

AI AS A BLACK SWAN or AI AS A TOOL?

Detecting misinformation and disinformation with AI

Dr. Petra Aczél, MOME

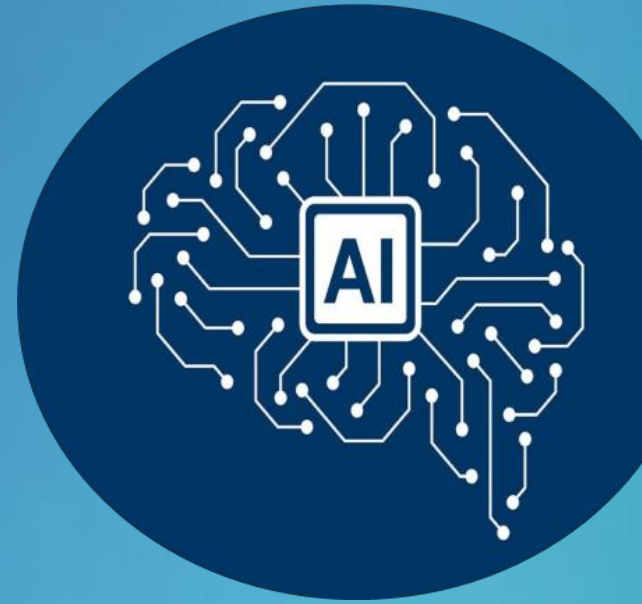


KNOWN KNOWNS

KNOWN UNKNOWN

UNKNOWN KNOWN

UNKNOWN UNKNOWN



How high was the number of visits of ChatGPT in November 2022? And in June 2024?



Which reached 1M in visitors first, ChatGPT or Threads?

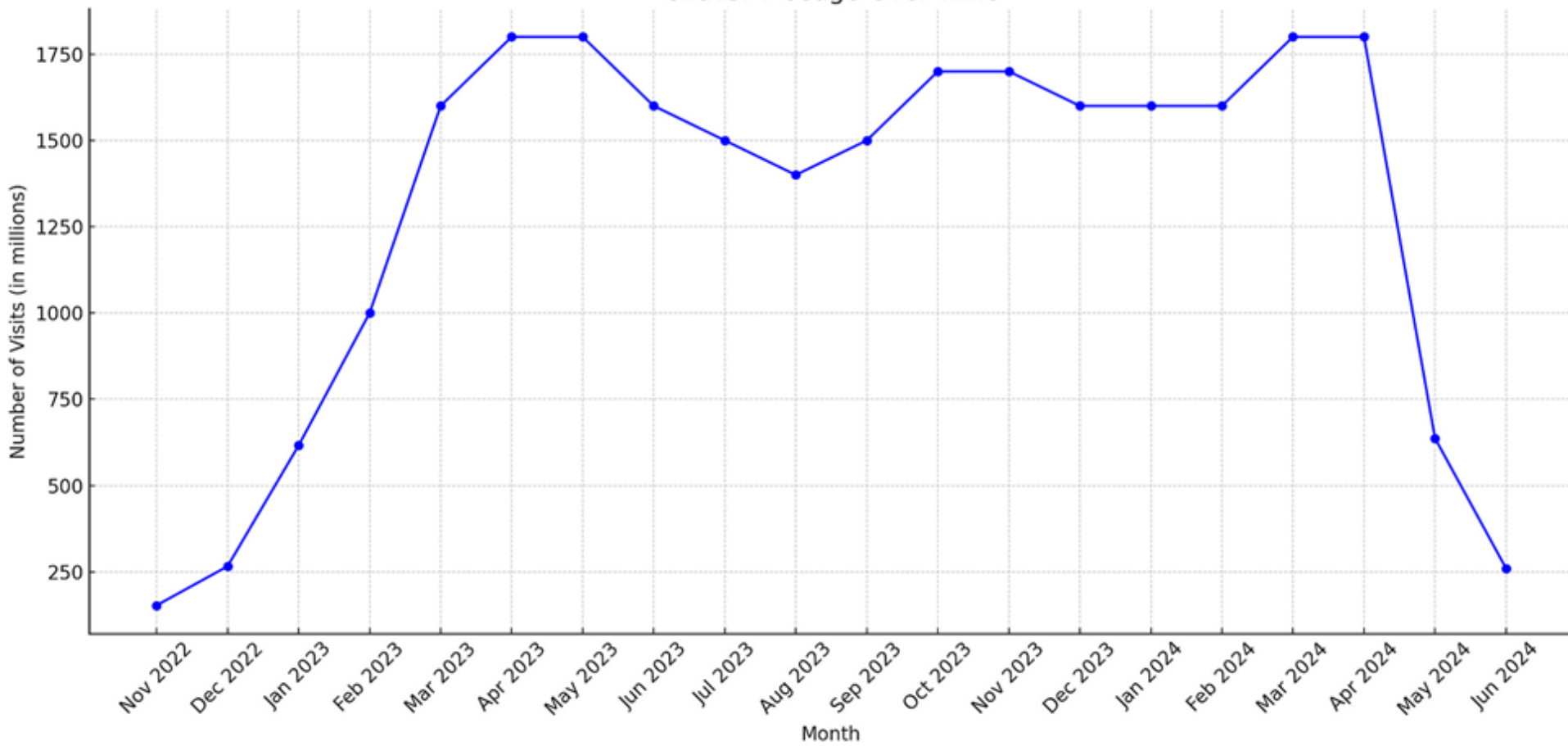
Humans or AI are/is more sensitive in moral decisions?

What is AI most used for by companies?

Which age group tends to be more AI-conscious, those between 15-29 or 30-49?

Do you personally consume media in a filter bubble?

ChatGPT Usage Over Time





delfi

trend-

analysis

horizon

scanning

foresighting

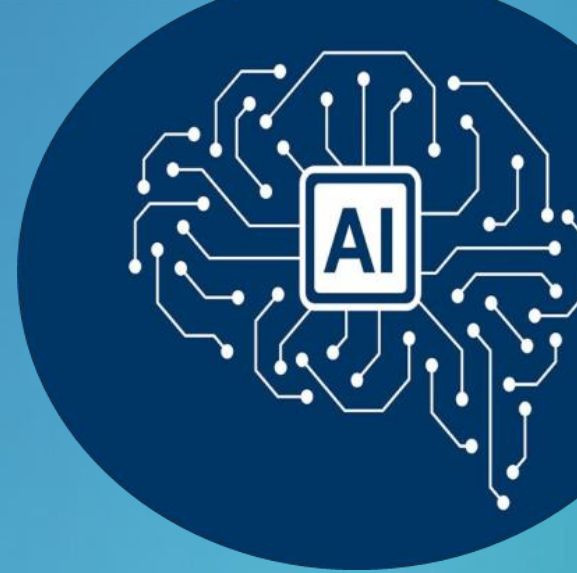
01.
Utopian

02.
Apocalyptic

03.
Enigmatic

04.
Pragmatic

CAN AI BE AVOIDED?



CAN FAKE/FALSE

INFORMATION BE AVOIDED?



DISCUSSION:

What is the difference between misinformation and disinformation?



What are the methods of faking and the categories of fake content?

How good are people at detecting lies in messages? What are the factors this skill depend on?

What media resources do people generally-globally gain information from?

How many of our 'everyday' digital devices use AI (percentage)?

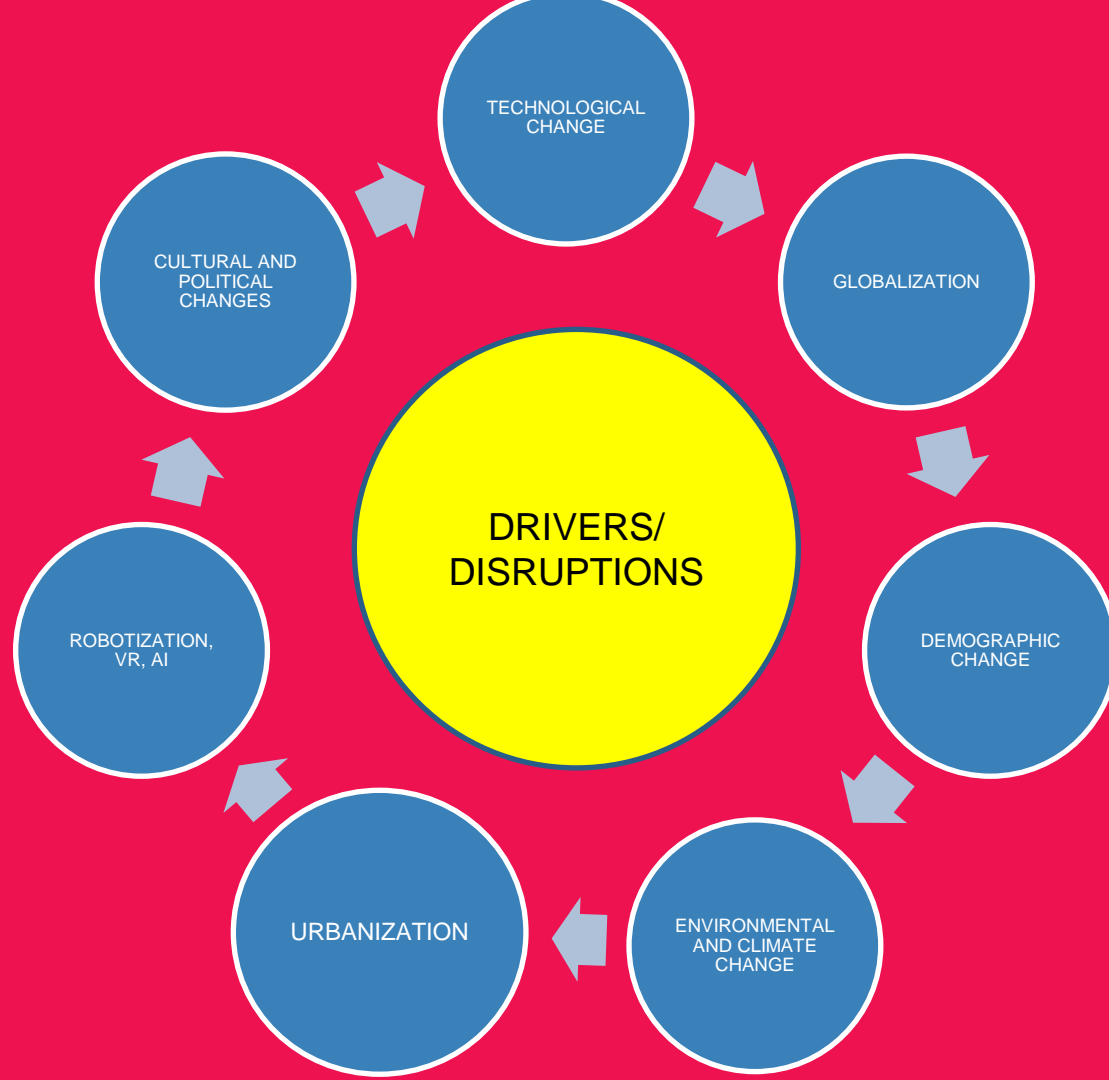
How active are people to check information using different resources themselves?

Does good fake detecting entail skills to create fake news?

What are the most fake-prone topics?

“

GLOBAL RISKS?



IFF 2011, 2016
Pearson, 2018, 2020,
WEF 2015, 2020,
McKinsey 2021

WEF 2024, Global Risks Perception Survey:

FIGURE B Current risk landscape

Please select up to five risks that you believe are most likely to present a material crisis on a global scale in 2024.

Risk categories

- Economic
- Environmental
- Geopolitical
- Societal
- Technological



1st
Extreme weather



2nd
AI-generated
misinformation
and disinformation



3rd
Societal and/or
political polarization



4th
Cost-of-living crisis



5th
Cyberattacks

Source

World Economic Forum Global Risks
Perception Survey 2023-2024.



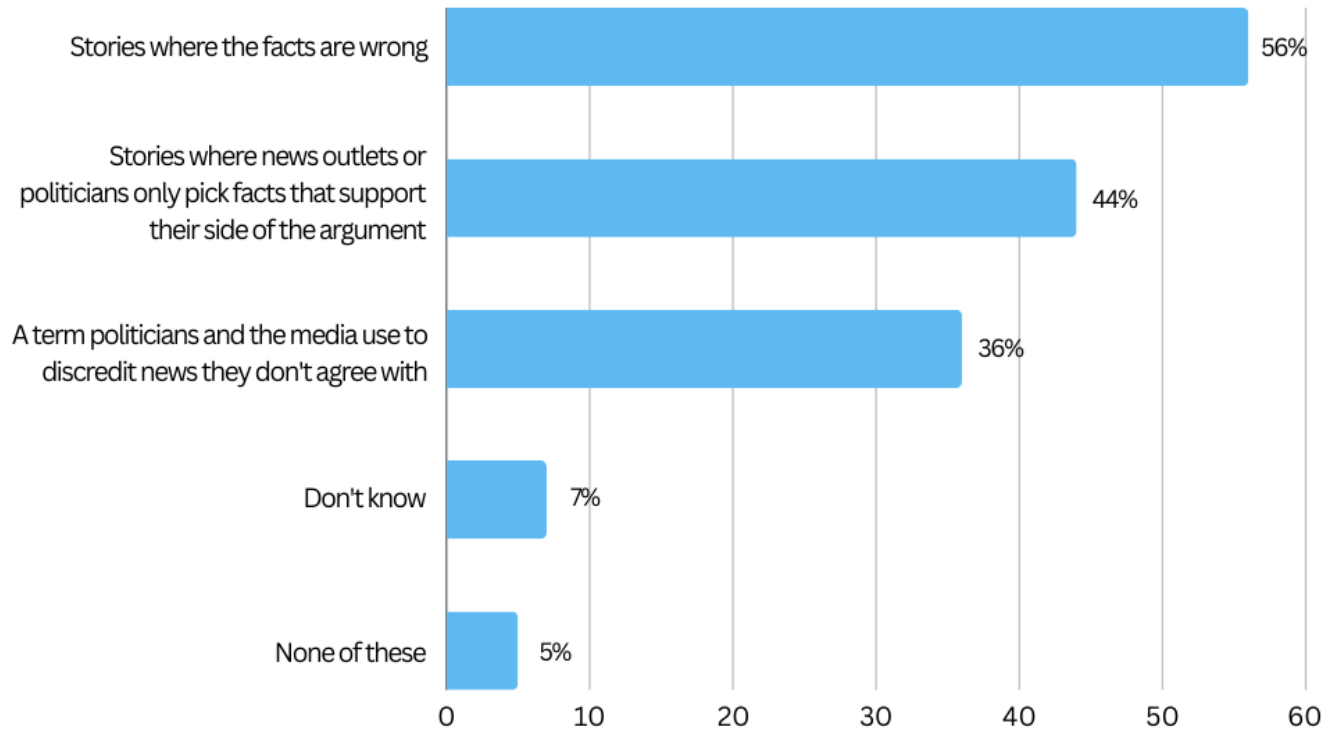
Withholding

- Silence
- Decontextualization
- Omission, erasure
- Non-references

Falsification:

- Exchange of facts
- Pseudo experts
- Wrong references
- Biased language use, wording

Respondents' Perceptions When They Hear the Term "Fake News"



Base: 13,000 adults across 26 countries



- There were 3x more video deepfakes and 8x more voice deepfakes in 2023 vs. 2022.

- In 2024 about 500,000 total video and voice deepfakes will be shared on social media globally .

- Free deepfake-programs are used for revenge-porn videos

DEEPPFAKE

- A hyper-realistic digital copy of a person that can be manipulated in all respects (Hughes et al. 2021)
- Deepfake is a special type of synthetic media where a person in a picture or video is replaced by the image of another person (Somers 2020)
- Manipulated or synthetic audio-visual media created by machine learning and deep learning AI that appears to be authentic and in which the persons appearing say or do what they did not say/do in reality (European

The Factual

content



Fabula AI

spreading

Compass by Blackbird AI

context

Grover AI

imitation

**FUTURE
7 principles
(Weick
2001)**

1. Uncertainty: insufficient facts – insufficient questions.
2. Fewer experts – more novices.
3. Decisions: more on those with the expertise to handle them, and less on people entitled by rank to make them.
4. Fewer attempts to capture the big picture and more attempts to capture the big story.
5. More focus on updating and plausibility and less on forecasting and accuracy.
6. More improvisation and fewer routines.
7. More humility and less hubris.