JULY 21, 2023

FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI

Voluntary commitments – underscoring safety, security, and trust – mark a critical step toward developing responsible AI

Biden-Harris Administration will continue to take decisive action by developing an Executive Order and pursuing bipartisan legislation to keep Americans safe

Since taking office, President Biden, Vice President Harris, and the entire Biden-Harris Administration have moved with urgency to seize the tremendous promise and manage the risks posed by Artificial Intelligence (AI) and to protect Americans' rights and safety. As part of this commitment, President Biden is convening seven leading AI companies at the White House today – Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI – to announce that the Biden-Harris Administration has secured voluntary commitments from these companies to help move toward safe, secure, and transparent development of AI technology.

Companies that are developing these emerging technologies have a responsibility to ensure their products are safe. To make the most of AI's potential, the Biden-Harris Administration is encouraging this industry to uphold the highest standards to ensure that innovation doesn't come at the expense of Americans' rights and safety.

These commitments, which the companies have chosen to undertake immediately, underscore three principles that must be fundamental to the future of AI – safety, security, and trust – and mark a critical step toward developing responsible AI. As the pace of innovation continues to accelerate, the Biden-Harris Administration will continue to remind these companies of their responsibilities and take decisive action to keep Americans safe.

There is much more work underway. The Biden-Harris Administration is currently developing an executive order and will pursue bipartisan legislation to help America lead the way in responsible innovation.

Today, these seven leading AI companies are committing to:

Ensuring Products are Safe Before Introducing Them to the Public

- **The companies commit to internal and external security testing of their AI systems before their release.** This testing, which will be carried out in part by independent experts, guards against some of the most significant sources of AI risks, such as biosecurity and cybersecurity, as well as its broader societal effects.
- **The companies commit to sharing information across the industry and with governments, civil society, and academia on managing AI risks.** This includes best practices for safety, information on attempts to circumvent safeguards, and technical collaboration.

Building Systems that Put Security First

- **The companies commit to investing in cybersecurity and insider threat safeguards to protect proprietary and unreleased model weights.** These model weights are the most essential part of an AI system, and the companies agree that it is vital that the model weights be released only when intended and when security risks are considered.
- **The companies commit to facilitating third-party discovery and reporting of vulnerabilities in their AI systems.** Some issues may persist even after an AI system is released and a robust reporting mechanism enables them to be found and fixed quickly.

Earning the Public's Trust

- **The companies commit to developing robust technical mechanisms to ensure that users know when content is AI generated, such as a watermarking system.** This action enables creativity with AI to flourish but reduces the dangers of fraud and deception.
- **The companies commit to publicly reporting their AI systems' capabilities, limitations, and areas of appropriate and inappropriate use.** This report will cover both security risks and societal risks, such as the effects on fairness and bias.
- **The companies commit to prioritizing research on the societal risks that AI systems can pose, including on avoiding harmful bias and discrimination, and protecting privacy.** The track record of AI shows the insidiousness and prevalence of these dangers, and the companies commit to rolling out AI that mitigates them.
- **The companies commit to develop and deploy advanced AI systems to help address society's greatest challenges.** From cancer prevention to mitigating climate change to

so much in between, AI—if properly managed—can contribute enormously to the prosperity, equality, and security of all.

As we advance this agenda at home, the Administration will work with allies and partners to establish a strong international framework to govern the development and use of AI. It has already consulted on the voluntary commitments with Australia, Brazil, Canada, Chile, France, Germany, India, Israel, Italy, Japan, Kenya, Mexico, the Netherlands, New Zealand, Nigeria, the Philippines, Singapore, South Korea, the UAE, and the UK. The United States seeks to ensure that these commitments support and complement Japan's leadership of the G-7 Hiroshima Process—as a critical forum for developing shared principles for the governance of AI—as well as the United Kingdom's leadership in hosting a Summit on AI Safety, and India's leadership as Chair of the Global Partnership on AI. We also are discussing AI with the UN and Member States in various UN fora.

Today's announcement is part of a broader commitment by the Biden-Harris Administration to ensure AI is developed safely and responsibly, and to protect Americans from harm and discrimination.

- Earlier this month, Vice President Harris **convened consumer protection, labor, and civil rights leaders** to discuss risks related to AI and reaffirm the Biden-Harris Administration's commitment to protecting the American public from harm and discrimination.

- Last month, President Biden **met with top experts and researchers** in San Francisco as part of his commitment to seizing the opportunities and managing the risks posed by AI, building on the President's ongoing engagement with leading AI experts.

- In May, the President and Vice President **convened** the CEOs of four American companies at the forefront of AI innovation—Google, Anthropic, Microsoft, and OpenAI—to underscore their responsibility and emphasize the importance of driving responsible, trustworthy, and ethical innovation with safeguards that mitigate risks and potential harms to individuals and our society. At the companies' request, the White House hosted a subsequent meeting focused on cybersecurity threats and best practices.

- The Biden-Harris Administration published a landmark **Blueprint for an AI Bill of Rights** to safeguard Americans' rights and safety, and U.S. government agencies have ramped up their efforts to protect Americans from the risks posed by AI, including through **preventing algorithmic bias** in home valuation and **leveraging existing**

enforcement authorities to protect people from unlawful bias, discrimination, and other harmful outcomes.

- President Biden signed an **Executive Order** that directs federal agencies to root out bias in the design and use of new technologies, including AI, and to protect the public from algorithmic discrimination.

- Earlier this year, the **National Science Foundation** announced a $140 million investment to establish seven new National AI Research Institutes, bringing the total to 25 institutions across the country.

- The Biden-Harris Administration has also released a **National AI R&D Strategic Plan** to advance responsible AI.

- The Office of Management and Budget will soon release draft policy guidance for federal agencies to ensure the development, procurement, and use of AI systems is centered around safeguarding the American people's rights and safety.

###