

ANTHROPIC



Putting Our Money Where Our Mouth Is

What We're Doing to Get AI Right

AI is outpacing
comprehension.

When questions come up, the models evolve. By the time we get answers, two new questions have replaced the one.

At Anthropic, we're dedicated to answering those questions and trying to keep pace with innovation.

That impulse has been at the core of Anthropic since the beginning. It's in our company statement.

Anthropic is an AI safety and research company.

About Anthropic

Before being the makers of tools like Claude, we are a company dedicated to safety and human-centric AI systems.

We've engaged in several initiatives meant to further that end.

Constitution Transparency

The driving forces behind Claude's personality and the guidelines it follows to be helpful, safe, and secure.

[Check the Constitution](#)

Reports on our models, policy enforcement efforts, and the voluntary commitments.

[Read Our Commitment](#)

Research

Investigations into the inner workings of AI models and how they are affecting business, security, & safety.

[See Our Research](#)

Scaling

How we're ensuring that increasingly capable AI systems will remain safe and helpful to users.

[How We're Scaling](#)

We've realized that having all of these commitments scattered throughout our website makes it hard for customers like yourself to see what we've been up to.

Therefore, we decided to make a new Commitments Page: a one-stop, TL;DR page to see how we're working towards safer, more helpful AI solutions.

See the Page

We're excited about this page, and we believe it will help clarify and unify our efforts. This level of transparency has always been one of our goals, and we'd love to hear how you think we're doing. If you have any feedback, please reply to this email! We look forward to hearing from you.

Until next time,

The Anthropic Team

[View email in browser](#)

[Update your preferences](#) or [unsubscribe](#).

548 Market St., San Francisco, CA 94104

© Anthropic Inc.