

## Anthropic, Google, Microsoft and OpenAI Launch Frontier Model Forum

- *Anthropic, Google, Microsoft and OpenAI are launching the Frontier Model Forum, an industry body focused on ensuring safe and responsible development of frontier AI models.*
- *The Forum aims to help (i) advance AI safety research to promote responsible development of frontier models and minimize potential risks, (ii) identify safety best practices for frontier models, (iii) share knowledge with policymakers, academics, civil society and others to advance responsible AI development; and (iv) support efforts to leverage AI to address society's biggest challenges.*
- *The Frontier Model Forum will establish an Advisory Board to help guide its strategy and priorities.*
- *The Forum welcomes participation from other organizations developing frontier AI models willing to collaborate toward the safe advancement of these models.*

**July 26, 2023** – Today, Anthropic, Google, Microsoft and OpenAI are announcing the formation of the Frontier Model Forum, a new industry body focused on ensuring safe and responsible development of frontier AI models. The Frontier Model Forum will draw on the technical and operational expertise of its member companies to benefit the entire AI ecosystem, such as through advancing technical evaluations and benchmarks, and developing a public library of solutions to support industry best practices and standards.

The core objectives for the Forum are:

1. **Advancing AI safety research** to promote responsible development of frontier models, minimize risks, and enable independent, standardized evaluations of capabilities and safety.
2. **Identifying best practices** for the responsible development and deployment of frontier models, helping the public understand the nature, capabilities, limitations, and impact of the technology.
3. **Collaborating with policymakers, academics, civil society and companies** to share knowledge about trust and safety risks.
4. **Supporting efforts to develop applications that can help meet society's greatest challenges**, such as climate change mitigation and adaptation, early cancer detection and prevention, and combating cyber threats.

### Membership criteria

The Forum defines frontier models as large-scale machine-learning models that exceed the capabilities currently present in the most advanced existing models, and can perform a wide variety of tasks.

Membership is open to organizations that:

- Develop and deploy frontier models (as defined by the Forum).
- Demonstrate strong commitment to frontier model safety, including through technical and institutional approaches.
- Are willing to contribute to advancing the Forum's efforts including by participating in joint initiatives and supporting the development and functioning of the initiative.

The Forum welcomes organizations that meet these criteria to join this effort and collaborate on ensuring the safe and responsible development of frontier AI models.

### What the Frontier Model Forum will do

Governments and industry agree that, while AI offers tremendous promise to benefit the world, appropriate guardrails are required to mitigate risks. Important contributions to these efforts have already been made by the US and UK governments, the European Union, the OECD, the G7 (via the Hiroshima AI process), and others.

To build on these efforts, further work is needed on safety standards and evaluations to ensure frontier AI models are developed and deployed responsibly. The Forum will be one vehicle for cross-organizational discussions and actions on AI safety and responsibility.

The Forum will focus on three key areas over the coming year to support the safe and responsible development of frontier AI models:

**Identifying best practices:** Promote knowledge sharing and best practices among industry, governments, civil society, and academia, with a focus on safety standards and safety practices to mitigate a wide range of potential risks.

**Advancing AI safety research:** Support the AI safety ecosystem by identifying the most important open research questions on AI safety. The Forum will coordinate research to progress these efforts in areas such as adversarial robustness, mechanistic interpretability, scalable oversight, independent research access, emergent behaviors and anomaly detection. There will be a strong focus initially on developing and sharing a public library of technical evaluations and benchmarks for frontier AI models.

**Facilitating information sharing among companies and governments:** Establish trusted, secure mechanisms for sharing information among companies, governments and relevant stakeholders regarding AI safety and risks. The Forum will follow best practices in responsible disclosure from areas such as cybersecurity.

**Kent Walker, President, Global Affairs, Google & Alphabet said:** “We’re excited to work together with other leading companies, sharing technical expertise to promote responsible AI innovation. We’re all going to need to work together to make sure AI benefits everyone.”

**Brad Smith, Vice Chair & President, Microsoft said:** “Companies creating AI technology have a responsibility to ensure that it is safe, secure, and remains under human control. This initiative is a vital step to bring the tech sector together in advancing AI responsibly and tackling the challenges so that it benefits all of humanity.”

**Anna Makanju, Vice President of Global Affairs, OpenAI said:** “Advanced AI technologies have the potential to profoundly benefit society, and the ability to achieve this potential requires oversight and governance. It is vital that AI companies—especially those working on the most powerful models—align on common ground and advance thoughtful and adaptable safety practices to ensure powerful AI tools have the broadest benefit possible. This is urgent work and this forum is well-positioned to act quickly to advance the state of AI safety.”

**Dario Amodei, CEO, Anthropic said:** “Anthropic believes that AI has the potential to fundamentally change how the world works. We are excited to collaborate with industry, civil society, government, and academia to promote safe and responsible development of the technology. The Frontier Model Forum will play a vital role in coordinating best practices and sharing research on frontier AI safety.”

### **How the Frontier Model Forum will work**

Over the coming months, the Frontier Model Forum will establish an Advisory Board to help guide its strategy and priorities, representing a diversity of backgrounds and perspectives.

The founding companies will also establish key institutional arrangements including a charter, governance and funding with a working group and executive board to lead these efforts. We plan to consult with civil society and governments in the coming weeks on the design of the Forum and on meaningful ways to collaborate.

The Frontier Model Forum welcomes the opportunity to help support and feed into existing government and multilateral initiatives such as the G7 Hiroshima process, the OECD's work on AI risks, standards, and social impact, and the US-EU Trade and Technology Council.

The Forum will also seek to build on the valuable work of existing industry, civil society and research efforts across each of its workstreams. Initiatives such as the [Partnership on AI](#) and [MLCommons](#) continue to make important contributions across the AI community, and the Forum will explore ways to collaborate with and support these and other valuable multi-stakeholder efforts.

### **Media contacts**

**Anthropic:** [press@anthropic.com](mailto:press@anthropic.com)

**Google:** [press@google.com](mailto:press@google.com)

**OpenAI:** [press@openai.com](mailto:press@openai.com)

**Microsoft:** [rapidresponse@we-worldwide.com](mailto:rapidresponse@we-worldwide.com)