

# What is AI and why does it matter for intangible cultural heritage?

**Bartosz Pielński**



UNIVERSITY  
OF WARSAW



# What is AI?

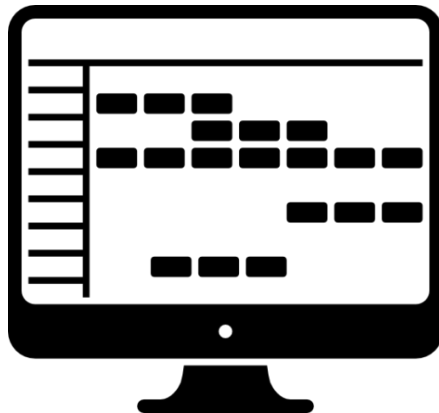
Algorithms



Nobel Prize  
2024



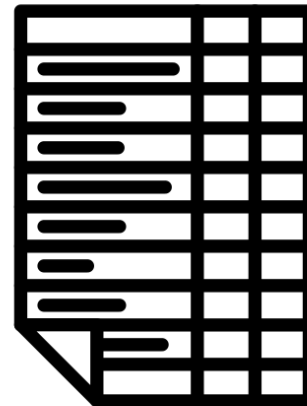
Computational  
power



LLM training  
1,300 MWh = 130 US homes



Data

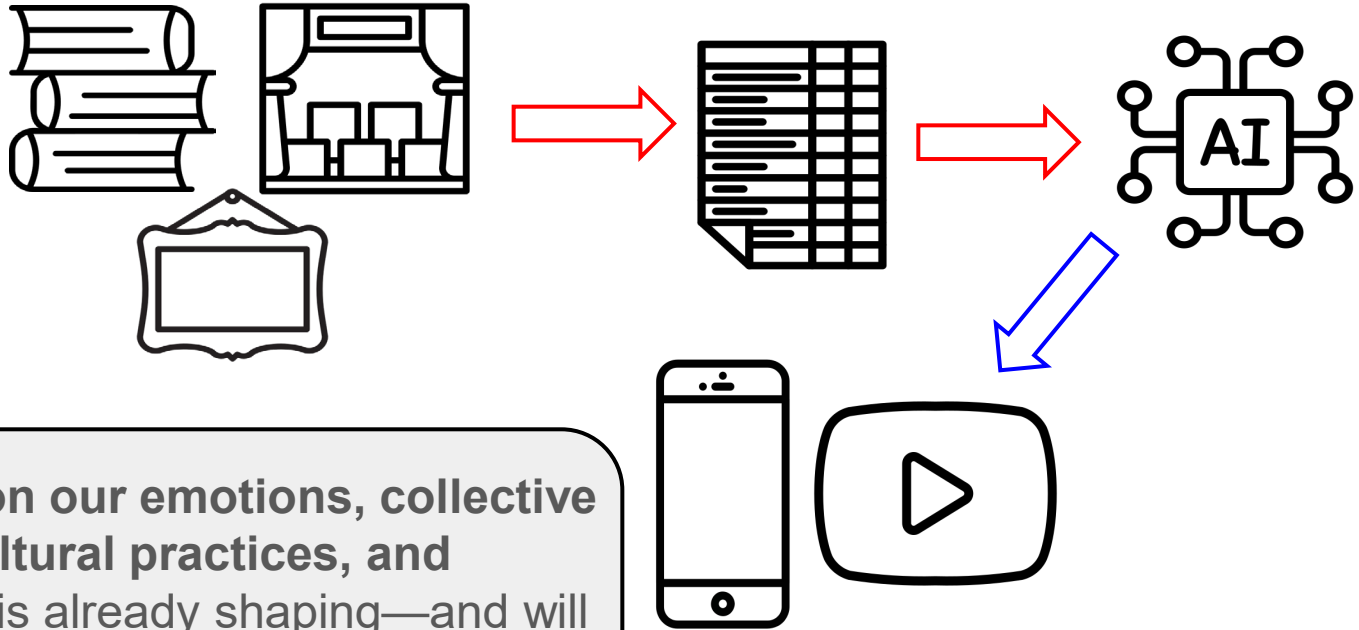


# How much data is required?

<b>Model</b>	<b>Token (Trillions)</b>	<b>Books (Approx.)</b>
GPT-3	0.3	3 billion
GPT-4	13	130 billion
Gopher	6.8	68 billion
Chinchilla	1.4	14 billion

Sources: <https://arxiv.org/pdf/2203.15556>;  
<https://www.kdnuggets.com/2023/07/gpt4-details-leaked.html>

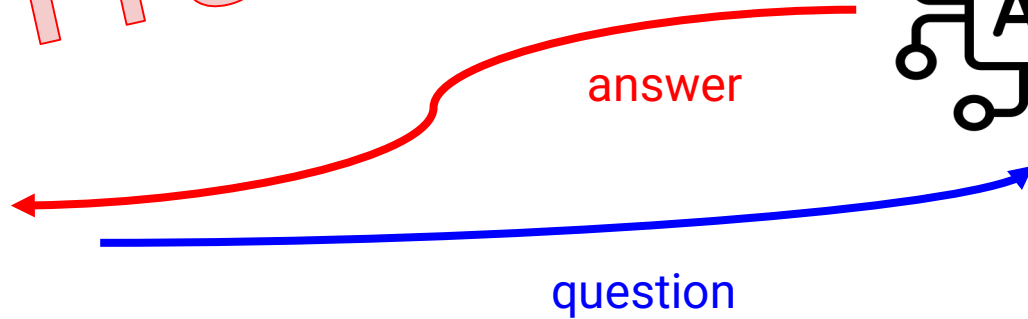
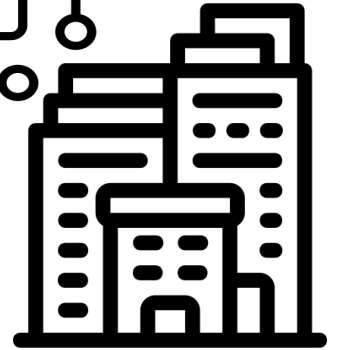
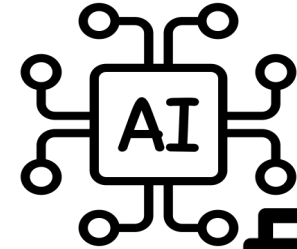
# The Importance of Data



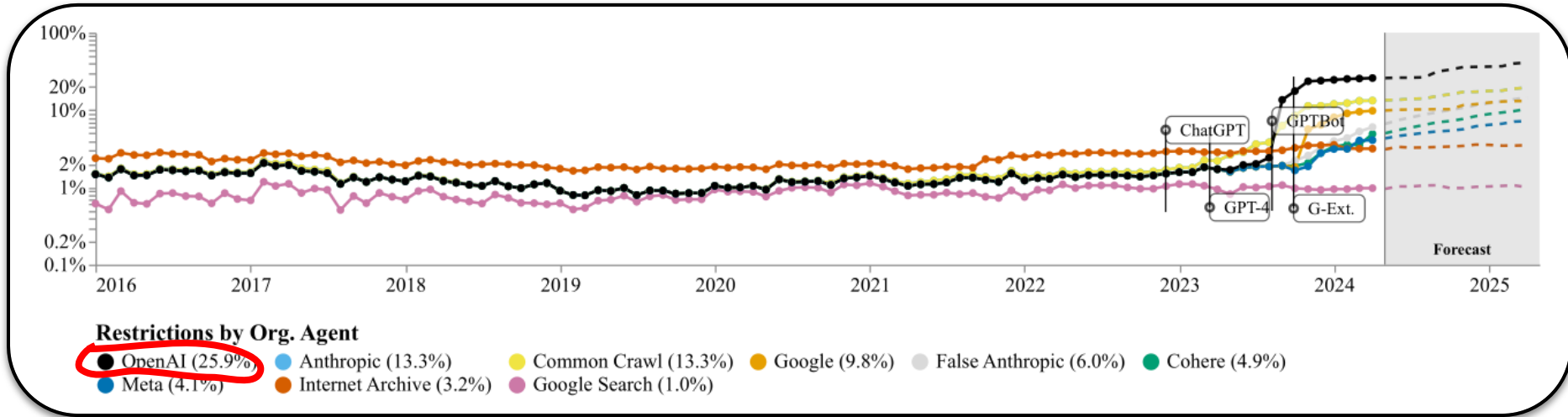
AI is built upon our emotions, collective memories, cultural practices, and **values**, and it is already shaping—and will continue to shape—these fundamental aspects of our existence.

# AI and data on ICH

Training Data



# Data scraping restrictions



Source: Longpre, S., Mahari, R., Lee, A., Lund, C., Oderinwale, H., Brannon, W., ... Pentland, S. (2024). Consent in Crisis: The Rapid Decline of the AI Data Commons. arXiv [Cs.CL]. Retrieved from <http://arxiv.org/abs/2407.14933>

# AI and ICH: some challenges and opportunities

## **Community as AI users:**

- AI can help communities innovate and transmit ICH.
- Risk of misappropriation and loss of attribution.

## **General users:**

- Make ICH more accessible and appealing;
- Misrepresentation of ICH.

## **AI training:**

- Bringing ICH to a broader audience.
- Unethical/unlawful use of data.

# Training data and languages

crawl		CC-MAIN-2024-30	CC-MAIN-2024-33	CC-MAIN-2024-38
language ↕	% ▾	% ↕	% ↕	% ↕
eng	English	43.7909	43.1787	44.1210
rus	Russian	6.1035	6.2242	6.1556
deu	German	5.2890	5.0895	5.4471
zho	Chinese	5.1358	5.1917	4.6266
jpn	Japanese	4.7980	4.8790	5.1119
spa	Spanish	4.5593	4.5819	4.4769
fra	French	4.2211	4.2439	4.4292
<unknown>		3.2009	3.2930	2.6706
ita	Italian	2.6077	2.4921	2.5224
por	Portuguese	2.1237	2.2179	2.2141
nld	Dutch	1.8216	1.7821	1.9149
pol	Polish	1.7963	1.9342	1.8868

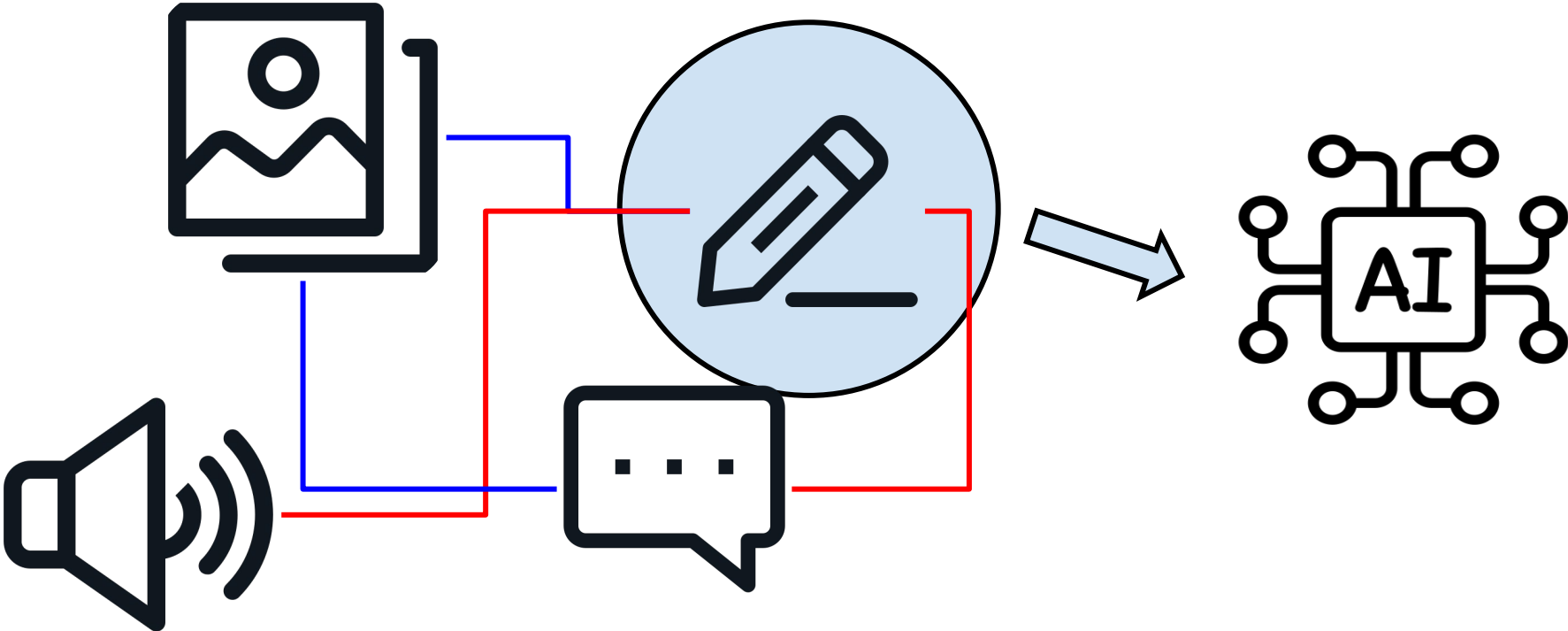
Over **73%**  
European  
languages

<https://commoncrawl.github.io/cc-crawl-statistics/plots/languages>

Bartosz Pielniński | 17.11.24 | AI&ICH



# Specificity of ICH



# Thank you!

We can't avoid AI, but we can and should guide it to safeguard our living heritage.

**Bartosz Pielński**

University of Warsaw

[www.pielinski.info](http://www.pielinski.info)

