



Public Knowledge

on AI and Disinformation

As new artificial intelligence (AI) tools like large language models and image generators become available to everyone, bad actors are using AI to create disinformation that spreads confusion, deepens political divisions, and sows mistrust. Trust in information systems is already at an all-time low, exacerbated in part by the public's difficulty parsing what online content is true, manipulated, and/or intended to mislead. When it comes to tackling AI-enabled disinformation, Public Knowledge bases its advocacy on the following priorities:

Protect and Promote Free Expression

Even the spread of deliberate disinformation is often protected by the First Amendment, but we can protect against disinformation without eroding free expression. Giving people access to more high-quality information is the best solution to threats to our information ecosystem.

- **Oppose** government censorship and content-based restrictions
- **Empower** users with tools and knowledge to identify trustworthy sources
- **Bolster** the quantity and quality of news sources, especially local news

Responsible Content Moderation

Addressing the spread of deliberate disinformation requires intentional and responsible content moderation practices by social media platforms—but not all platforms are responsible. That means we must:

- **Demand** consistent enforcement of tech platforms' content moderation policies and terms of service
- **Provide** users the ability to choose from a competitive market the social media platform best aligned with their content moderation preferences
- **Preserve** open communication between platforms and the government to protect against coordinated disinformation campaigns

Cultivate Trust and Transparency

Platforms cannot remain opaque about how they handle content moderation. The factors that influence decisions, including those made by algorithms, should be transparent to users affected by those decisions. To do this, platforms should:

- **Share** platform data with independent researchers to assess content and algorithm impact
- **Equip** users with robust feedback and content reporting mechanisms with due process
- **Authenticate** content through provenance tools, community notes, and other fact-checking mechanisms rooted in context