication approval process work in my department?”

Sub-Query: “What is the first step to initiate a publication approval?”

Sub-Query: “What documents should I submit to get my publication approved?”

Multi-agent / multi-model workflows ?



WHAT IS THE CHALLENGE ABOUT ?

Criteria	Weight
Quality (how complete and relevant is the extracted information?)	60%
Efficiency (what are the size and compute requirements?)	30%
Accompanying documentation and ease of use	10%

Completeness

$$\frac{\sum_{i=1}^{n} (w_i \times presence_i)}{\sum_{i=1}^{n} w_i}$$

Precision:

$$\tau\ (retrieved\ results,\ ground\ truths)$$

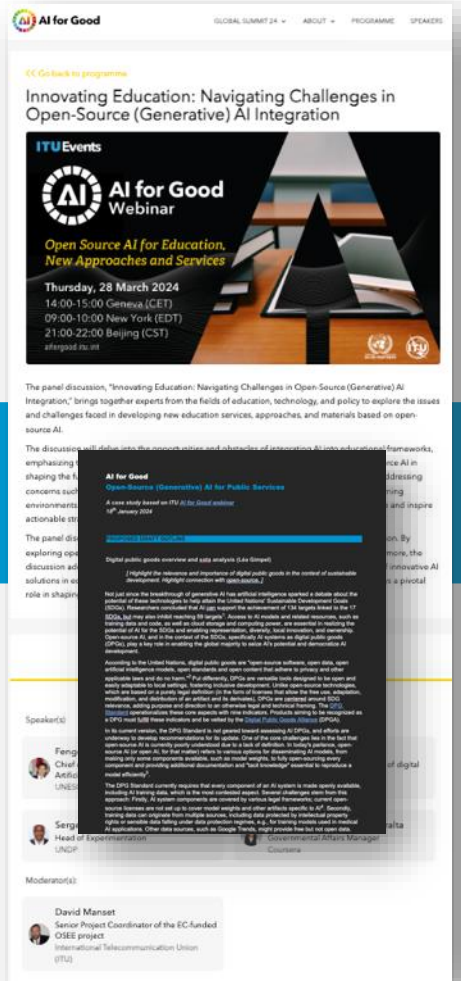
Noise:

$$\frac{(length\ top\ 5 - length\ groumd\ in\ top\ 5)}{length\ top\ 5}$$

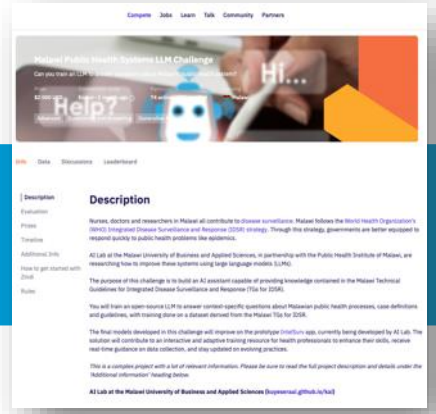
Where:

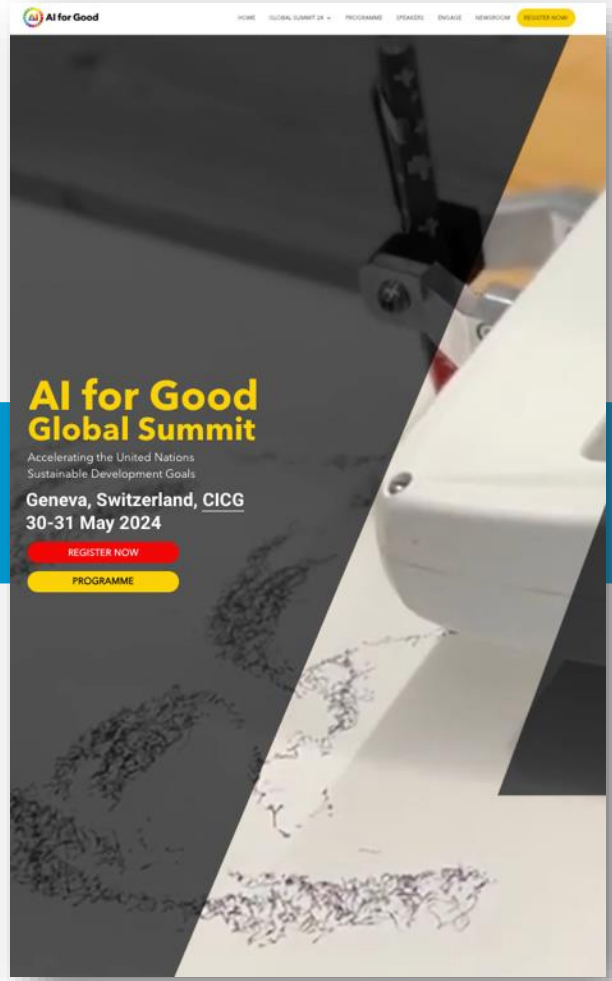
- w_i - is weight of the i-th ground truth;
- presence_i - is 1 if the i-th ground truth appears in the top five retrieved results, 0 otherwise;
- n - is the total number of ground truths in the reference

Focussing on the AI track…



Create a community





Prototype Implementation for Public Services

AI Global Summit 2024

Unleashing the Power of Open-Source AI: Transforming Digital Public Services for a Better Tomorrow / 31st May 2024 8.30am to 12.15pm CEST

Abstract / Description

As governments worldwide increasingly recognize the transformative potential of Artificial Intelligence (AI), ensuring ethical, sustainable, and cost-effective approaches to its implementation is paramount. This session will delve into the crucial role of open-source AI technology and frameworks in enabling safe and efficient adoption, use, and scaling of AI-based services and applications within the public sector. It is co-organized by the International Telecommunication Union (ITU) and the German Development Cooperation initiative FAIR Forward implemented by GIZ on behalf of the German Federal Ministry for Economic Cooperation and Development (BMZ).

The session will foster dialogue among diverse stakeholders, sharing experiences, best practices, and cooperation opportunities to enhance AI capacity in public administration. It will feature real-world use cases from Kenya, Rwanda, and Kazakhstan, offering insights into lessons learned and different stakeholders' perspectives. To further illustrate the potential of AI applications in public administration, a prototype implementation of a Generative AI (GenAI) solution, leveraging open-source software and Large Language Models (LLMs), will be showcased to tackle basic public administration use-cases such as citizen chatbot and process automation.

A critical aspect of the discussion will be the alignment of concrete use-cases, tools, and experiences with open-source AI policy considerations, including regulations, norms, and practices. In a concluding panel discussion, the session will explore approaches to fostering national and international ecosystems conducive to the emergence of ethical open-source AI tools, which can be shared and re-utilized as Digital Public Goods (DPGs). Examining open-source AI within the framework of data governance, international standards, and DPGs, this session equips policy- and decision-makers with insights to drive impact-oriented action. By promoting AI solutions that positively impact the public sector and beyond, whilst equally highlighting risks and harms to avert, the session aims to catalyze meaningful progress towards inclusive and sustainable development goals.

Hackathon competition result @ AI Global Summit

Other **thematic instances**

Integration to **OSEE Framework**
As hands-on training

AIntuition Prototype Implementation
MVP for digital public services
Integration to GovStack sandbox