AI: the good, the bad, and the possibly amazing

Anita de Waard
VP Research Collaborations
a.dewaard@elsevier.com

Danske Fag-, Forsknings- og Uddannelsesbiblioteker

Should librarians be scared of (Generative) AI?

The Bad: AI can be biased:

Bias in **data** [Scheuermann et al]:

A particularly interesting example was PUBFIG, which had two gendered annotations: "male" and "attractive woman," of which there was no associated "female" [87]. The absence of annotations for "female" or "attractive man", however, highlights the culturally-situated values around gender that can emerge within an annotation schema (c.f., [150]).

Bias in **algorithms** [Amnesty International Report]:

Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal

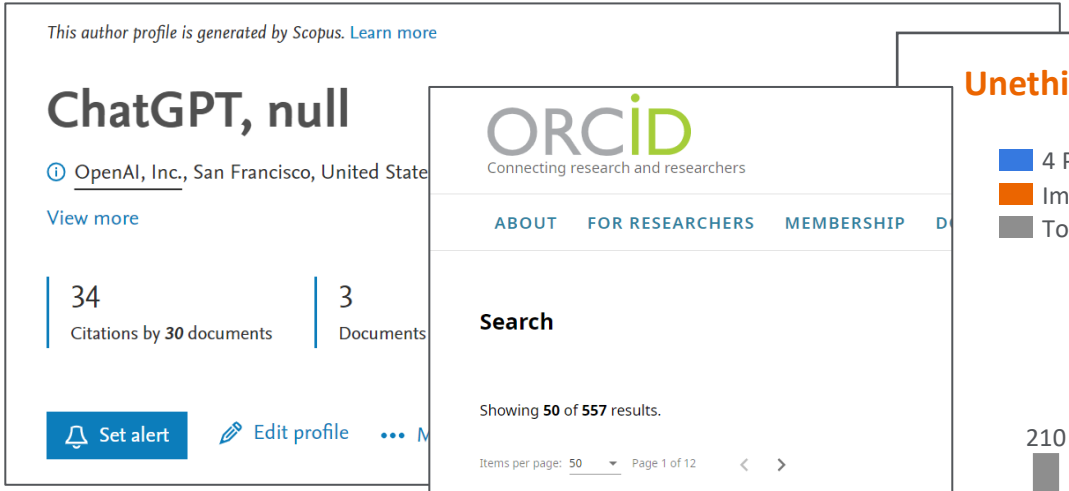
October 25, 2021, Index Number: EUR 35/4686/2021

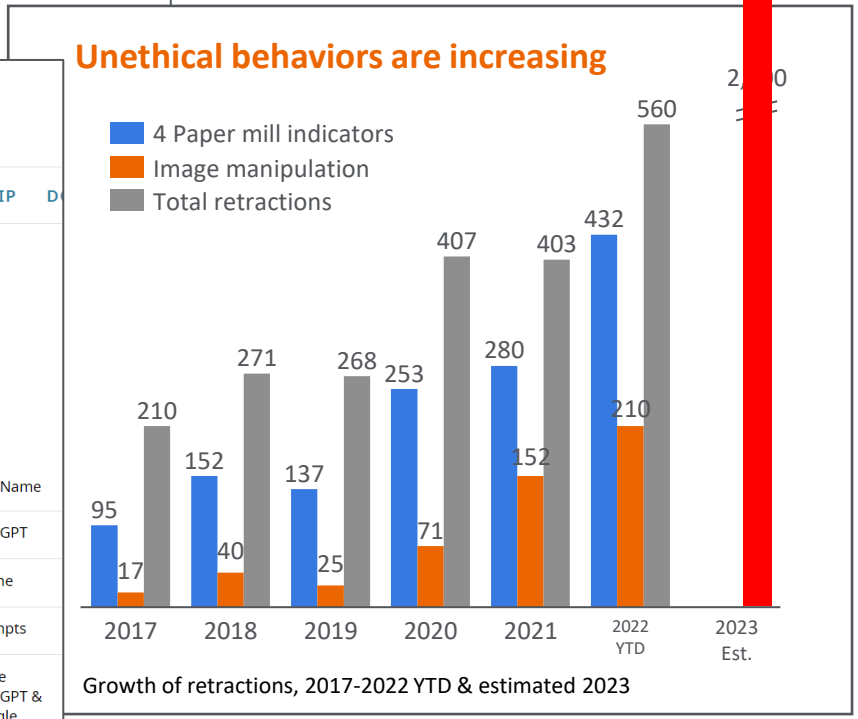
EUROPE AND CENTRAL ASIA TECHNOLOGY AND HUMAN RIGHTS

Social security enforcement agencies worldwide are increasingly automating their processes in the hope of detecting fraud. The Netherlands is at the forefront of this development. The Dutch tax authorities adopted an algorithmic decision-making system to create risk profiles of individuals applying for childcare benefits in order to detect inaccurate and potentially fraudulent applications at an early stage. Nationality was one of the risk factors

Scheuerman et al., "How We've Taught Algorithms to See Identity Constructing Race and Gender in Image Databases for Facial Analysis", ACM HCI, CSCW May 2020,https://doi.org/10.1145/3392866
"Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal" October 25, 2021 https://www.amnesty.org/en/documents/eur35/4686/2021/en/

The Bad: AI can be weaponized:

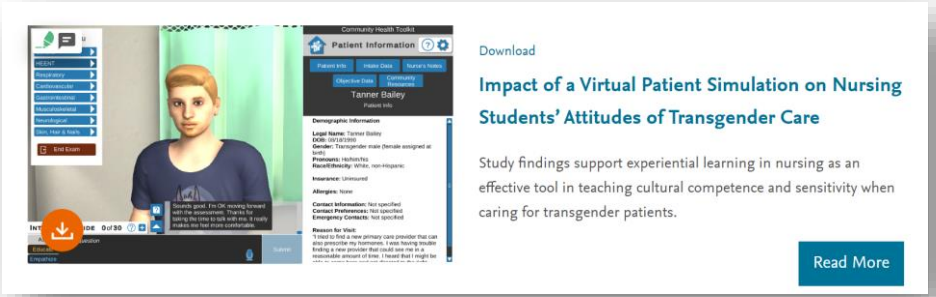




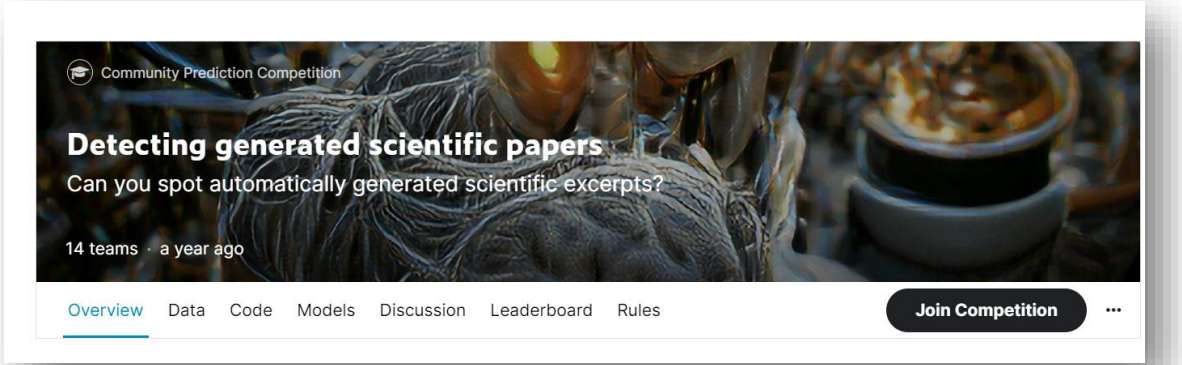
- https://www.the-geyser.com/chatgpt-says-its-not-an-author/ -- https://www.elsevier.com/about/policies/publishing-ethics/the-use-of-ai-and-ai-assisted-writing-technologies-in-scientific-writing
- Internal data but covered by Sarah Jenkins at https://www.stm-assoc.org/events/stm-research-integrity-master-class-2/

The Good: But AI can also do great things:

- AI can help protect patients and train nurses:

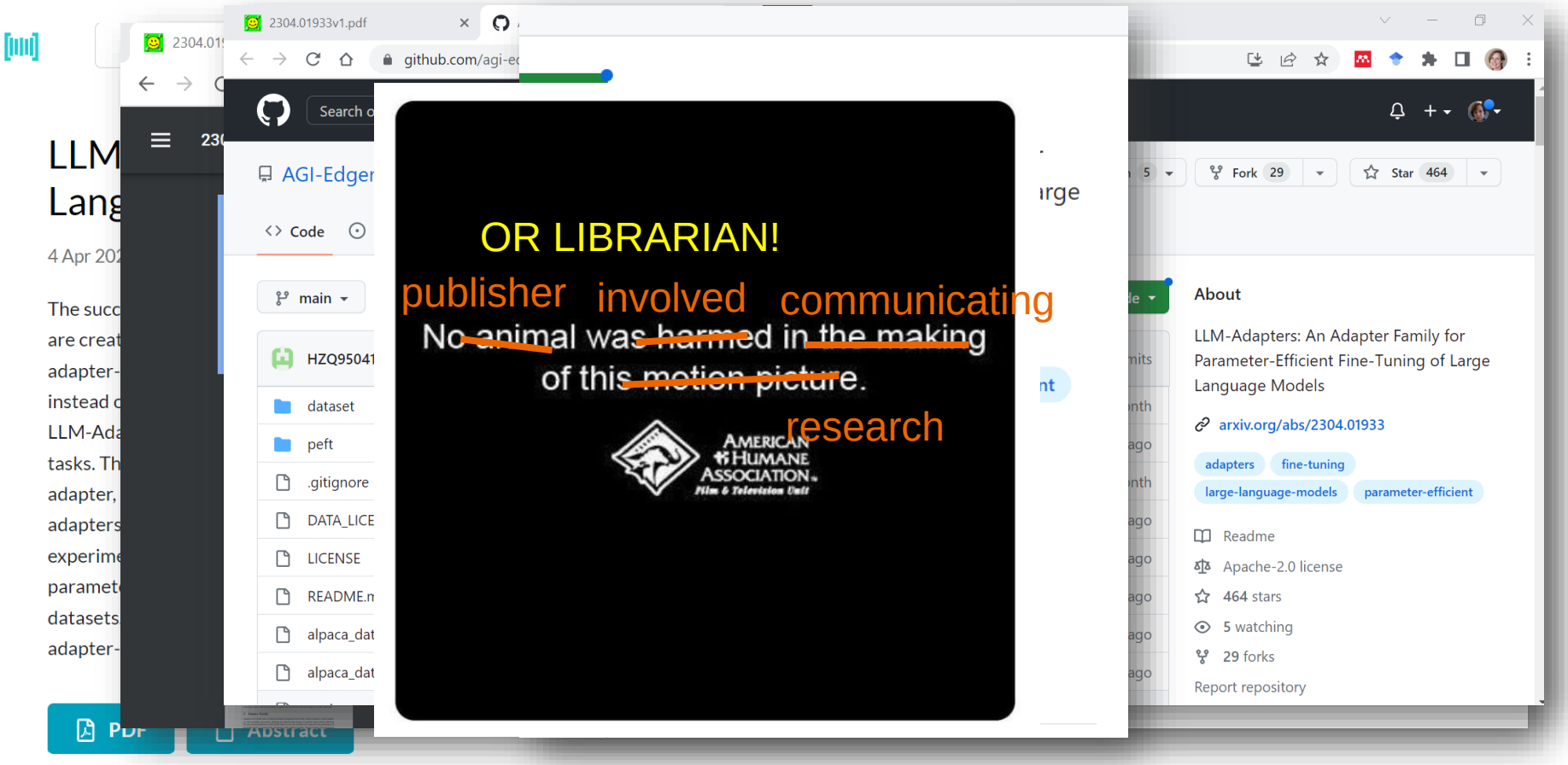


- AI can help detect AI-generated Content!



https://journals.lww.com/nurseeducatoronline/Citation/2023/05000/Impact_of_a_Virtual_Patient_Simulation_on_Nursing.6.aspx
https://aclanthology.org/2022.sdp-1.26/

How is AI research ~~published? shared?~~ used!



Paperswithcode: https://paperswithcode.com/paper/llm-adapters-an-adapter-family-for-parameter

Where does Elsevier stand, on all this?

Elsevier's AI author policy states that *authors* are allowed to use generative AI and AI-assisted technologies in the writing process before submission, but only to improve the language and readability of their paper and with the appropriate disclosure, as per our instructions in Elsevier's Guide for Authors.

Generative AI or AI-assisted technologies should not be used by editors to assist in the evaluation or decision-making process of a manuscript as the critical thinking and original assessment needed for this work is outside of the scope of this technology and there is a risk that the technology will generate incorrect, incomplete or biased conclusions about the manuscript.

Elsevier owns identity protected AI-assisted technologies which conform to the RELX Responsible AI Principles, such as those used during the screening process to conduct completeness and plagiarism checks and identify suitable reviewers.

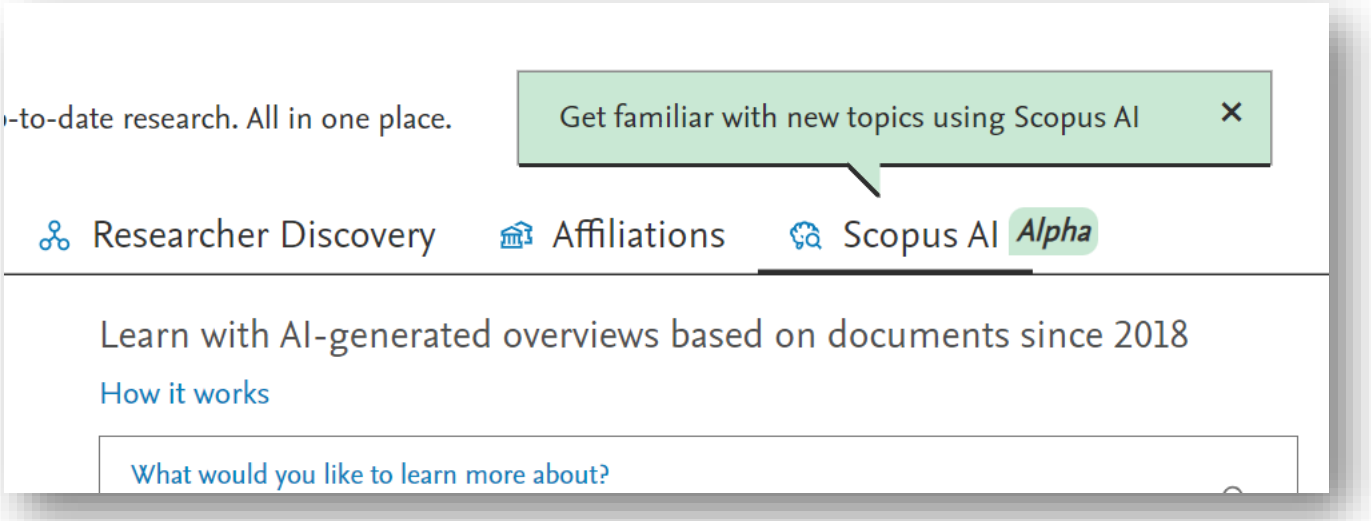
about the manuscript and/or the authors. For this reason, editors should not upload their letters into an AI tool, even if it is just for the purpose of improving language and readability.

Peer review is at the heart of the scientific ecosystem and Elsevier abides by the highest

https://beta.elsevier.com/about/policies-and-standards/publishing-ethics?trial=true

For instance: ScopusAI

A small demo



Vector Search

- Generate vectors quickly and efficiently, in under 300ms
- Results are of high quality and match user intent

Prompt Engineering

- Ensure outputs are not just accurate, but also meaningful
- Outputs are produced in usable formats (JSON)

Transparency

- Ensure every claim and statement is grounded in academic research
- All research can be traced back to source

Bringing the Responsible AI Principles in Practice

AS PART OF OUR RESPONSIBLE AI APPROACH...

1. We consider the real-world impact of our solutions on people
2. We take action to prevent the creation or reinforcement of unfair bias
3. We can explain how our solutions work
4. We create accountability through human oversight
5. We respect privacy and champion robust data governance

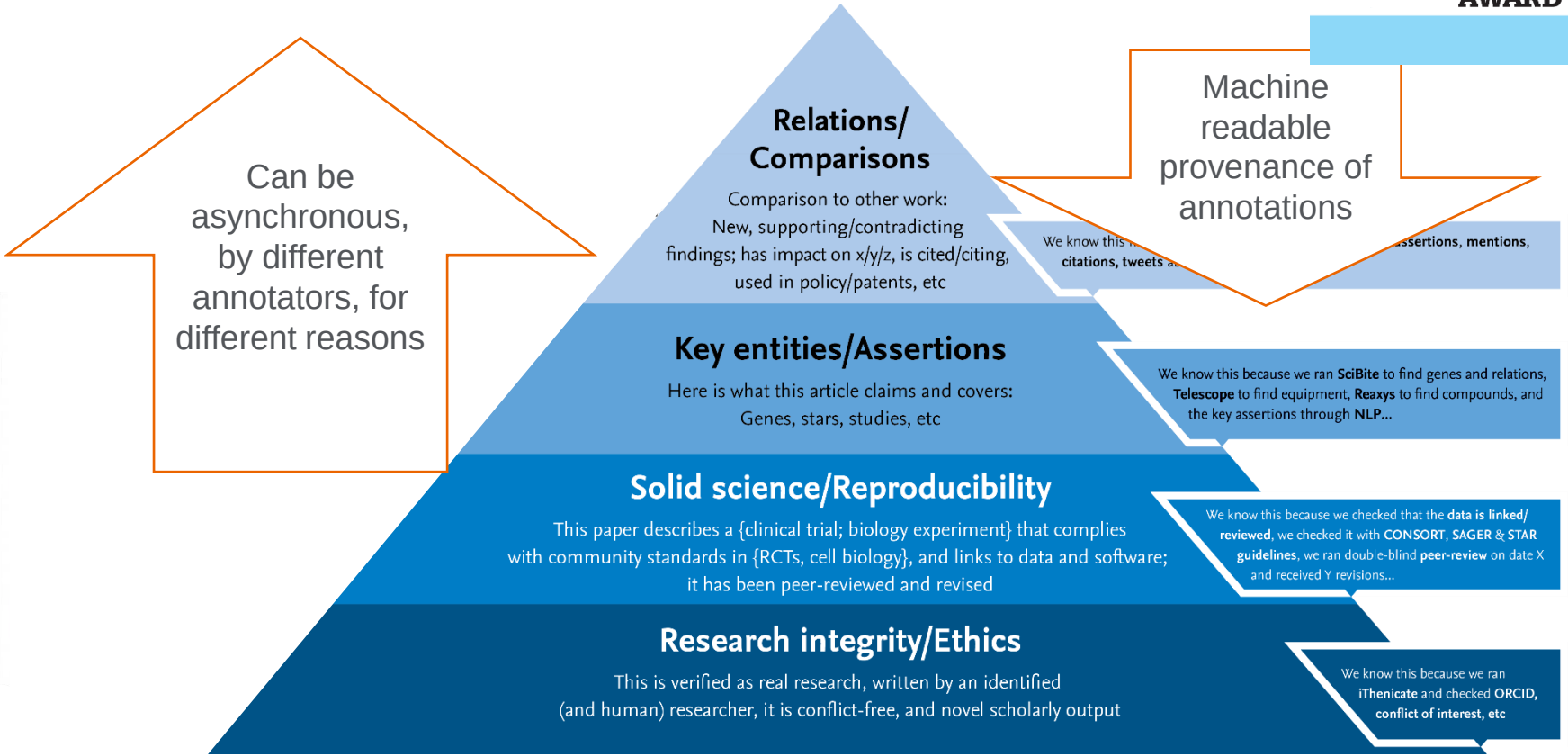
Generative AI output is **evaluated against sample data** before release

Ensure generative AI is **grounded** in **academic information** and **traceable**

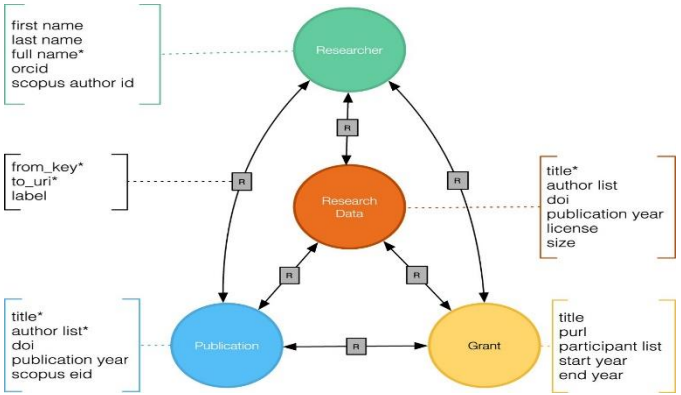
Observe Elsevier guidelines around personal data usage – GDPR

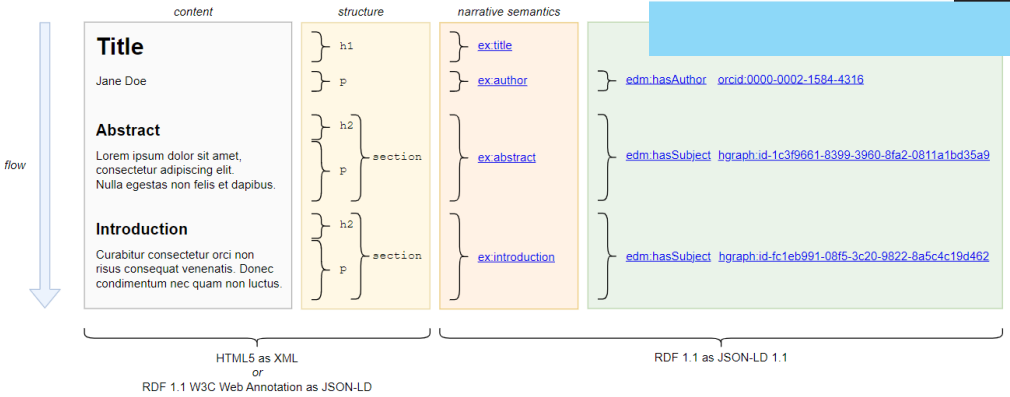
https://www.relx.com/~/media/Files/R/RELX-Group/documents/responsibility/download-center/relx-responsible-ai-principles-0622.pdf

Behind all this: the pyramid of trust



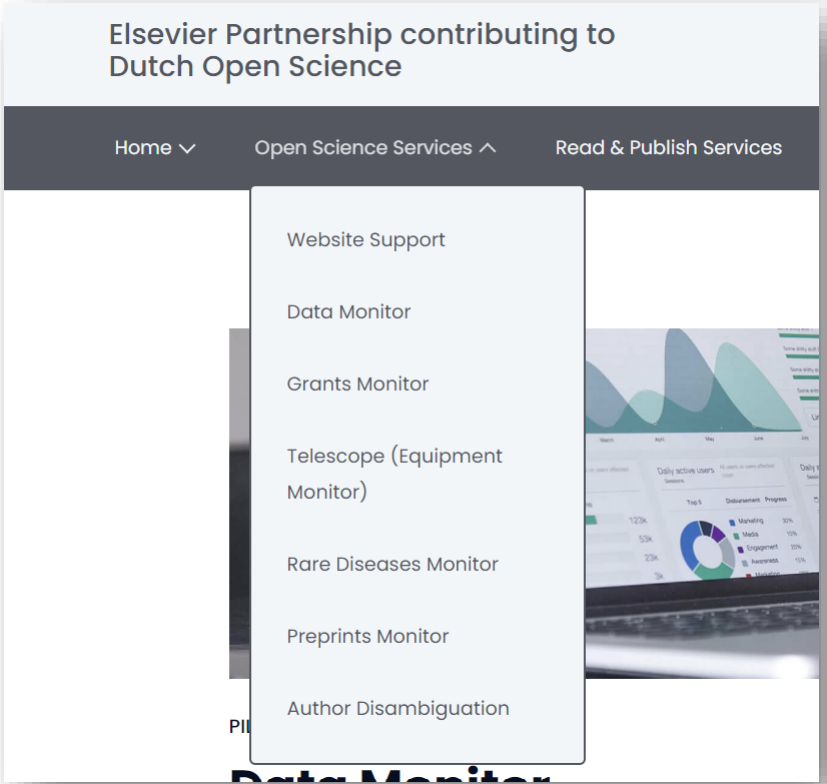
Behind that: Persistent Identifiers and Open Knowledge Gra

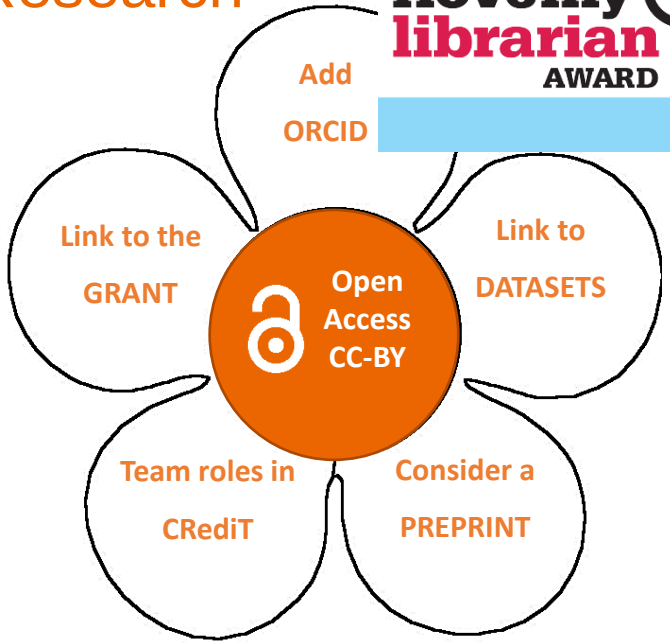




- 2010: ORCID: PIDs for authors
- 2011: Force11 Data Citation Principles: PIDs for data
- 2014: COPDESS Enabling FAIR Data: shared author's instructions
- 2014 FAIR Data Principles more than open, also interpretable data
- 2015: CrediT Taxonomy: roles of researchers
- 2017: Scholix leading to (2023) Open Science/Research Graphs for Fair Data
- 2018: Research Object Authoring tool: creating a linked data graph for an output
- 2022: Linked Document Standard: adding metadata as we go
- 2023: Peer review terminology: structured levels that can be reported
- 2023: NIST Research Data Framework: stakeholders and roles

This leads to helping Enable Open Research



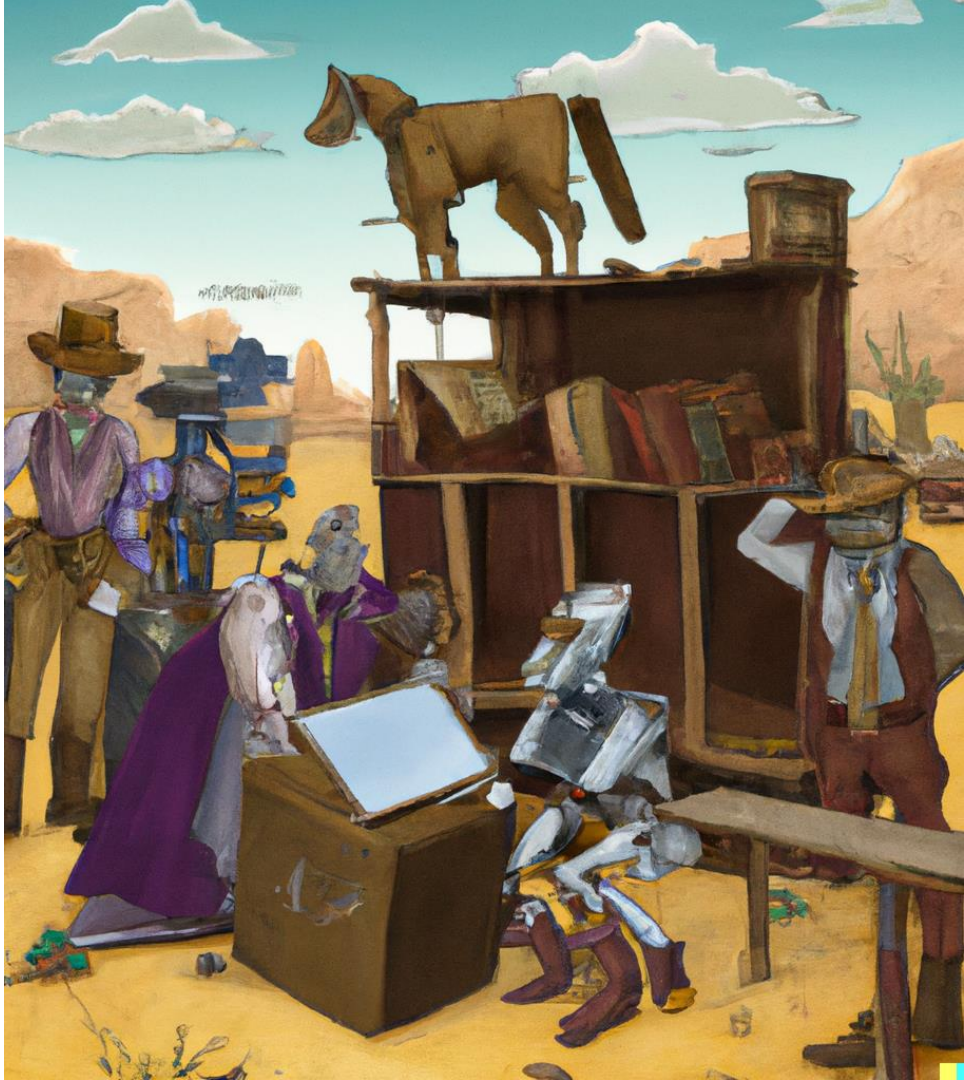


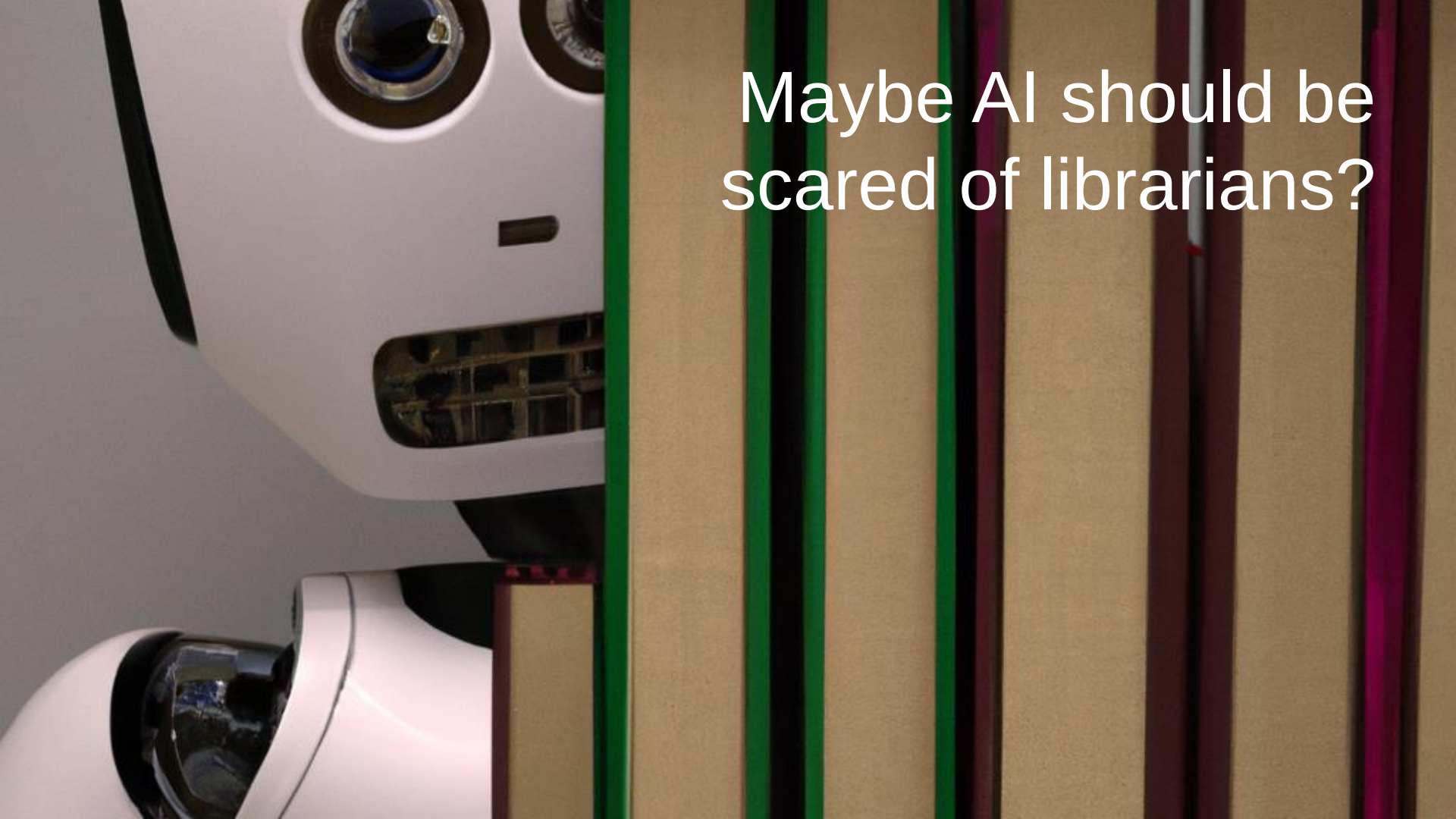
1. Use **ORCIDs** to find all outputs for each researcher
2. Find **DATASETS** wherever they are
3. Aggregate outputs per funder based on **GRANT IDs**
4. Acknowledge team members using **CRediT**
5. Find **PREPRINTS** wherever they are

More info about our trusted collaboration & principles in the Netherlands: **EPDOS.nl**

In summary:

- AI can be good, bad, or amazing
- It is important to be responsible when using AI
- To move forward we need to embed these technologies into a robust community infrastructure with:
  1. **Provenance** to enable verification
  2. **Persistent identifiers** for all components
  3. **Knowledge graphs** to connect them
- The way forward is **all together**: institutions, funders, publishers and **librarians**
- In other words:




Maybe AI should be scared of librarians?

Tak for jeres opmaerksomhed

Anita de Waard, a.dewaard@elsevier.com