

AI

AI at Google: our principles

Sundar Pichai

CEO

Published Jun 7, 2018



At its heart, AI is computer programming that learns and adapts. It can't solve every problem, but its potential to improve our lives is profound. At Google, we use AI to make products more useful—from email that's spam-free and [easier to compose](#), to a digital assistant you can [speak to naturally](#), to photos that [pop the fun stuff out](#) for you to enjoy.

Beyond our products, we're using AI to help people tackle urgent problems. A pair of high school students are building AI-powered sensors to [predict the risk of wildfires](#). Farmers are using it to monitor the [health of their herds](#). Doctors are starting to use AI to help [diagnose cancer](#) and [prevent blindness](#). These clear benefits are why Google invests heavily in AI research and development, and makes AI technologies widely available to others via our tools and open-source code.

We recognize that such powerful technology raises equally powerful questions about its use. How AI is developed and used will have a significant impact on society for many years to come. As a leader in AI, we feel a deep responsibility to get this right. So today, we're announcing seven principles to guide our work going forward. These are not theoretical concepts; they are concrete standards that will actively govern our research and product development and will impact our business decisions.

The Keyword

Objectives for AI applications

We will assess AI applications in view of the following objectives. We believe that AI should:

1. Be socially beneficial.

The expanded reach of new technologies increasingly touches society as a whole. Advances in AI will have transformative impacts in a wide range of fields, including healthcare, security, energy, transportation, manufacturing, and entertainment. As we consider potential development and uses of AI technologies, we will take into account a broad range of social and economic factors, and will proceed where we believe that the overall likely benefits substantially exceed the foreseeable risks and downsides.

AI also enhances our ability to understand the meaning of content at scale. We will strive to make high-quality and accurate information readily available using AI, while continuing to respect cultural, social, and legal norms in the countries where we operate. And we will continue to thoughtfully evaluate when to make our technologies available on a non-commercial basis.

2. Avoid creating or reinforcing unfair bias.

AI algorithms and datasets can reflect, reinforce, or reduce unfair biases. We recognize that distinguishing fair from unfair biases is not always simple, and differs across cultures and societies. We will seek to avoid unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief.

3. Be built and tested for safety.

We will continue to develop and apply strong safety and security practices to avoid unintended results that create risks of harm. We will design our AI systems to be appropriately cautious, and seek to develop them in accordance with best practices in AI safety research. In appropriate cases, we will test AI technologies in constrained environments and monitor their operation after deployment.

4. Be accountable to people.

We will design AI systems that provide appropriate opportunities for feedback, relevant explanations, and appeal. Our AI technologies will be subject to appropriate human direction and control.

5. Incorporate privacy design principles.

The Keyword

of data.

6. Uphold high standards of scientific excellence.

Technological innovation is rooted in the scientific method and a commitment to open inquiry, intellectual rigor, integrity, and collaboration. AI tools have the potential to unlock new realms of scientific research and knowledge in critical domains like biology, chemistry, medicine, and environmental sciences. We aspire to high standards of scientific excellence as we work to progress AI development.

We will work with a range of stakeholders to promote thoughtful leadership in this area, drawing on scientifically rigorous and multidisciplinary approaches. And we will responsibly share AI knowledge by publishing educational materials, best practices, and research that enable more people to develop useful AI applications.

7. Be made available for uses that accord with these principles.

Many technologies have multiple uses. We will work to limit potentially harmful or abusive applications. As we develop and deploy AI technologies, we will evaluate likely uses in light of the following factors:

Primary purpose and use: the primary purpose and likely use of a technology and application, including how closely the solution is related to or adaptable to a harmful use

Nature and uniqueness: whether we are making available technology that is unique or more generally available

Scale: whether the use of this technology will have significant impact

Nature of Google's involvement: whether we are providing general-purpose tools, integrating tools for customers, or developing custom solutions

AI applications we will not pursue

In addition to the above objectives, we will not design or deploy AI in the following application areas:

1. Technologies that cause or are likely to cause overall harm. Where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks, and will incorporate appropriate safety constraints.
2. Weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people.

The Keyword

-
4. Technologies whose purpose contravenes widely accepted principles of international law and human rights.

We want to be clear that while we are not developing AI for use in weapons, we will continue our work with governments and the military in many other areas. These include cybersecurity, training, military recruitment, veterans' healthcare, and search and rescue. These collaborations are important and we'll actively look for more ways to augment the critical work of these organizations and keep service members and civilians safe.

AI for the long term

While this is how we're choosing to approach AI, we understand there is room for many voices in this conversation. As AI technologies progress, we'll work with a range of stakeholders to promote thoughtful leadership in this area, drawing on scientifically rigorous and multidisciplinary approaches. And we will continue to share what we've learned to improve AI technologies and practices.

We believe these principles are the right foundation for our company and the future development of AI. This approach is consistent with the values laid out in our original Founders' Letter back in 2004. There we made clear our intention to take a long-term perspective, even if it means making short-term tradeoffs. We said it then, and we believe it now.

POSTED IN:

[AI](#)

Related stories

AI

Researchers can use quantum simulation to explore quantum algorithms

The Keyword

Australian wildlife in
fire-affected areas

Dec 15, 2020



Dec 7, 2020



Questo sito utilizza i cookie di Google per erogare i propri servizi e analizzare il traffico. [Informazioni.](#)

OK