AI search's high costs could be a vicious cycle as Big Tech eyes profitability

Article

The data: The generative AI-powered search rivalry comes at a steep cost.

- Training **GPT-3**, the AI model underlying **ChatGPT**, required **1,287 MWh** of energy and **contributed over 550 tons of CO2 emissions** to the environment, per Wired.

- For context, the typical car emits [4.6 tons of CO2 annually](#), **so it would take almost 120 years for the emissions to match that of AI model training**.

- **Powering search with generative AI uses at least four to five times more computing power** than standard search, according to **QScale** cofounder **Martin Bouchard**. He says current data center infrastructure won't be able to cope with the demand.

- Integrating the technology into search has significant energy and emissions implications— [ChatGPT](#) has about **13 million** users per day, according to UBS data. **Microsoft Bing** crunches half a billion searches daily and **Google 8.5 billion**.

  Why it could backfire: [Microsoft](#), Google, [Baidu](#), and **Opera** are making AI-powered search available to consumers. The problem is that the associated energy costs and carbon emissions add to the [litany of generative AI's problems](#).

- Widespread reports of AI chatbot errors and limitations means companies will be **steadily training new models and retraining existing ones**.

- With data centers already contributing 1% of the world's greenhouse gas emissions, according to the IEA, we can expect generative AI will add pressure to political controversy around [tech infrastructure expansion in Europe](#) and elsewhere.

- The technology could find itself in the crosshairs of a global energy crisis exacerbated by war and natural disasters and [could contribute to cloud outages](#) during heatwaves.
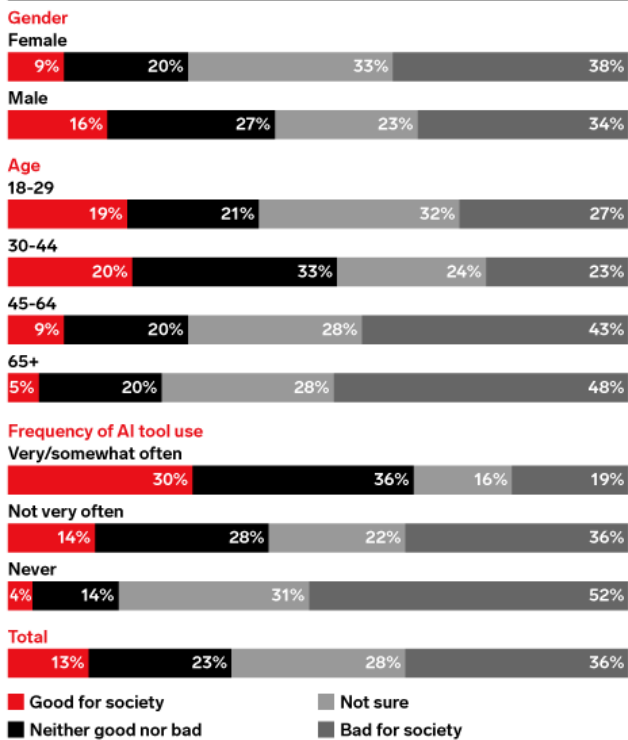
  A rushed job: The steep environmental costs aren't inevitable. Making data centers and neural networks run more efficiently could reduce the fallout. The problem is that tech companies are deploying technology supported by a weak foundation.

- To ease the computational workload of **Bard**, Google is initially using a scaled-back version of its **LaMDA** AI model, which might have contributed to an error that cost the tech giant **[$100 billion](#) in market value**.

- Constantly retraining models is expensive, which is likely the reason **OpenAI** has been operating a version of ChatGPT that uses data from 2021 and earlier.

  The high compute and energy costs of the technology make profitability uncertain and could contribute to a vicious cycle for tech companies. **Launching scaled back systems to cut costs means that the tech might not live up to the hype, undermining the consumer confidence** these companies need to make it viable.

Impact That Generative AI Text Will Have on Society According to US Adults, by Demographic, Jan 2023

% of respondents in each group

Gender

Female

| 9% | 20% | 33% | 38% |

Male

| 16% | 27% | 23% | 34% |

Age

18-29

| 19% | 21% | 32% | 27% |

30-44

| 20% | 33% | 24% | 23% |

45-64

| 9% | 20% | 28% | 43% |

65+

| 5% | 20% | 28% | 48% |

Frequency of AI tool use

Very/somewhat often

| 30% | 36% | 16% | 19% |

Not very often

| 14% | 28% | 22% | 36% |

Never

| 4% | 14% | 31% | 52% |

Total

| 13% | 23% | 28% | 36% |

■ Good for society ■ Not sure
■ Neither good nor bad ■ Bad for society

Note: numbers may not add up to 100% due to rounding
Source: YouGov as cited in company blog, Feb 1, 2023

280123   InsiderIntelligence.com

This article originally appeared in Insider Intelligence's Connectivity & Tech Briefing—a daily recap of top stories reshaping the technology industry. Subscribe to have more hard-hitting takeaways delivered to your inbox daily.

- Are you a client? *Click here to subscribe*.

- Want to learn more about how you can benefit from our expert analysis? *Click here*.