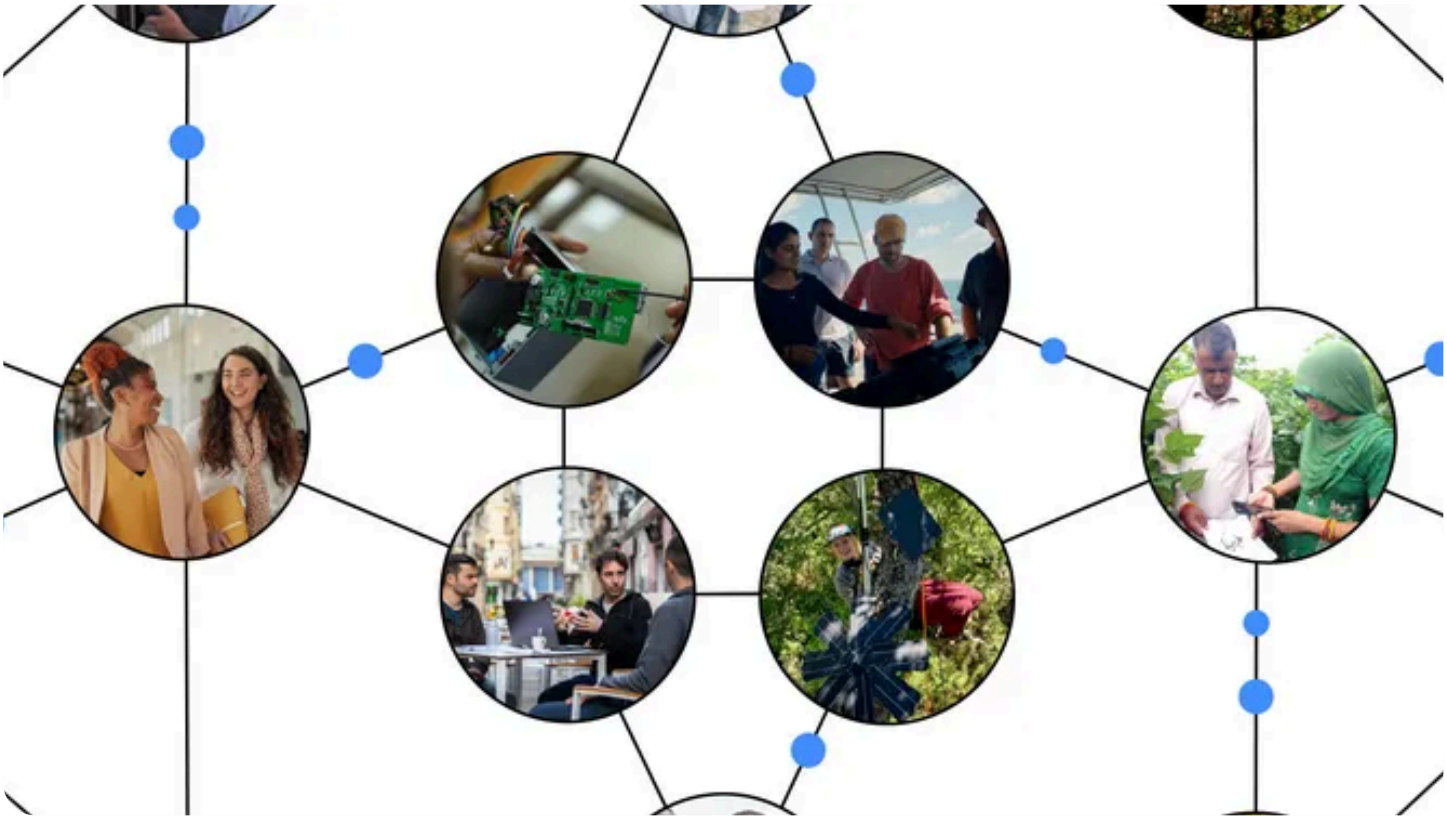
Being bold on AI means being responsible from the start

May 10, 2023 | 5 min read

From breakthroughs in products to science to tools to address misinformation, how Google is applying AI to benefit people and society

James Manyika
SVP, Technology & Society



We believe our approach to AI must be both bold and responsible. To us that means developing AI in a way that maximizes the positive benefits to society while addressing the challenges, guided by our AI Principles. While there is natural tension between the two, we believe it's possible — and in fact critical — to embrace that tension productively. The only way to be truly bold in the long-term is to be responsible from the start.

We're boldly applying AI in our groundbreaking products used by people everywhere, in our contributions to scientific advances that benefit people, and in helping to address societal challenges.

In our products

AI is already in many products millions (and in some cases billions) of people already use such as Google Maps, Google Translate, Google Lens and more. And now we are bringing AI to help people ignite and assist in their creativity with Bard, increase their productivity with Workspace tools, and revolutionize the way they access knowledge with Search Generative Experience. Several other early examples and experiments of useful applications can be found in Google Labs.

In helping to address societal challenges

We are applying AI to mitigating and adapting to climate change: by providing critical flood forecasts — now to more than 20 countries, tracking wildfire boundaries in real time, and helping to reduce carbon emissions by decreasing stop-and-go traffic. We're applying it to improve healthcare — including maternal care, cancer treatments and tuberculosis screening. And we recently announced a new large language model that could be a helpful tool for clinicians: Med-PaLM. Later this year, Data Commons — a system that organizes data from hundreds of sources to inform approaches to major societal challenges from sustainability to healthcare to jobs and the economy in many countries — will be accessible through Bard, making it even more useful.

In our field-defining research

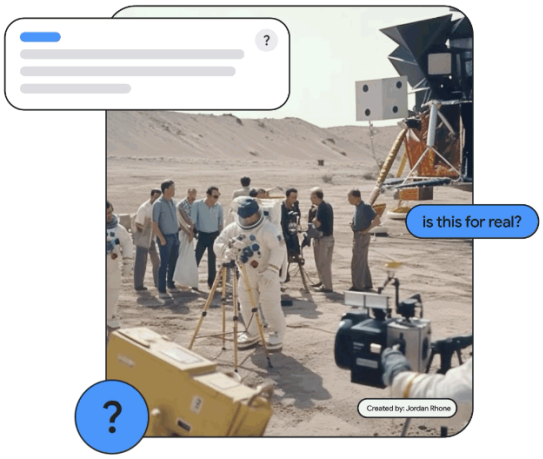
AI is helping scientists make bold advances across many fields from physics, material science and healthcare that will benefit society — take for example Google DeepMind's AlphaFold program. AlphaFold can accurately predict the 3D shape of 200 million proteins, nearly all cataloged proteins known to science — an accomplishment that gave us the equivalent of nearly 400 million years of research progress in just weeks. AI is also powering progress in making the world's information accessible to people everywhere. For example, it's enabling ambitious undertakings like our moonshot 1,000 Languages Initiative — we've made exciting progress towards our goal to support the 1,000 most spoken languages with a Universal Speech Model trained on over 400 languages.

Challenges

While it's exhilarating to see these bold breakthroughs, we know AI is a still-emerging technology — and there is still so much more to do. It is also important to acknowledge that AI has the potential to worsen existing societal challenges — such as unfair bias — and pose new challenges as it becomes more advanced and as new uses emerge, as our own research and that of others has highlighted. That's why we believe it's imperative to take a responsible approach to AI, guided by the AI Principles we first established in 2018. Each year we issue progress reports on how we are putting our AI Principles into practice that deep dive into examples. This work continues, as AI becomes more capable and we learn from users and new uses of the technologies, and we share what we learn.

Evaluating information

One area that is top of mind for many — including us — is misinformation. Generative AI makes it easier than ever to create new content, but it also raises additional questions about trustworthiness of information online. This is why we're continuing to develop and provide people with tools to evaluate online information. In the coming months, we're adding a new About this image tool in Google Search. About this image will provide helpful context such as when and where similar images may have first appeared and where else it's been seen online, including news, fact checking and social sites. Later this year, About this image will be available in Chrome and Google Lens.



AI Principles in action

As we apply our AI Principles to our products, we also start to see potential tensions when it comes to being bold and responsible. For example, Universal Translator is an experimental AI video dubbing service that helps experts translate a speaker's voice and match their lip movements. This holds enormous potential for increasing learning comprehension; yet knowing the risks it could pose in the hands of bad actors, we've built the service with guardrails to limit misuse and made it accessible to authorized partners only.

Another way we live up to our AI Principles is with innovations to tackle challenges as they emerge. For example, we're one of the first in the industry to automate adversarial testing using LLMs, which has significantly improved the speed, quality and coverage of testing, allowing safety experts to focus on more difficult cases. To help address misinformation, we'll soon be integrating new innovations in watermarking, metadata, and other techniques into our latest generative models. We're also making progress on tools to detect synthetic audio — in our AudioLM work, we trained a classifier that can detect synthetic audio in our own AudioLM model with nearly 99% accuracy.

A collective effort

We know building AI responsibly must be a collective effort involving researchers, social scientists, industry experts, governments, creators, publishers and people using AI in their daily lives.

We're sharing our innovations with others to increase impact, as in the case of Perspective API, which was originally developed by our researchers at Jigsaw to mitigate toxicity found in online comments. We're now applying it to our large language models (LLMs) — including every model mentioned at I/O — and academic researchers used it to create an industry-standard evaluation used by all significant LLMs, including models from OpenAI and Anthropic.

We believe that everyone benefits from a vibrant web ecosystem, today and in the future. To support that, we'll be working with the web community on ways to give web publishers choice and control over their web content.

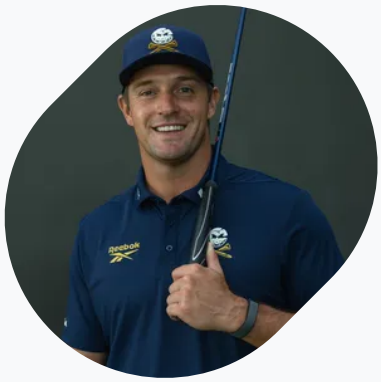
Part of what makes AI such an exciting area of focus is that the potential to benefit people and society everywhere is immense and palpable, as the imperative to develop and use it responsibly. So much is changing and evolving as AI advances, and more people experience, share, develop and use it. We constantly learn from our research, experiences, users and the wider community — and incorporate what we learn into our approach. Looking forward, there's so much we can accomplish and so much we must get right — together.

I would like to thank and acknowledge the inspiring and complex work of my colleagues on the Responsible AI, Responsible Innovation, Google.org, Labs, Jigsaw, Google Research and Google DeepMind teams.

POSTED IN:

AI Products

Related stories

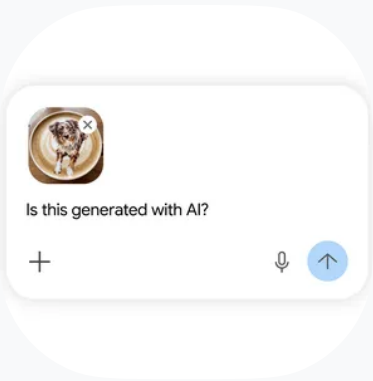


Google Health

Bryson DeChambeau partners with Google Health.

NotebookLM

Dive deeper into I/O 2026 with NotebookLM.



AI Products

Making it easier to understand how content was created and edited

By Laurie Richardson & Pushmeet Kohli