Navigating AI's Future: Ethical Commitments of America's AI Giants

By Kevin LaGrandeur, Ph.D. ;
Director of Research, Global AI Ethics Institute
Aco Momcilovic, EMBA
President of Global AI Ethics Institute

Contributors to this report via the survey (in no particular order): Josh Gellers, Tomoko Mitsuoka, Mahendra Samarawickrama, Thiago Felipe Avanci, Aco Momcilovic, Ebru Goek, Azadeh Williams, Emma Ruttkamp, Elvis Thierry Sounna Vofo, Kaushik Chaudhuri, Alec Crawford, Alex Antic, Arlette Roman

Abstract: Recently, the 7 biggest companies making AI in the USA committed to the US government that they would voluntarily abide by a set of 8 rules in making future AI. This report is a summary of the opinions of the members of our Institute regarding the effectiveness of this agreement.

The question:

In July, 2023, the Biden administration of the United States government convinced the seven largest corporations working on AI to commit to voluntary rules that they would follow to make sure that the AI they develop would be safe. The authors of this report consequently made a survey of our Institute's membership to assess how effective they thought these rules and the corporations' commitment to them would be. The survey was sent out in August, 2023.

The 8 rules that the 7 large AI manufacturers said that they would voluntarily follow:

1. The companies commit to internal and external security testing of their A.I. systems before their release.

2. The companies commit to sharing information across the industry and with governments, civil society and academia on managing A.I. risks.

3. The companies commit to investing in cybersecurity and insider-threat safeguards to protect proprietary and unreleased model weights (coding that runs the AI).

4. The companies commit to facilitating third-party discovery and reporting of vulnerabilities in their A.I. systems.

5. The companies commit to developing robust technical mechanisms to ensure that users know when content is A.I. generated, such as a watermarking system.

6. The companies commit to publicly reporting their A.I. systems' capabilities, limitations, and areas of appropriate and inappropriate use.

7. The companies commit to prioritizing research on the societal risks that A.I. systems can pose, including on avoiding harmful bias and discrimination and protecting privacy.

8. The companies commit to develop and deploy advanced A.I. systems to help address society's greatest challenges.

Here are the four questions the survey asked of the respondents, who are all experts of our Institute and who live all over the world in various cultures:

- Which country are you from?
- Do you think these eight commitments are an effective start to good ethical standards for making AI? Why or why not?
- Do you think they would work in your country? Why or why not?
- Any other thoughts on this topic (Safety of AI and its future)?

Aggregate opinion regarding these rules:

Where participants were from: Those who contributed responses to the survey were from the following countries: USA, Japan, Australia, India, South Africa, Italy, Croatia, Germany, Switzerland, and Brazil.

Responses regarding the first question: Are these eight commitments are an effective start to good ethical standards for making AI?

The responses to the first question about the rules and their potential effectiveness were mixed. Pretty much everyone thought that these rules were a good place to start, but we all thought there were problems—some of them big.

The biggest doubt almost everyone had about these rules was the fact that they are voluntary. In other words, there are no teeth in these rules, no real enforceable consequences for breaking them. Because of this, most were dubious that the companies in question would actually hold to them in the face of profit motives and in the heat of the competition to develop AI further. And, as of now, there is no US regulatory body assigned to monitoring AI development.

Some also had problems with these rules not being detailed enough; others thought that the rules could be expanded and that, on a positive note, there was room to do that within this framework. One problem that we see is that most of these rules are things that the companies were already doing anyhow, or intended to do in the future. In other words, there was nothing here suggested by outside third parties or ethical boards, such as regulating how data is collected that is then used to train AI.

So, in sum, these rules are a good start, but the fact that they are voluntary poses problems, as does their somewhat vague, generalized construction.

Regarding the second question: Do you think these voluntary rules would work in your country?

Perhaps not surprisingly, the answers to this question depended on which country the contributors were from. Those of us from the USA, where these rules are supposed to be in effect, were very skeptical that this would work in our country because they are just voluntary rules with no enforcement capabilities behind them. And some of them have already been violated, especially those concerning data bias and privacy.

An issue that comes with cultural differences is tension between the need for such rules as these, and how dependent a country and its people might be on AI; such tension might inhibit or degrade the commitment to rules such as the ones here. For example, one contributor from Japan was doubtful that these rules would work for his country because use of data is much looser there than in the US, which would undermine the general effectiveness of the rules. Also, that Japanese contributor noted that worries about the country's national economic future are extreme, and this could cause a much more relaxed view of AI governance. Indeed, it is worth remembering that Japan has had a very marked decline in the population of their young, and so they are looking to robots and AI as a possible solution to who will take care of the aging population and who or what will provide the necessary labor pool in the future. It is notable that other economically developed countries, such as Italy, are having the same problem with population reduction, though not as extreme as Japan's. So they may also consider taking a more relaxed regulatory stance toward AI in order to solve the problem of replacing a declining working population.

Contributors in the European Union often mentioned that these voluntary rules are already included in the pending AI Act that is likely to become law there very soon. As such, they approved of their general adoption in European countries.

One other worry that some contributors had was that over-regulation might slow economic progress or hurt small businesses. This concern did not appear to be based on culture so much as economic and political leanings, because the objection was made by people from various countries across the cultural spectrum. But as businesses themselves were involved in making these rules, there does not seem to be much reason to worry that they will hurt business progress.

Regarding the third question: Any other thoughts on this topic?

One of the biggest concerns with these rules is that they did not take into account cultural differences. In particular, some worried that the rules were being defined by Western nations and their philosophies and religious beliefs. As one member, Arlette Roman, put it, "countries that value data sharing [and] innovation over privacy or those with low privacy concerns" might not want to worry so much about those issues in their regulation. One member from Japan, Tomoko Mitsuoka, said that he worried that there would be a hegemony of western values in

the making of regulations, and said that he hoped regulations would not be “based on only Christian…ethical framework[s] nor power politics, since ethics are very different culture to culture, religion to religion.”

As with question #2, a number of people reiterated the concern that economic differences among various countries would affect the implementation and efficacy of AI regulations. Kaushik Chaudhuri, a contributor from India noted, “it is uncertain that countries, specifically developing economies, have the capacity to handle the cascading employment impacts,” and thus it would be incumbent upon business leaders to provide more support for helping workers to “adapt to AI-induced changes in employment.” Ebru Goek added a detailed objection in this vein. As she pointed out,

> “The topic of AI Sustainability in a global context is crucial but often overlooked. Rules and guidelines aimed at ensuring safe AI applications often overlook the entire supply chain, including the natural raw materials sourced from the Global South, which are essential for the technology's development.
>
> Therefore, it is necessary to extend standard regulations and guidelines to guarantee the safety of procurers in the Global South. This is significant because numerous international tech companies from the Global North have their AI hubs in the Global South, where they unfairly benefit from the local communities, potentially causing harm.”

The overall consensus was that AI is very promising as an aid to society, as long as some guardrails are set up for AI makers, and that they are made mandatory rather than voluntary.

GAIEI 2024