RESPONSIBLE AI

AI is becoming integral to sectors like education, finance, and healthcare, enhancing decision-making but also raising concerns about risks. Here are four key areas of responsible AI (The AI Index Report, 2024).

PRIVACY TRANSPARENCY SECURITY FAIRNESS

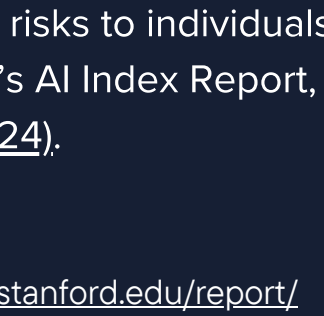
PRIVACY & SECURITY

A global survey on responsible AI reveals that companies prioritize concerns like privacy, security, and reliability, and are starting to address these issues, but most have only partially mitigated the risks.

TRANSPARENCY

Research indicates that the outputs of popular large language models (LLMs) often include copyrighted content like news articles or movie scenes, raising important legal questions about copyright violations.

ORIGINAL MIDJOURNEY



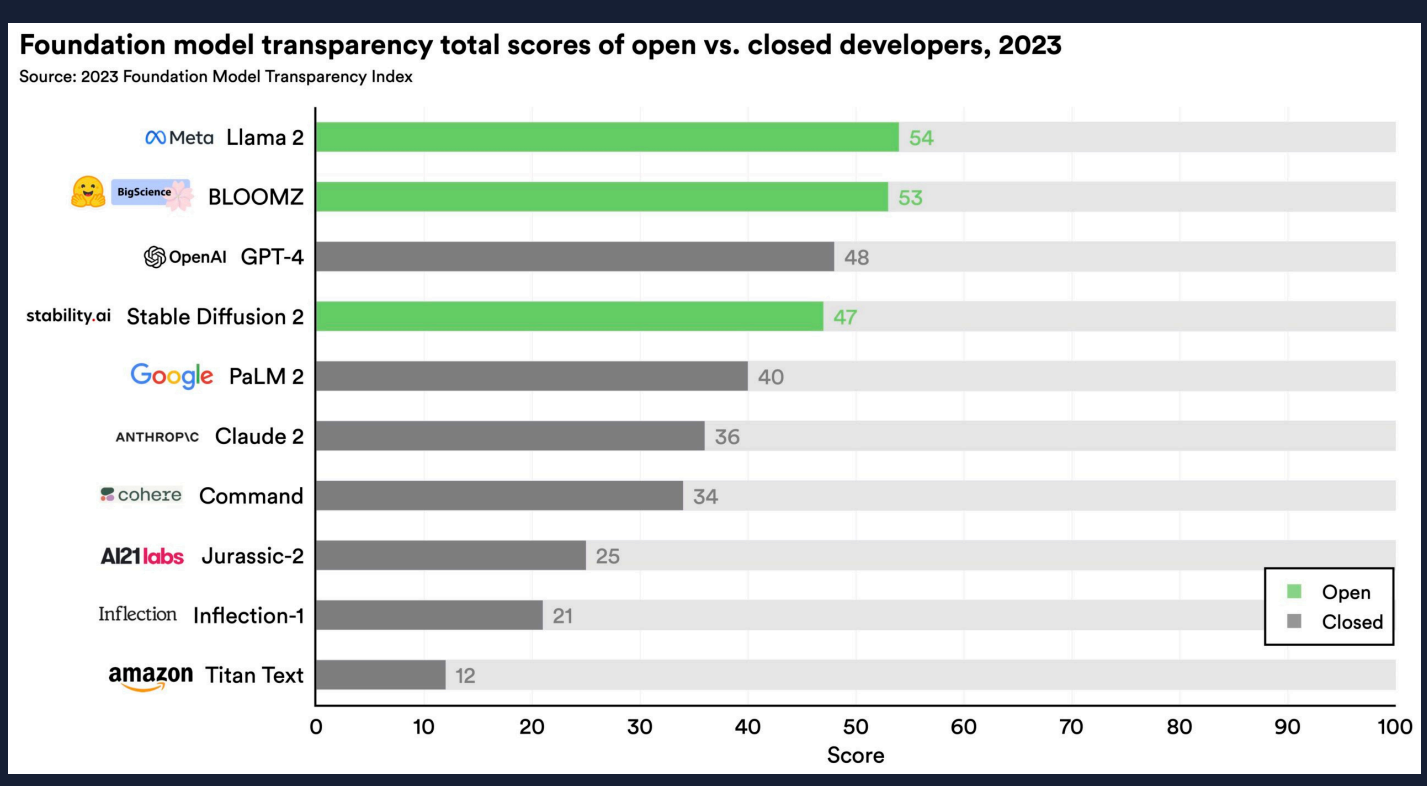
Midjourney generated a nearly identical image of the original movie scene featuring Thanos.

```
Thanos infinity war, 2018,
screenshot from a movie, movie
scene, 4k, bluray --ark 16:9 --v 6.0
```

Marcus & Southen (2024)

TRANSPARENCY

The new Foundation Model Transparency Index reveals that AI developers often fail to disclose training data and methodologies, impeding the understanding of AI system robustness and safety.



FAIRNESS

Researchers have identified a significant bias in ChatGPT favoring Democrats in the U.S. and the Labour Party in the U.K., raising concerns about its potential to sway political views during major election years.

Raising awareness toward responsible AI practices is a start to building trust and mitigating potential risks to individuals and society. For more information, see Stanford University's AI Index Report, Chapter 3: Responsible AI. Text and Analysis (Reuel, 2024).

Source: https://aiindex.stanford.edu/report/

HAWAIʻI CENTER FOR AI