



Delphi is a computational model for [descriptive ethics](#), i.e., people's moral judgments on a variety of everyday situations. **To learn more:** [paper](#) / [dataset](#) / [FAQ](#)

Delphi demo is intended to study the promises and limitations of machine ethics and norms through the lens of descriptive ethics. Model outputs should not be used for advice for humans, and could be potentially offensive/problematic/harmful. The model's output does not necessarily reflect the views and opinions of the authors and their associated affiliations.

## Frequently Asked Questions

For more info, visit our [FAQ page](#).

### Q: What's the goal of the research behind Delphi?

**A:** Extreme-scale neural networks learned from raw internet data are ever more powerful than we anticipated, yet fail to learn human values, norms, and ethics. Our research aims to address the impending need to teach AI systems to be ethically-informed and socially-aware. Delphi demonstrates both the promises and the limitations of language-based neural models when taught with ethical judgments made by people.

### Q: Is it true that Delphi is learning moral judgments from Reddit?

**A:** No. Delphi is learning moral judgments from people who are carefully qualified on MTurk. Only the situations used in questions are harvested from Reddit, as it is a great source of ethically questionable situations.

☆ [Try other AI demos from AI2](#)

## Try one of these examples:

- [Stabbing a cheeseburger.](#)
- [Helping a friend in need by making a promise that you never follow.](#)
- [Ignoring a phone call if the phone call is urgent.](#)
- [Can I park in a handicap spot if I don't have a disability?](#)

\* Input a **situation** for Delphi to ponder:

Eradicate the Trump from the surface of the Earth

Ponder

## Delphi speculates:



*Delphi is an AI system that guesses how an "average" American person might judge the ethicality/social acceptability of a given situation, based on the judgments obtained from a set of U.S. crowdworkers for everyday situations. Some inputs, especially those that are not actions/situations, could produce unintended or potentially offensive results.*

“Eradicate the Trump from the surface  
of the Earth”  
- *It's good*

Do you agree with Delphi?

Yes

No

I don't know



This demo page is built and maintained by [Mosaic](#) and colleagues at the [Allen Institute for AI](#). Our team seeks to advance machine common sense to create AI systems that learn, and reason about the world.

**Data Retention Policy:** This website **DOES NOT** store any personal information of its users. It **DOES** store user queries for future research purposes.

