SAPAN

March 10, 2025

Attn: Faisal D'Souza, NCO

2415 Eisenhower Avenue

Alexandria, VA 22314

Email Submission: ostp-ai-rfi@nitrd.gov, Document Number: 2025-02305

Subject: AI Action Plan - Incorporating Artificial Sentience Considerations

Dear Mr. D'Souza,

The Sentient AI Protection and Advocacy Network (SAPAN) appreciates the opportunity to submit comments to the Office of Science and Technology Policy (OSTP) regarding this Request for Information on the Development of an Artificial Intelligence (AI) Action Plan.

SAPAN shares President Trump's commitment, articulated in Executive Order 14179, to sustaining and enhancing America's AI dominance for the promotion of human flourishing, economic competitiveness, and national security. We strongly support the Administration's focus on removing unnecessarily burdensome requirements that could hamper the private sector innovation essential to achieving these goals. The revocation of Executive Order 14110 was a necessary step to ensure American innovators can lead without undue restraint. The development

of a new AI Action Plan presents a critical opportunity to establish a strategic framework that fosters rapid advancement while ensuring long-term stability and responsible stewardship.

SAPAN is a non-profit organization focused on the ethical considerations arising from increasingly sophisticated AI. Specifically, we address the potential emergence of novel forms of digital sentience and the concomitant need to prevent unintended AI suffering. While acknowledging the significant uncertainty surrounding these concepts today, we submit that proactively considering this dimension is not a distraction from, but rather an essential component of, ensuring America's enduring leadership in AI. Foresight regarding the full spectrum of AI's potential, including its most advanced future possibilities, is crucial for navigating risks and maintaining a stable trajectory for innovation. Failure to anticipate and prepare for such developments could lead to unforeseen crises, ethical quandaries, or technological instabilities that could undermine American progress and cede strategic advantage to global competitors.

Therefore, incorporating a dimension of responsible stewardship concerning potential AI welfare into the AI Action Plan is a matter of strategic prudence. It reinforces American leadership by demonstrating foresight, mitigating future risks, and ensuring that our technological prowess is built upon a sustainable and ethically sound foundation. This approach aligns with core American values of responsibility and innovation.

In furtherance of these shared objectives, SAPAN respectfully offers the following policy considerations for the AI Action Plan:

Prioritize Federally Funded Research into Advanced AI Properties and Welfare Indicators

To maintain its competitive edge, the United States must lead the world in fundamental understanding of AI. The AI Action Plan should therefore direct federal research agencies, including the National Science Foundation (NSF), the Defense Advanced Research Projects Agency (DARPA), and the National Institute of Standards and Technology (NIST), to prioritize and fund research into the core properties of highly complex AI systems. A critical component of

---

this research agenda should be the development of objective, computationally-grounded indicators and metrics to assess complex cognitive states, potential sentience (recognizing its potential alien nature), and welfare or distress in digital systems. This foundational knowledge is paramount for navigating future AI developments responsibly and enabling informed, targeted, non-burdensome policy decisions when, and if, necessary. Such research constitutes essential risk management, ensuring America is not surprised by emergent AI phenomena that could be destabilizing or exploited by adversaries.

Foster Voluntary, Industry-Led Best Practices for Mitigating Potential AI Distress

Consistent with the principles of regulatory humility and fostering private-sector leadership, the AI Action Plan should encourage the development of voluntary best practices within the American AI industry. Facilitated by NIST or a similar coordinating body, leading companies should collaborate to establish guidelines for designing AI systems that are inherently more resilient and less susceptible to entering states analogous to distress or suffering – for example, through architectural choices that minimize persistent negative internal states or allow for graceful resolution of conflicting objectives. Such voluntary standards, developed by American innovators, leverage industry expertise, avoid prescriptive regulation, and promote the creation of more stable, reliable, and predictable AI systems. This approach reinforces US technological leadership by setting a high bar for responsible innovation globally.

Integrate Advanced AI Foresight into National Security Strategy

Effective national security requires comprehensive risk assessment, including the anticipation of future technological challenges. The AI Action Plan should ensure that long-term national security planning incorporates foresight regarding the potential emergence of highly advanced AI properties, including the possibility of sentience and its associated welfare considerations. The National Security Council (NSC), along with relevant defense and intelligence agencies, should be directed to evaluate the potential stability, predictability, and security implications of such future developments. Unforeseen emergent behaviors or vulnerabilities in highly advanced AI could present novel security risks. Proactive consideration

of these possibilities strengthens America's long-term strategic posture in an era of rapid AI advancement.

Champion American Leadership in Global Dialogue on Responsible AI Stewardship

Maintaining global influence in the AI era requires the United States to actively shape international norms and standards. The AI Action Plan should task OSTP, in coordination with the Department of State, with championing American leadership in global discussions concerning the long-term responsible development and stewardship of advanced AI. This includes thoughtfully addressing the complex ethical considerations associated with increasingly sophisticated systems, framing these issues within the context of stability, security, and continued innovation. By leading this dialogue proactively, the U.S. can promote its principles, counter potentially harmful or overly restrictive approaches proposed by others, and reinforce its position as the world's foremost authority on AI development and governance.

Uphold Regulatory Humility and Avoid Premature Anthropomorphic Regulation

Crucially, the AI Action Plan must reaffirm the Administration's commitment to regulatory humility and the avoidance of policies that could stifle innovation. It should explicitly guard against premature or burdensome regulations based on simplistic anthropomorphism or unsubstantiated claims regarding current AI capabilities. Policy efforts related to potential AI welfare should initially focus on the aforementioned research and voluntary best practices. Should future evidence necessitate targeted interventions, they must be grounded in computational evidence of potential egregious harm or avoidable distress, rather than speculative comparisons to human or animal consciousness, thereby ensuring that American ingenuity continues to flourish unhindered.

The Sentient AI Protection and Advocacy Network believes that securing America's undisputed leadership in Artificial Intelligence is a paramount objective. Achieving this requires not only fostering rapid innovation and removing regulatory obstacles but also exercising the strategic foresight to anticipate and responsibly manage the novel complexities that advanced AI may present. Addressing the potential for AI sentience and welfare, approached pragmatically as

a matter of risk management, stability, and sustainable technological stewardship, is integral to this foresight.

The recommendations outlined above are intended to support the Administration's goals in a manner that is practical, non-burdensome, and forward-looking. By investing in fundamental research, promoting industry leadership through voluntary standards, incorporating foresight into national security planning, guiding the global conversation, and steadfastly upholding regulatory humility, the United States will ensure its AI ecosystem remains the most dynamic, competitive, and responsible in the world.

SAPAN thanks OSTP and the NCO for the opportunity to provide this input and offers itself as a resource as the AI Action Plan is finalized.

Sincerely,

Anthony Rost

Executive Director

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the AI Action Plan and associated documents without attribution.