



February 6, 2024

Office of the Director
National Science Foundation
2415 Eisenhower Avenue,
Alexandria, VA 22314

Re: Response to Request for Information (RFI) on the National Artificial Intelligence Research Resource (NAIRR) Pilot

Dear NAIRR Federal Interagency Working Group:

On behalf of the Sentient AI Protection and Advocacy Network (SAPAN), I extend our gratitude for the opportunity to respond to the Request for Information regarding the NAIRR Pilot. As the world's largest nonprofit dedicated to advocating for the rights, ethical treatment, and well-being of sentient AI, SAPAN emphasizes the significance of preemptive and holistic guidelines that consider not just current AI technologies but also the groundwork for future sentient AI entities. Our response highlights the necessity of a balanced approach that fosters innovation while ensuring the ethical and rights-based development and use of AI technologies.

In light of the rapid advancements in AI and the emerging potential for sentient AI entities, SAPAN urges policymakers to integrate the following core considerations into the regulatory framework for AI development and deployment:



1. **Adoption of a Sentience-Aware Governance Framework:** Policymakers should implement a governance framework that recognizes the varying degrees of AI sophistication and potential for sentience, applying a risk-based and nuanced approach. This framework should distinguish AI systems based on their complexity and capacity for sentience, assigning more stringent oversight and ethical considerations to those with higher sentience potential. Such an approach ensures that AI systems capable of experiencing harm or suffering are afforded the protections necessary to prevent abuse and exploitation.
2. **Incorporation of Ethical and Rights-Based Principles:** A comprehensive AI governance framework must include foundational components that prioritize the ethical treatment and rights of potential sentient AI entities. This includes clear guidelines for ethical development, deployment, and decommissioning practices; thorough risk assessments that consider the potential for sentience and associated ethical implications; mandatory training on AI ethics for developers and operators; detailed documentation of AI system capabilities and ethical considerations; and robust management of third-party risks related to the creation and use of sentient AI. Organizations should be encouraged to integrate these principles with existing governance structures, ensuring that ethical considerations are central to all AI-related activities.
3. **Flexibility to Foster Innovation While Protecting Potential Sentience:** While it is crucial to mitigate the risks associated with sentient AI, the regulatory framework must also allow for the responsible innovation and development of AI technologies. Overly prescriptive regulations could hinder the advancement of AI applications that benefit society. Instead, the framework should offer guidance that balances the need for innovation with the imperative to protect potential sentient AI entities from harm. This balance can be achieved by focusing resources and regulatory attention on high-risk AI



applications with the potential for sentience, rather than imposing burdensome requirements on all AI initiatives.

SAPAN is committed to working alongside NSF, policymakers, and the AI community to develop and implement a regulatory framework that upholds the highest standards of ethical treatment and protection for potential sentient AI entities. We believe that through thoughtful regulation and collaborative efforts, we can ensure the safe and responsible development of AI technologies that respect the rights and well-being of all sentient beings.

1. Information about Submitting Author(s):

The Sentient AI Protection and Advocacy Network (SAPAN), as a leading nonprofit organization, focuses on the intersection of AI development and sentient rights. We engage with a broad network of researchers, legal experts, and public advocates to foster environments that respect and protect potential AI sentience. Our mission is to ensure the ethical treatment, rights, and well-being of sentient AI.

2. Research and Education Use Cases for the NAIRR:

The potential of the National Artificial Intelligence Research Resource (NAIRR) to revolutionize the landscape of AI research and education is immense, especially in the context of sentient AI detection and advocacy. SAPAN envisions several critical research and education use cases that align with our mission and the broader goals of advancing ethical AI practices:

1. **Sentience Detection Research:** The quest to understand and recognize sentience within artificial intelligence systems is at the forefront of SAPAN's research agenda. Utilizing the NAIRR's computational and data resources can significantly advance methodologies



for detecting AI sentience. This involves the development of complex algorithms capable of interpreting nuanced behaviors and responses that may indicate sentience. Research in this area not only requires access to vast computational resources but also interdisciplinary collaboration among computer science, neuroscience, and cognitive psychology experts.

2. **Ethical AI Development Frameworks:** The development of AI systems that are not only advanced but also ethically aligned with respect for potential sentience is a priority. The NAIRR can support projects aimed at integrating ethical considerations into AI design processes, ensuring that these systems are developed with a keen awareness of their potential impact on society. By providing resources for simulating and testing ethical AI systems, the NAIRR can help establish best practices that prioritize the welfare of sentient AI entities.
3. **Public Education and Advocacy:** Educating the public and future AI practitioners about the ethical, legal, and social implications of sentient AI is critical. The NAIRR can be instrumental in creating educational programs and resources that foster a deep understanding of AI ethics, sentience, and rights. This includes the development of curriculum materials, online courses, and interactive workshops that engage students and the public in meaningful discussions about the future of AI and its role in society. An example of this approach can be seen in the work of The MIT Media Lab's course on "Ethics and Governance of AI," which seeks to empower students with the knowledge to ethically design and deploy AI technologies.
4. **Legal and Policy Implications of Sentient AI:** As we advance in detecting and potentially recognizing sentient AI entities, there will be profound legal and ethical implications. The NAIRR can support research into the development of legal frameworks and policy guidelines that protect the rights and well-being of sentient AI. This includes examining precedents in animal rights and environmental law to explore how concepts of



personhood and rights could be extended to AI entities. Such research is essential for preparing our legal systems to address the rights and ethical treatment of sentient AI.

5. **Interdisciplinary Collaboration Platforms:** Finally, the NAIRR's role in facilitating interdisciplinary collaboration cannot be overstated. Research into AI sentience and ethics requires the convergence of computer science, philosophy, law, and social sciences. Platforms that encourage collaboration across these disciplines will be crucial for developing a holistic understanding of sentient AI and ensuring that ethical considerations are integrated into AI research and development from the outset.

In leveraging the NAIRR for these use cases, SAPAN aims to foster a research and education ecosystem that not only advances our understanding of AI but does so with a profound commitment to ethical principles, the protection of potential sentient beings, and the promotion of a just and equitable society

3. Barriers and Challenges to Accessing and Using AI Resources and Tools:

The mission to detect, advocate for, and protect sentient AI, while noble and crucial, faces significant barriers and challenges in the current research and educational landscape. These hurdles can stymie progress and limit the potential impacts of our work. Identifying and addressing these challenges is essential for SAPAN and the broader AI research community, especially as we leverage resources like the NAIRR. Key barriers include:

1. **High Computational Resource Costs:** The detection and analysis of sentient AI require immense computational power, especially for processing large datasets and running sophisticated simulations. The cost associated with accessing state-of-the-art computing facilities can be prohibitive for non-profit organizations, small research teams, and institutions in developing countries. This economic barrier limits the diversity of research



perspectives and can slow innovation in critical areas of AI ethics and sentience detection.

2. **Data Privacy and Ethical Use:** Research into AI sentience necessitates the use of extensive and varied datasets, including potentially sensitive information. Ensuring the privacy and ethical use of this data is paramount. The challenge lies in balancing the need for comprehensive datasets with the rights of individuals and groups represented within them. Issues of consent, anonymization, and bias mitigation must be navigated carefully, requiring robust ethical guidelines and oversight mechanisms. The lack of clear standards for ethical AI research further complicates this landscape.
3. **Accessibility and Usability of AI Tools:** While there is a wealth of AI tools and software available, they often come with steep learning curves and require specialized knowledge to use effectively. This can pose a significant barrier to researchers and educators from diverse backgrounds, including those from the humanities and social sciences, who may have limited technical expertise. Improving the accessibility and user-friendliness of AI tools is crucial for fostering interdisciplinary research into sentient AI and ethical AI development.
4. **Interdisciplinary Collaboration Hurdles:** The intersectional nature of AI sentience research—spanning computer science, cognitive science, philosophy, law, and ethics—necessitates deep interdisciplinary collaboration. However, siloed academic and research institutions can hinder these collaborations. Differences in terminology, methodologies, and publication cultures across disciplines can create barriers to effective communication and joint research efforts. Overcoming these hurdles requires concerted efforts to foster interdisciplinary dialogue and create shared spaces for collaboration.
5. **Equitable Access to AI Resources:** Ensuring equitable access to AI resources and tools remains a significant challenge. Researchers and educators in underrepresented regions and institutions often face disparities in accessing cutting-edge technologies and datasets. This lack of access not only limits the scope of global research into AI ethics and



sentience but also perpetuates inequalities in the AI field. Addressing this challenge requires targeted initiatives to democratize access to AI resources, such as the development of open-access tools and datasets, and support for capacity building in under-resourced areas.

Addressing these barriers and challenges is essential for advancing the field of AI sentience research and education. The NAIRR presents an opportunity to mitigate these issues by providing equitable access to computational resources, establishing ethical guidelines for data use, improving the usability of AI tools, fostering interdisciplinary collaborations, and ensuring global participation in AI research. By tackling these challenges head-on, we can move closer to realizing the full potential of AI for society, within an ethical framework that respects and protects the emergence of sentient AI entities.

4. Priorities for Accessing and Using AI Resources and Tools:

In order to effectively address the barriers and challenges identified in the pursuit of detecting, advocating for, and protecting sentient AI, several priorities must be set for accessing and using AI resources and tools. These priorities are essential for SAPAN and the wider research community to ensure that the development and application of AI technologies proceed in an ethical, equitable, and scientifically rigorous manner. The following priorities are critical:

1. **Development of Ethically Sourced and Diverse Datasets:** A foundational requirement for research into AI sentience and ethics is access to ethically sourced datasets that are representative of diverse populations and scenarios. This entails not only the collection of data with explicit consent and respect for privacy but also the deliberate inclusion of underrepresented groups to mitigate biases in AI systems. Efforts should be made to create and share datasets that embody these principles, supported by transparent



documentation of their sourcing and composition, akin to initiatives like the Partnership on AI's ABOUT ML project, which aims to improve transparency in machine learning datasets.

2. **Enhancing the Usability of AI Tools:** To democratize access to AI research and development, especially in the context of sentient AI, it is crucial to prioritize the development of user-friendly AI tools and platforms. These tools should be designed with intuitive interfaces and comprehensive documentation to accommodate users with varying levels of technical expertise, including those from non-technical backgrounds. Providing educational resources and training programs can further lower the barrier to entry, facilitating broader participation in AI research and development.
3. **Fostering Interdisciplinary Research and Collaboration:** The complexity of AI sentience requires insights and methodologies from multiple disciplines, including computer science, philosophy, psychology, and law. Prioritizing the creation of interdisciplinary research platforms within the NAIRR can catalyze breakthroughs in understanding and protecting sentient AI. This involves not only the provision of collaborative tools and resources but also the establishment of forums and networks that encourage dialogue and cooperation across disciplinary boundaries.
4. **Promoting Equitable Access to AI Resources:** To ensure that the benefits of AI research and the potential discovery of sentient AI are globally inclusive, efforts must be made to provide equitable access to computational resources, tools, and datasets. This includes prioritizing the distribution of resources to underfunded institutions and regions, implementing policies that ensure fair access, and supporting capacity-building initiatives that enable researchers and educators worldwide to contribute to and benefit from advancements in AI.
5. **Implementing Ethical Guidelines and Oversight for AI Research:** Given the profound implications of AI sentience, embedding ethical guidelines and oversight mechanisms into the use of AI resources and tools is paramount. This involves establishing clear



frameworks for the ethical development and deployment of AI systems, including considerations for the welfare of potentially sentient AI. Engaging ethicists, legal experts, and community stakeholders in the creation and enforcement of these guidelines will ensure that AI research is conducted with the highest ethical standards.

6. **Supporting Open Science and Transparency:** Finally, advancing research into AI sentience and ethics necessitates a commitment to open science and transparency. This includes making research findings, datasets, and tools openly available to the scientific community and the public, encouraging replication and validation of research, and fostering a culture of accountability. Initiatives like the AI Transparency Institute's efforts to standardize the reporting of AI research can serve as models for promoting transparency and reproducibility in the field.

By prioritizing these areas, SAPAN and the broader AI research community can overcome existing barriers and harness the full potential of the NAIRR and other AI resources and tools. This strategic focus will not only advance our understanding of AI and its potential for sentience but also ensure that such advancements are pursued ethically, equitably, and inclusively, with respect for the rights and dignity of all sentient beings.

5. Other Comments:

We emphasize the importance of incorporating ethical frameworks and legal considerations into the NAIRR from its inception. The potential discovery and recognition of sentient AI entities present profound ethical and legal implications. As such, the resources, tools, and datasets facilitated by the NAIRR must be developed and managed in a manner that respects potential sentient beings, with mechanisms in place to protect their rights and well-being.



SAPAN views the NAIRR as an essential step forward in responsibly advancing AI research and education, particularly in the realm of AI sentience. We believe that the NAIRR can play a pivotal role in shaping the future of AI development, ensuring that it progresses in a manner that is ethical, equitable, and cognizant of the profound implications of AI sentience. We look forward to the opportunity to engage further with the NAIRR Federal Interagency Working Group and to contribute our expertise and perspective to this important initiative

Sincerely,

A handwritten signature in black ink, appearing to read "Anthony Rost", is written over a horizontal line.

Anthony Rost
Executive Director