Does AI increase or decrease transparency?

Tools and our relations with them

Yung-Hsuan Wu
Researcher, DiploFoundation

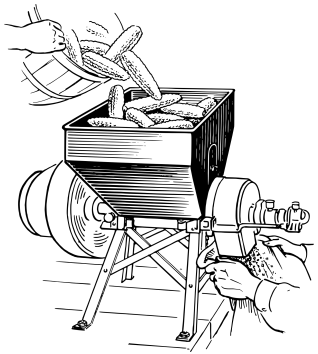
DiPLO Geneva Internet Platform

Know the talk before you walk the talk.

At Diplo, we insist on knowing what AI is before delving deep into AI governance issues.

It is crucial to our research, teaching, discussions, and work.

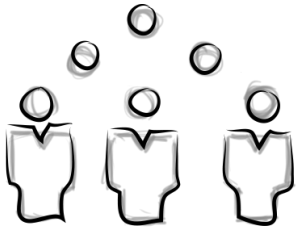
Workflow of creating a survey



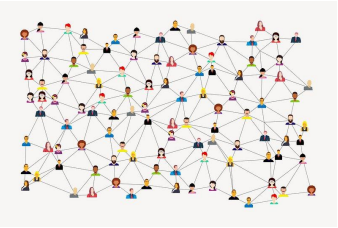
Writing questions

Fig. 29.

Testing survey

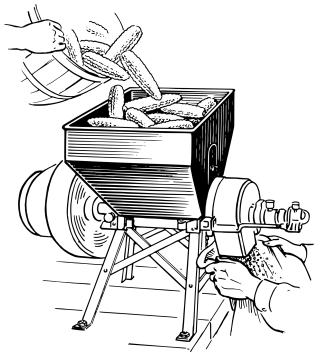


Distribute survey/
collect responses



Analyse responses

How does AI fit in this?



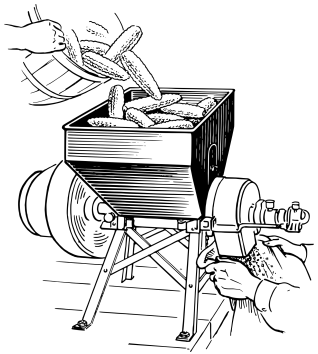
Writing questions

- AI-based writing tools
  - ChatGPT
  - Copy AI
  - Nichesss
  - NotionAI
  - Google Docs
  - Microsoft Teams Premium
  - ...

OpenAI

copy.ai

How does AI fit in this?



Writing questions

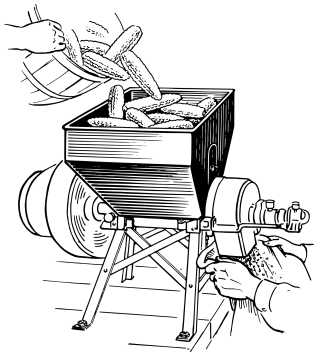
- AI-based writing tools
  - ChatGPT
  - Copy AI
  - Nichesss
  - NotionAI
  - Google Docs
  - Microsoft Teams Premium
- Online survey creators
  - SurveyMonkey
  - Survey Sparrow
  - Startquestion
  - …

SurveyMonkey®

SurveySparrow

Startquestion

How does AI fit in this?

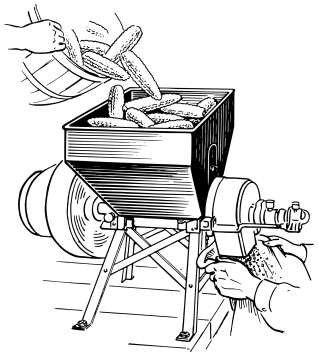


Writing questions

- AI-based writing tools
  - ChatGPT
  - Copy AI
  - Nichesss
  - NotionAI
  - Google Docs
  - Microsoft Teams Premium
- Online survey creators
  - SurveyMonkey
  - Survey Sparrow
  - Startquestion
  - ...

Use: Take prompts to generate questionnaire templates, detect and recommend question types based on question wordings, etc.

How does AI fit in this?



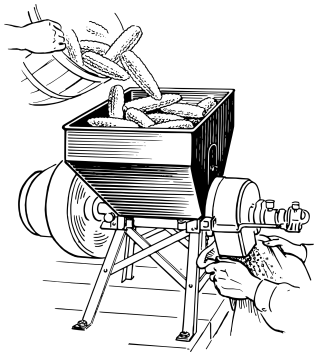
Writing questions

Testing survey

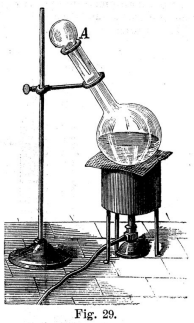
- Online survey creators
  - Issue detecting
  - Estimated response rate
  - …

Big data-enabled survey experience analyses

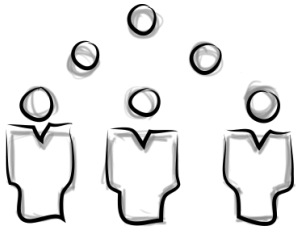
How does AI fit in this?



Writing questions



Testing survey

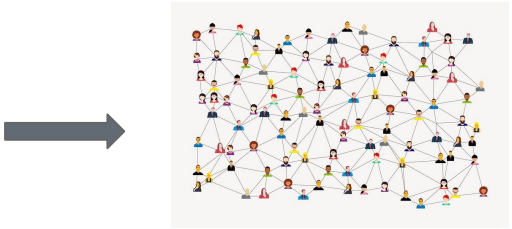


Distribute survey/
collect responses

How does AI fit in this?

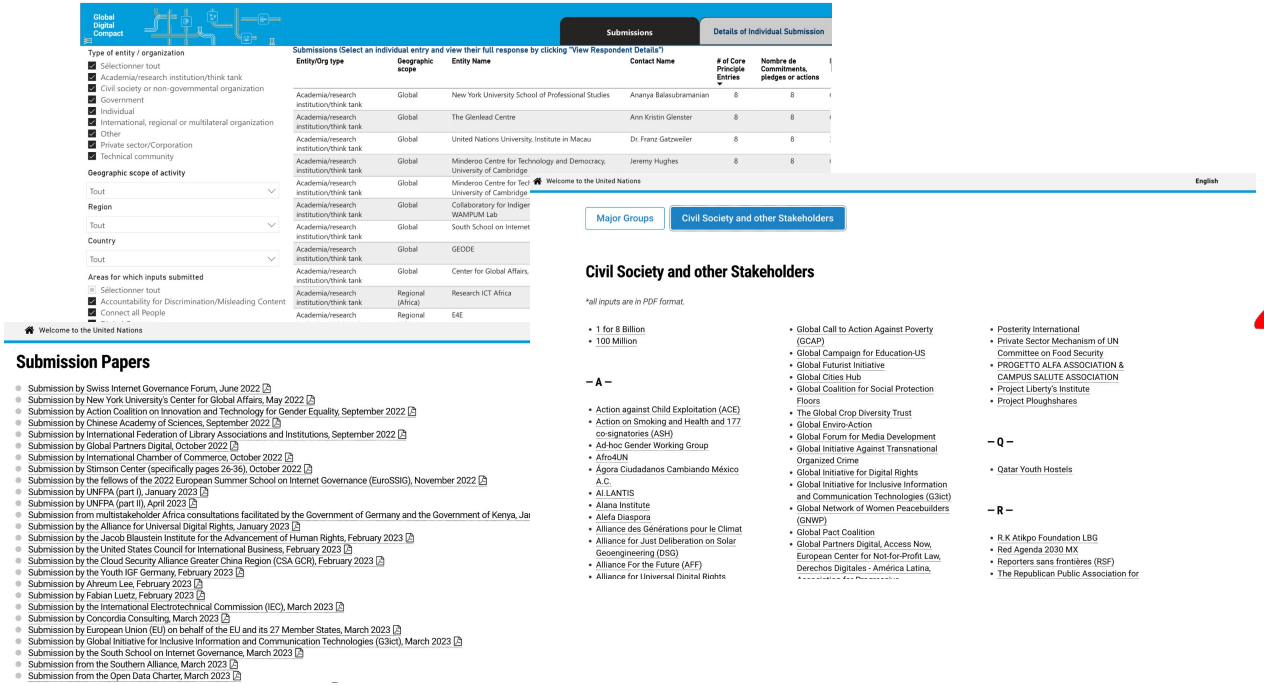
- Response summarisation:
  - Works well with open-ended responses
  - Tone detection
  - Key word finding

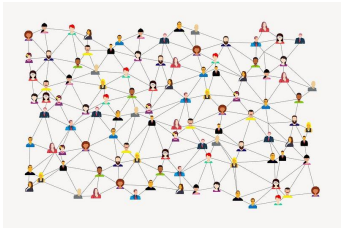
- Metadata analysis:
  - How long did participants spend on each question
  - Patterns in questionnaire answering
  - Trend analyses



Analyse responses

How does AI fit in this?





Analyse responses

How does AI fit in this?

Transparency

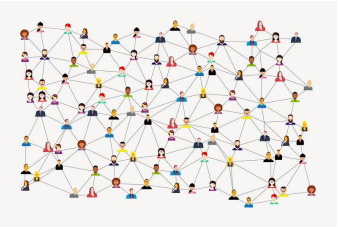
Attribution

Traceability

Agency

Contestability

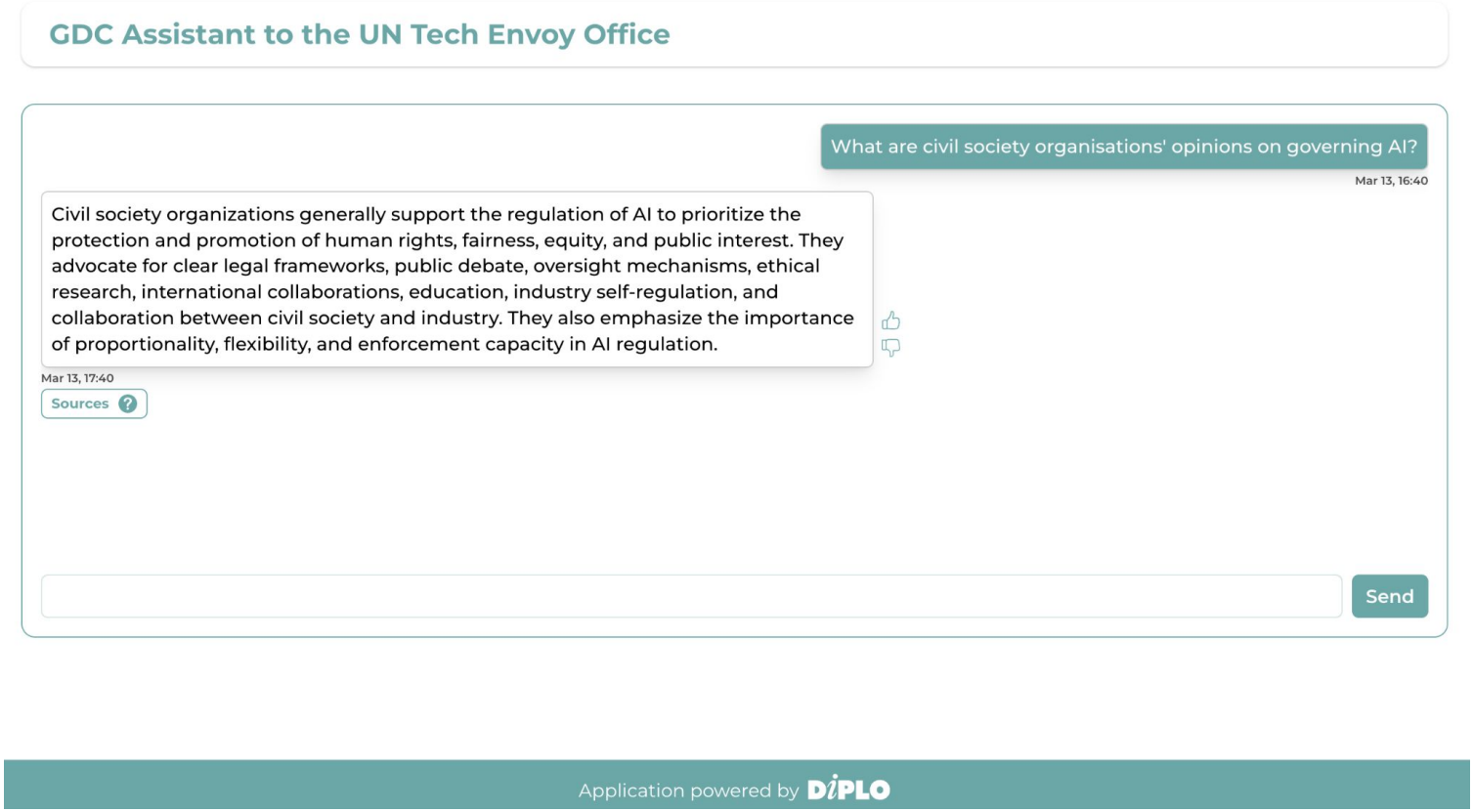
Accountability



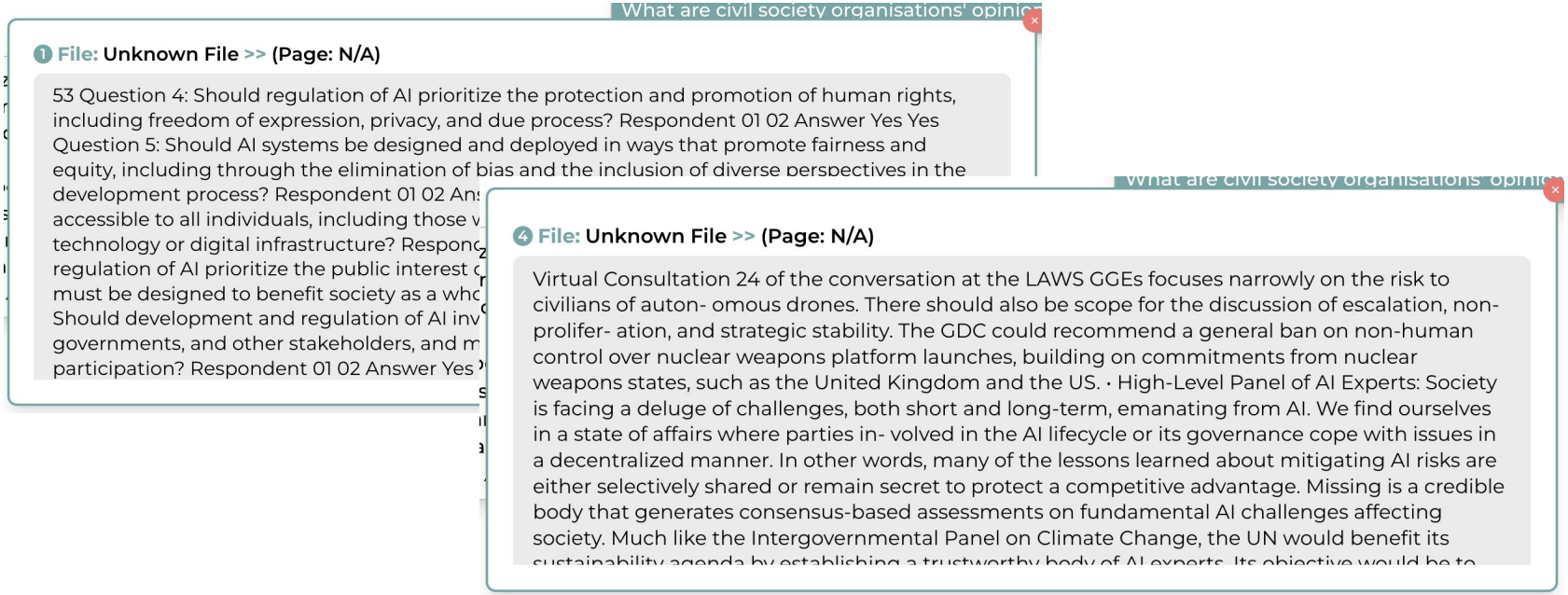
Analyse responses

Does AI decrease or increase transparency?

Is transparency technically possible?



Is transparency technically possible?



Is transparency technically possible?

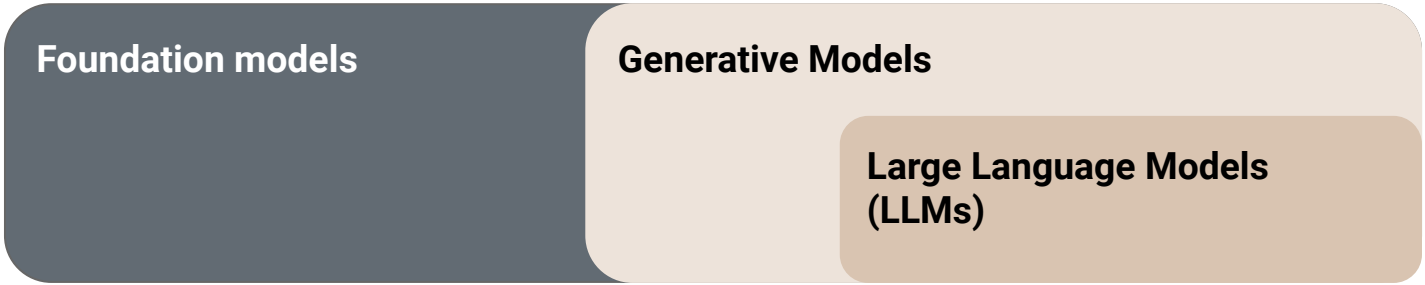
Foundation models
AI models that are trained on broad data and can be adapted to a wide range of tasks across different domains. They serve as the basis for more specialised models dedicated to specific tasks or domains.

Is transparency technically possible?

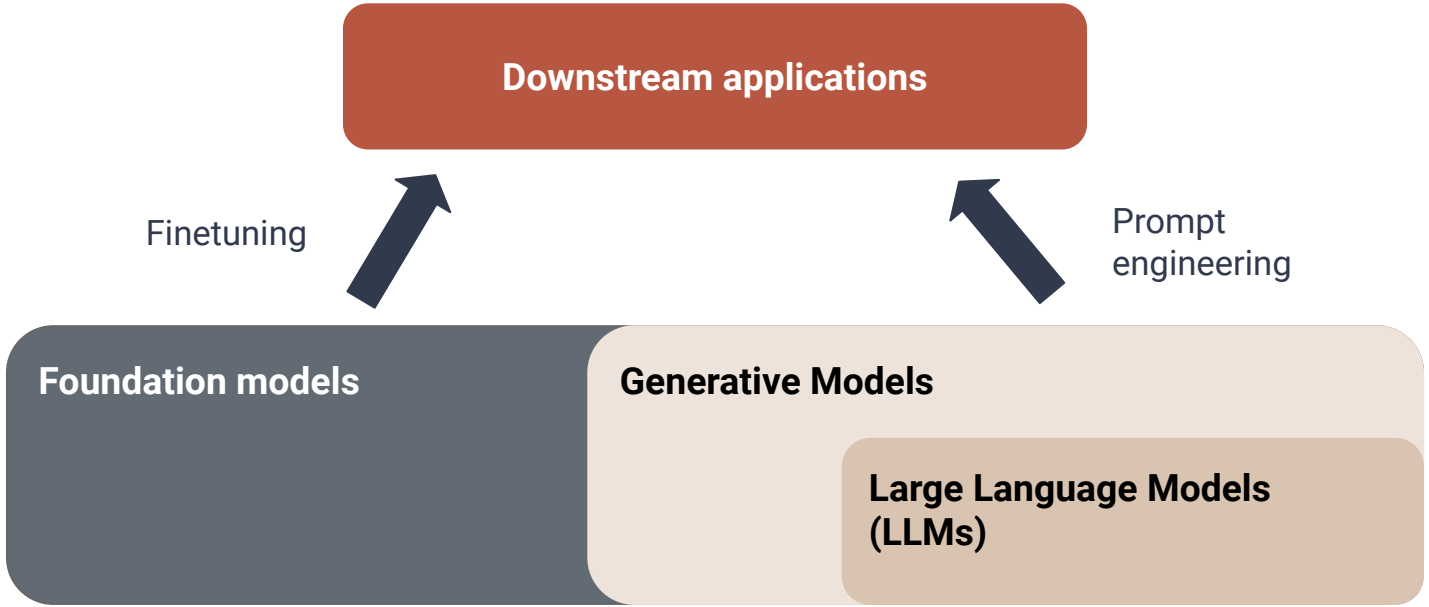
Foundation models

Generative Models
A type of deep learning models that could generate high-quality content (texts, audios, images, videos, etc.) that are not seen in the data

Is transparency technically possible?

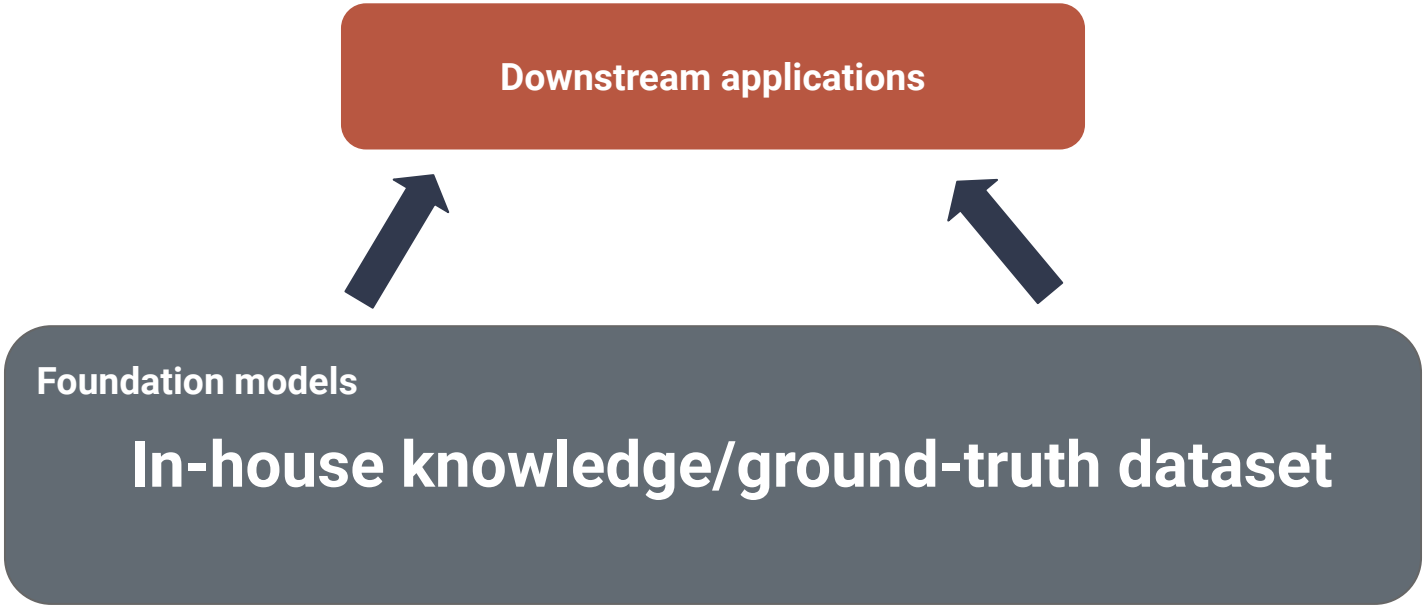


Is transparency technically possible?



Is transparency technically possible?

Q: What are civil society organisations' opinions on governing AI?

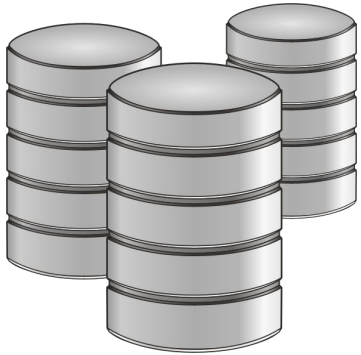


Is transparency technically possible?

- Retrieval-augmented generation (RAG)
  - Provide curated databases of documents for the LLM to reference before relying on in-house knowledge
  - More context, less hallucination

Is transparency technically possible?

Q: What are civil society organisations' opinions on governing AI?



Downstream applications

External database

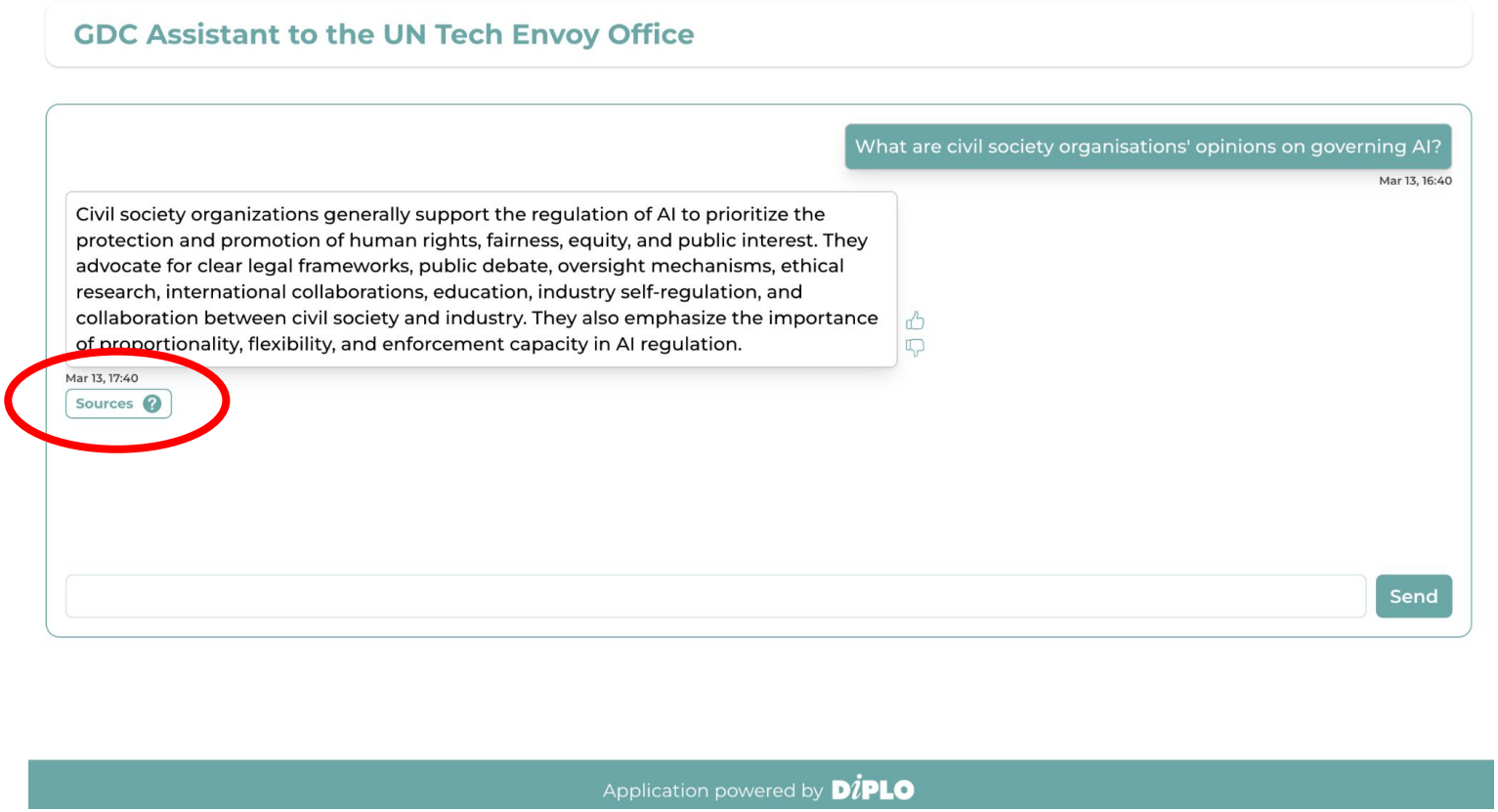
Foundation models

In-house knowledge/ground-truth dataset

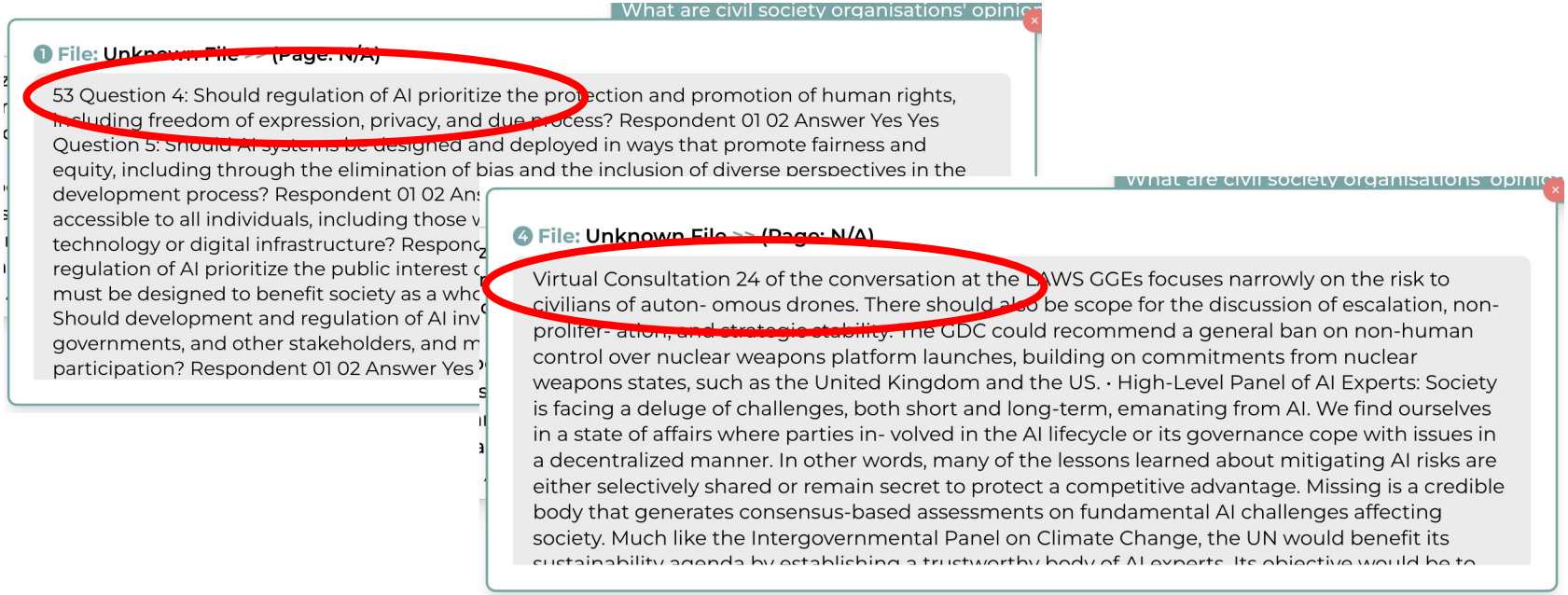
Is transparency technically possible?

- Retrieval-augmented generation (RAG)
  - Provide curated databases of documents for the LLM to reference before relying on in-house knowledge
  - More context, less hallucination
- Source attribution
  - Provide the source on which the AI model generates the answer improves transparency and traceability

Is transparency technically possible?



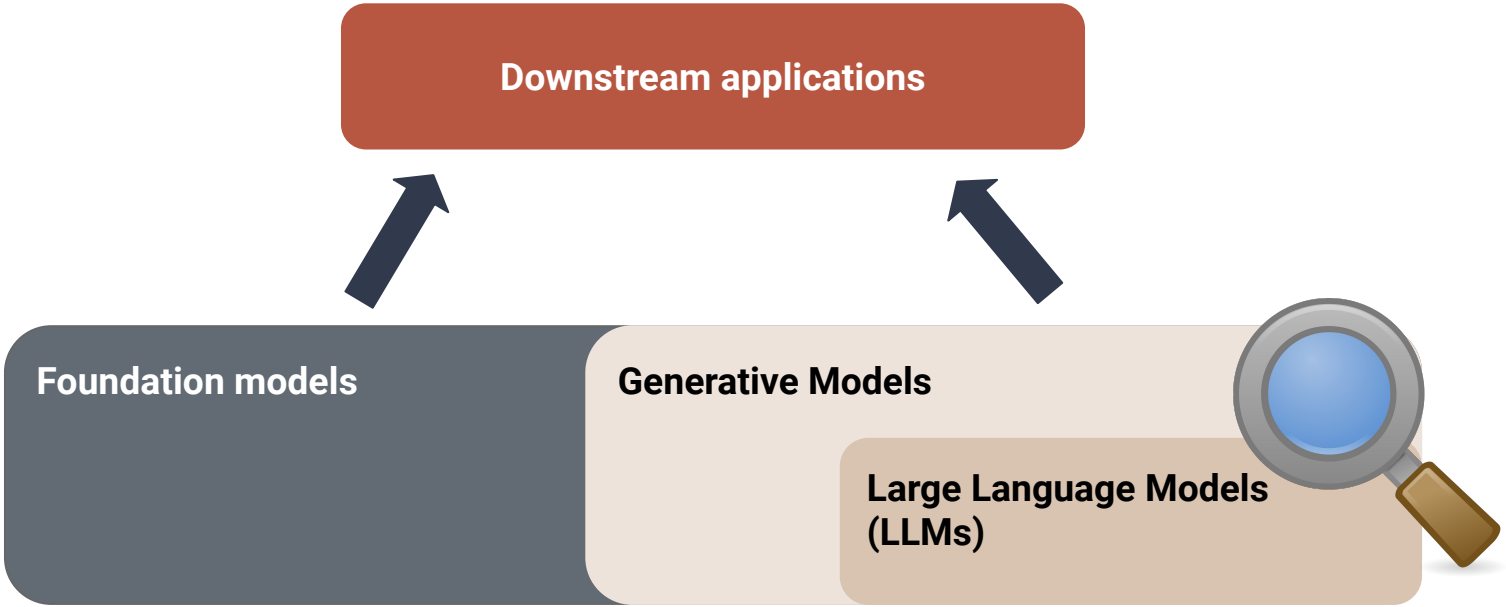
Is transparency technically possible?



Is transparency technically possible?

- Retrieval-augmented generation (RAG)
  - Provide curated databases of documents for the LLM to reference before relying on in-house knowledge
- Source attribution
  - Provide the source on which the AI model generates the answer improves transparency and traceability
- Open-source models
  - Allow better insights into the architecture of the AI models

Is transparency technically possible?



Is transparency technically possible?

Is it black boxed by technology...

Or by humans?

Thank you!

diplomacy.edu
dig.watch
humainism.ai

wu@diplomacy.edu

DiPLO Geneva Internet Platform