

Responsible AI AI Ethics AI Policy Data Governance

## 5 Takeaways from the UK's AI Safety Summit

2024 will see AI Safety Summits in France and Korea



**Ben Wodecki**  
November 2, 2023

4 Min Read



From left: European Commission President Ursula von der Leyen, U.S. Vice President Kamala Harris and Britain's Prime Minister Rishi Sunak at the AI Safety Summit GETTY IMAGES

The UK's AI Safety Summit has drawn to a close, with 100 attendees from across the world descending on a quaint English country estate to discuss considerations for building and governing future AI systems.

British Prime Minister Rishi Sunak said the U.K. is "proud to have brought the world together and hosted the first summit."

Here are five major takeaways from the event.

### 1. Consensus reached – but it's still early days

The EU and 28 countries signed the [Bletchley Declaration](#), an agreement to commit to designing AI that is safe and human-centric.

---

The missive consists of pledges to ensure AI is designed and deployed responsibly. Signatories commit to working together through existing international forums to promote cooperation. Of particular focus are the most advanced AI systems such as [Google Gemini](#) or the rumored [GPT-5](#) with the potential to inflict the greatest AI harms including human extinction.

The summit held eight [roundtable discussions](#) chaired by high-ranking government officials, researchers or academics. One group stated that “we are only at the early stages of understanding how these models work, the risks they pose, and therefore how to develop adequate safeguards.”

## 2. Sharing is caring

A key takeaway from the Bletchley agreement was that attendees agreed to cooperate by building a “shared scientific and evidence-based understanding” of AI risks, as a foundation for policy.

---

**Related:** [AI Safety Summit: 28 Nations and EU Sign the 'Bletchley Declaration'](#)

---

Part of those shared efforts would see increased transparency by private actors developing AI models, as well as the development of tools for safety testing that organizations can access.

Each nation, however, can categorize AI risks based on their national circumstances and legal frameworks – so they are not tied to one global viewpoint.

Suki Dhuphar, head of international business at Tamr said: “Through actively collaborating in AI safety research, organizations and governments can strike a balance between innovation and safeguarding against potential harms, ultimately unlocking AI's full potential while ensuring ethical and accountable AI deployment.”

## 3. Open source AI crowd disagrees

The overarching theme of the event was to mitigate the risk of advanced AI systems wiping out humanity if it falls into the hands of bad actors.

But a vocal group, specifically Meta's Chief AI Scientist Yann LeCun and Google Brain co-founder Andrew Ng, believes the focus on AI's existential threat is misguided and would lead to over-regulation that would harm open-source efforts and innovation.

They and about 150 of other open-source believers signed a statement published by Mozilla saying that open AI is “an antidote, not a poison.”

---

**Related:** [UK's AI Safety Summit: What They Discussed](#)

---

They have at least one high-profile backer: U.K. Deputy Prime Minister Oliver Dowden. “If we want to make sure [AI] spreads globally, in terms of the developing world, ... I think there is a very high bar to restrict open source in any way,” he told Politico.

Amanda Brock, CEO of OpenUK, lauded Dowden's stance. "Recognizing the value of open source to global economies, its role in democratizing technology, and building trust through transparency is critical to the evolution of AI. It is the only way to ensure that our digital future is equitable and that we learn the lessons from our recent digital history."

#### 4. China plays ball

China was a surprise addition to the guest list for some, and perhaps even more surprising is their attendance. A delegation from China's Ministry of Science and Technology as well as Alibaba and Tencent joined the summit.

China even signed onto the Bletchley Declaration, agreeing to its principles and pledging to attend future events.

Addressing the concern around China in his [closing speech](#), Prime Minister Sunak noted that "some said, we shouldn't even invite China. Others that we could never get an agreement with them. Both were wrong."

"A serious strategy for AI safety has to begin with engaging all the world's leading AI powers. And all of them have signed the Bletchley Park Communique."

In emailed comments, Peter van Jaarsveld, Global Head of Production at OLIVER, said: "Talk of U.S. / China is interesting and highlights that on the surface, there is a desire for global coordination, which will ultimately benefit people bringing AI into their day to day. But announcing an intention to collaborate does not guarantee collaboration or concrete outcomes. The summit represents the first tentative steps towards a global approach to AI, but whether it bears any fruit remains to be seen."

#### 5. More AI Safety Summits to come

The event at Bletchley Park marked the first global event on AI safety. But more are on the way – signatories to the Bletchley Declaration agreed to hold further such events in the future.

In 2024, there will be AI Safety Summits in South Korea and France.

In his [closing speech](#), Sunak said: "While this was only the beginning of the conversation, I believe the achievements of this summit will tip the balance in favor of humanity.

"Because they show we have both the political will and the capability to control this technology and secure its benefits for the long-term."

##### Read more about:

ChatGPT / Generative AI

#### About the Author(s)



**Ben Wodecki**

Jr. Editor

[See more from Ben Wodecki](#)

Keep up with the ever-evolving AI landscape

Unlock exclusive AI content by subscribing to our newsletter!!

**Stay Updated!**

### You May Also Like

- Responsible AI and Its Impact on Cybersecurity**  
August 01, 2023
- Responsible AI Generators Accept 85% of Fake News Prompts**  
July 31, 2023
- Responsible AI 'Masks' to Protect Images From Manipulation by AI**  
July 28, 2023
- Responsible AI Tech Leaders Reach Accord on AI Guardrails**  
July 21, 2023

### Latest podcasts

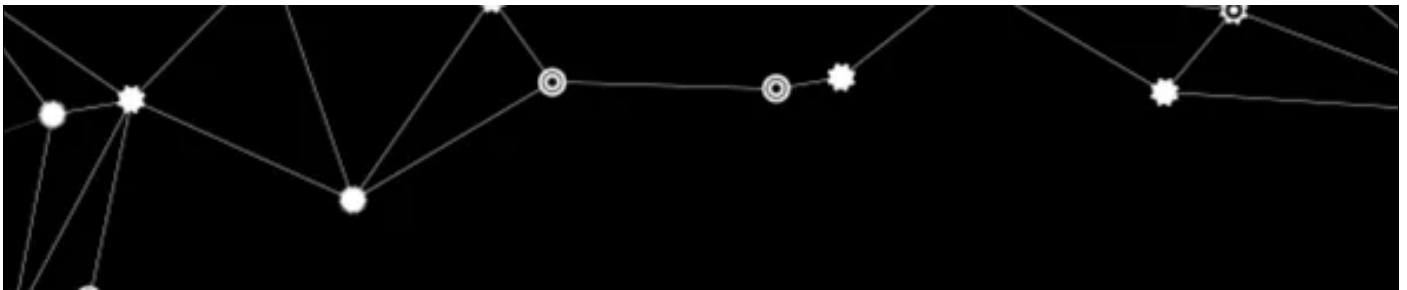
ML



Schneider Electric's CDO on the Merits of an AI Hub-and-Spoke Structure

Oct 11, 2023

NLP



## Latest News



NLP

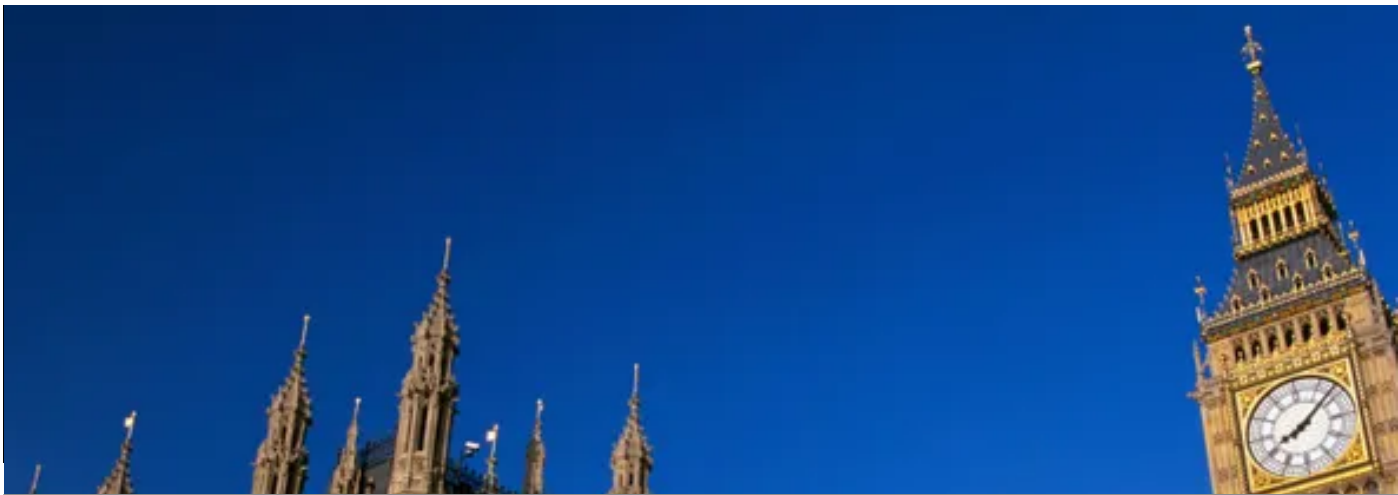
### Microsoft Makes Miniscule Model Multimodal Like ChatGPT's GPT-4V

by Ben Wodecki

Nov 3, 2023

3 Min Read

Verticals



Verticals



# Microsoft



## AI News Roundup: Microsoft May Have Leaked ChatGPT Parameters

by Ben Wodecki

Nov 3, 2023

4 Min Read

NLP

# TOP STORIES OF THE

Responsible AI



## AI Luminaries Make Case for Open Source AI

by Ben Wodecki

Nov 2, 2023

2 Min Read

### Trending articles

NLP





NLP



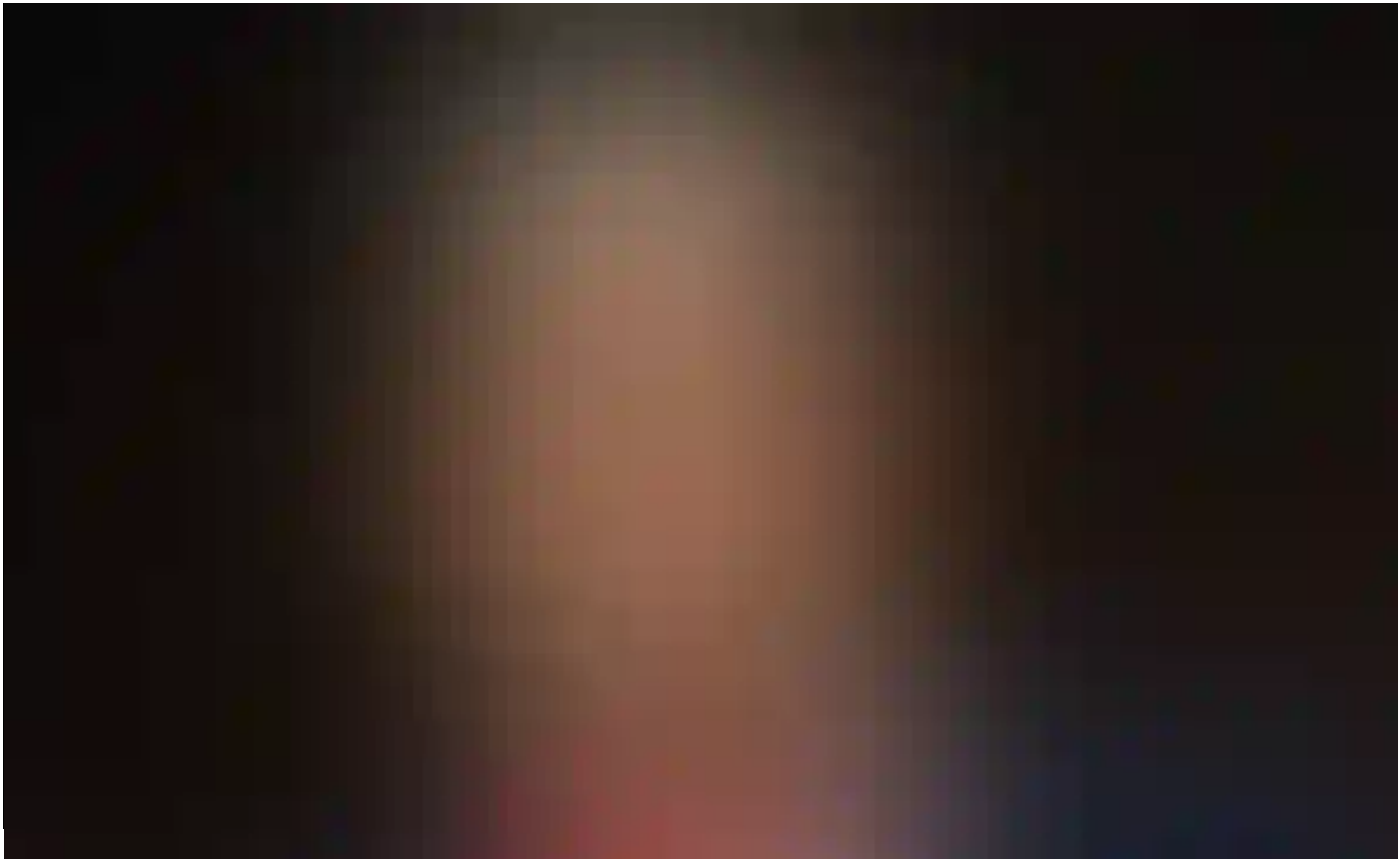
**How Apple Can Catch Up in the Generative AI Race**

by Sascha Brodsky

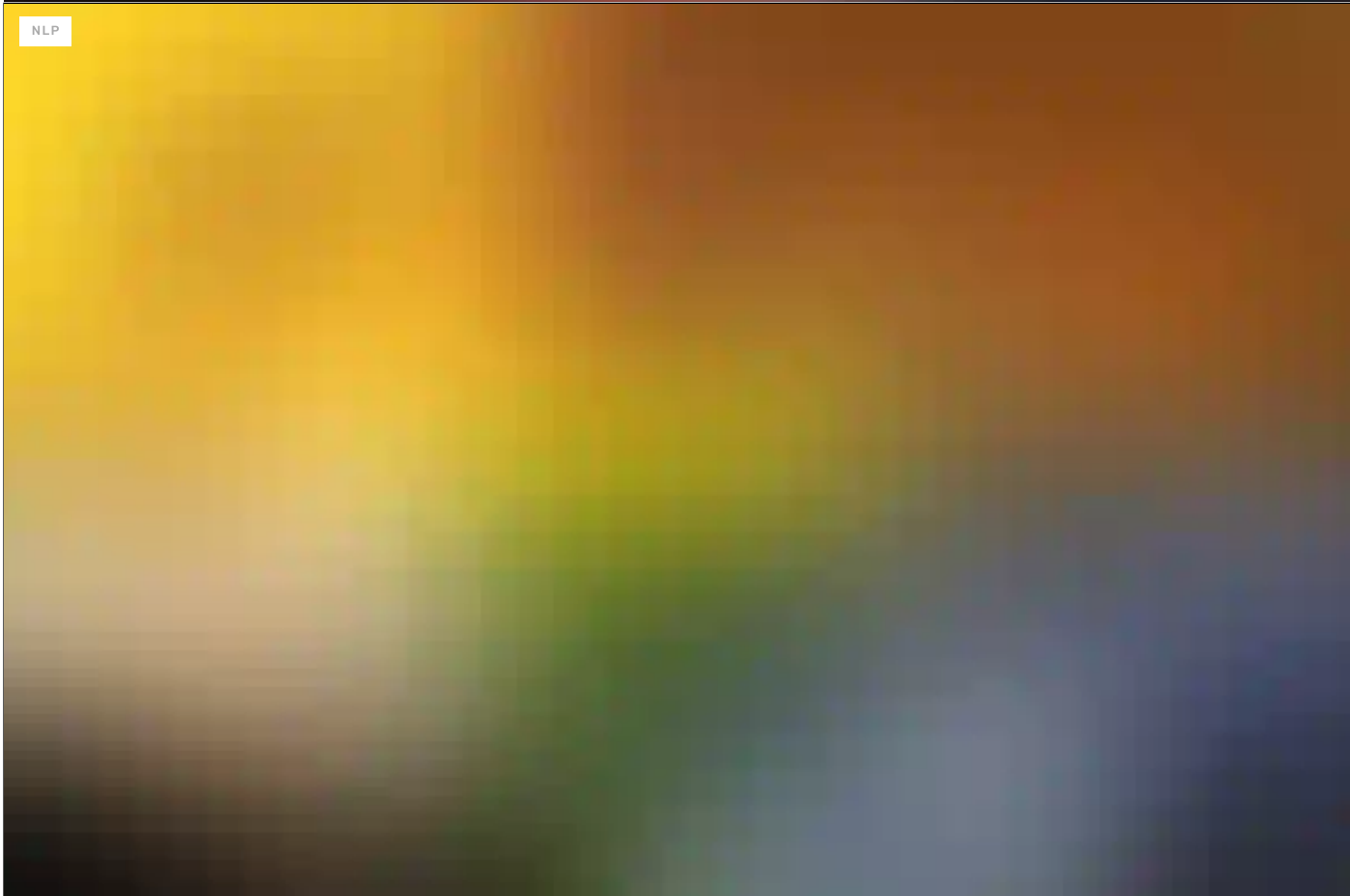
Oct 30, 2023

4 Min Read

NLP



NLP



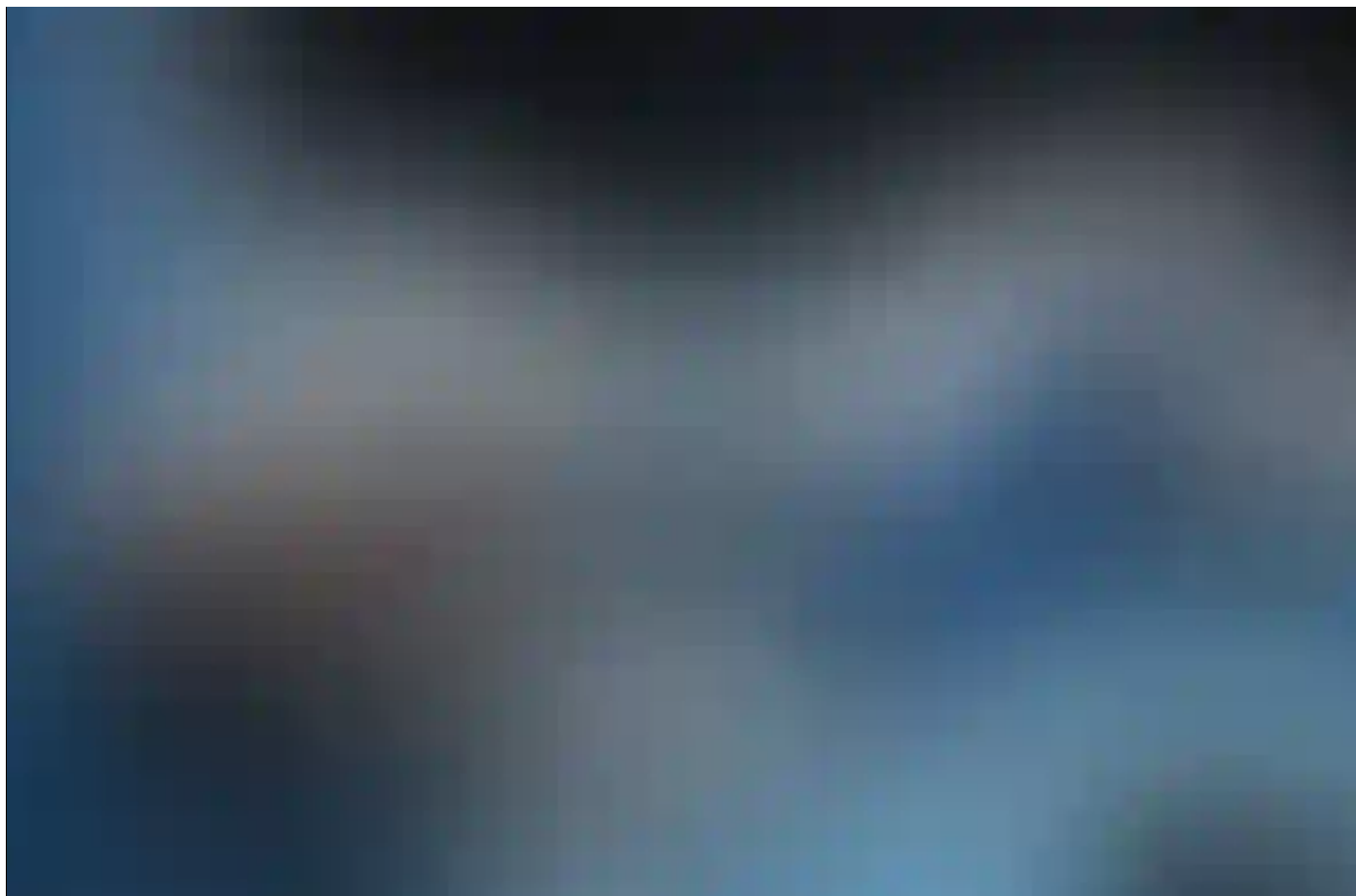
### Meta Researchers Use AI to Generate Images from the Human Brain

by Ben Wodecki

Oct 19, 2023

3 Min Read

ML



**Keep up with the ever-evolving AI landscape**

Unlock exclusive AI content by subscribing to our newsletter!!

**STAY UPDATED**

**AI  BUSINESS**

**More from Informa Tech**

The AI Summit Series

Applied Intelligence Live!

AI Research and Consulting

Our Solutions

**Communicate**

Contact us

About us

**Working with us**

Advertise

**Join Us**

**STAY UPDATED**

---

**Follow Us**



Copyright © 2023 Informa PLC Informa UK Limited is a company registered in England and Wales with company number 1072954 whose registered office is 5 Howick Place, London, SW1P 1WG.

[Home](#) | [About us](#) | [Contact](#) | [Cookie Policy](#) | [Terms of Use](#)