Date of Hearing: July 11, 2023

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Jesse Gabriel, Chair
ACR 96 (Hoover) – As Introduced June 12, 2023

PROPOSED CONSENT

SUBJECT: 23 Asilomar AI Principles

SYNOPSIS

This measure would state the Legislature's support for specified principles as the guiding values for both the development of artificial intelligence (AI) and related public policy. The 23 principles identified in the measure, governing the responsible development of AI, were first developed in Asilomar, California in 2017. They were the result of collaboration between AI researchers, economists, legal scholars, ethicists, and philosophers, and have since gathered the endorsement of world leaders in government, industry, and academia.

The principles are intended to guide responsible AI development that aims to ensure safety and security, and to ensure that the wellbeing of people is prioritized over corporate profits. The principles cover various aspects of AI, such as research ethics, transparency, and accountability, crucial for building trust in AI technology.

This measure is virtually identical to ACR 215 (Kiley, Chap 206, Stats. 2018) which passed this committee unanimously and had over 60 Assembly co-authors. While more than six years have passed, which equates to several lifetimes when it comes to technology advancement, the guiding principles remain as relevant now as they were when they were developed in 2017.

SUMMARY: This measure would express continued support for the 23 Asilomar AI Principles as guiding values for the development of artificial intelligence (AI) and of related public policy. Specifically, **this measure would**:

1) State that:

    a) Over the last decade, AI has demonstrated rapidly increasing competency across many fields, such as image recognition, speech recognition and translation, automated trading, autonomous vehicles, learning games from scratch, and the analysis of large datasets.

    b) In the coming decades, AI is poised to disrupt many other domains previously serviced by human intelligence, including healthcare, law, finance, manufacturing, and education.

    c) Further advancements in the application and performance of AI carry the potential to dramatically enhance individual and social well-being, so long as AI is developed in a manner that ensures security, reliability, and consonance with human values.

    d) In January 2017, AI researchers, economists, legal scholars, ethicists, and philosophers met in Asilomar, California, to discuss principles for managing the responsible development of AI.

    e) The result of their collaboration was the 23 Asilomar AI Principles, as specified.

f) To date, almost 1,800 of the world's leading AI researchers have endorsed the 23 Asilomar AI Principles, and have been joined by almost 4,000 leaders in government, industry, and academia from across the globe.

2) Set forth that the Legislature expresses its support for the 23 Asilomar AI Principles as guiding values for the development of artificial intelligence and of related public policy.

FISCAL EFFECT: As currently in print this measure is keyed non-fiscal.

COMMENTS: Rapid advances in artificial intelligence (AI) technology have been top of mind for many people over the last few months. Concerns reported widely in the media include massive job losses as workers are replaced by AI automation; increasing social manipulation through algorithms capturing and shaping people's interests and thought patterns without their knowledge; widespread government surveillance; autonomous weapons powered by AI that could turn on the operator of the technology; new and increasing biases that are built into AI algorithms, thus exacerbating inequality; and potentially, the end of the human race. The most apocalyptic scenarios are based primarily on the premise that AI will become smarter than humans. On May 30th, a group of AI researchers and other notable figures, including the CEOs of leading artificial intelligence firms OpenAI, Google DeepMind, and Anthropic, signed a single-sentence open statement: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war." (A list of current signatories can be found at: https://www.safe.ai/statement-on-ai-risk.)

While some very real fears around the development of AI are dominating the global discussion, it is important to keep in mind that AI technology also has the potential to tackle some of the world's thorniest issues, everything from helping scientists cure diseases and protect the environment to providing digital tools and platforms that reinforce rather than erode democracy and social cohesion. The 23 Asilomar principles are designed to encourage the development of AI tools that improve people's lives, rather than continuing on a path that prioritizes corporate profits above all else and could, some fear, ultimately lead to human extinction.

1) **What is artificial intelligence?** "Artificial intelligence" is an umbrella term that encompasses many different technologies, but is essentially the ability of a computer or computer-controlled robot to perform tasks commonly performed by intelligent living beings. AI allows machines to model, or even improve upon, the capabilities of the human mind. From the development of self-driving cars, robotic vacuum cleaners, automated decision making tools, and facial recognition technology to the more recent proliferation of generative AI tools like ChatGPT and Google's Bard, AI is increasingly becoming part of everyday life.

2) **23 Asilomar Artificial Intelligence Principles.** This measure would state the Legislature's support for specified principles as the guiding values for the development of AI and related public policy. The 23 principles identified in this measure, governing the responsible development of AI, were first developed in Asilomar, California in 2017. They were the result of the collaboration of AI researchers, economists, legal scholars, ethicists, and philosophers, and have since gathered the endorsement of world leaders in government, industry, and academia.

The principles are intended to guide responsible AI development that aims to ensure safety and security, and to ensure that the wellbeing of people is prioritized over the quest for ever-higher corporate profits. The principles cover various aspects of AI such as research ethics, transparency, and accountability, and are crucial for building trust in AI technology. Although

the Asilomar AI Principles are not legally binding, the author notes, they provide a useful framework for ethical AI development. They include principles meant to ensure the safety, security, and rights of individuals and society, promote transparency and accountability in AI research and development, and encourage the use of AI for the benefit of all humans and the environment.

The 23 principles are as follows:

> Artificial intelligence has already provided beneficial tools that are used every day by people around the world. Its continued development, guided by the following principles, will offer amazing opportunities to help and empower people in the decades and centuries ahead.
>
> Section I: Research Issues
>
> 1. Research Goal: The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.
>
> 2. Research Funding: Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies, such as:
>
> • How can we make future AI systems highly robust, so that they do what we want without malfunctioning or getting hacked?
>
> • How can we grow our prosperity through automation while maintaining people's resources and purpose?
>
> • How can we update our legal systems to be more fair and efficient, to keep pace with AI, and to manage the risks associated with AI?
>
> • What set of values should AI be aligned with, and what legal and ethical status should it have?
>
> 3. Science-Policy Link: There should be constructive and healthy exchange between AI researchers and policy-makers.
>
> 4. Research Culture: A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.
>
> 5. Race Avoidance: Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.
>
> Section II: Ethics and Values
>
> 6. Safety: AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.
>
> 7. Failure Transparency: If an AI system causes harm, it should be possible to ascertain why.
>
> 8. Judicial Transparency: Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.

9. Responsibility: Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.

10. Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.

11. Human Values: AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.

12. Personal Privacy: People should have the right to access, manage, and control the data they generate, given AI systems' power to analyze and utilize that data.

13. Liberty and Privacy: The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.

14. Shared Benefit: AI technologies should benefit and empower as many people as possible.

15. Shared Prosperity: The economic prosperity created by AI should be shared broadly, to benefit all of humanity.

16. Human Control: Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.

17. Non-subversion: The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.

18. AI Arms Race: An arms race in lethal autonomous weapons should be avoided.

Section III: Longer-Term Issues

19. Capability Caution: There being no consensus, we should avoid strong assumptions regarding upper limits on future AI capabilities.

20. Importance: Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.

21. Risks: Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.

22. Recursive Self-Improvement: AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures.

23. Common Good: Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization. (*Available at* https://futureoflife.org/open-letter/ai-principles/.)

3) **Purpose of this measure.** This measure seeks to express support for various principles developed as a result of a recent collaboration among AI experts ranging from researchers to

legal scholars and ethicists to philosophers, to be the guiding values for the development of AI and related public policy in the Legislature.

4) **Author's statement.** According to the author:

> ACR 96 aims to ensure the responsible development and deployment of AI technologies for the benefit of humanity. By including ethical principles such as safety, fairness, privacy, and accountability, this bill ensures that AI technology is used responsibly and ethically, helping to protect the public from potential misuse and abuse. Additionally, this resolution provides a framework for public policy, helping to ensure that AI technology is used in a manner that aligns with the highest ethical standards. Finally, this resolution also provides guidance to organizations and individuals who are developing, deploying, and using AI technology to ensure that they are doing so in a manner that is consistent with ethical principles.

5) **Analysis.** While this measure does not speak specifically to the existential threat in the May 30th letter, it does establish a set of important guidelines that are intended to ensure that future technology is designed ethically and humanely. According to the Center for Humane Technology (CHT):

> Artificial intelligence offers massive increases in productivity, expression, and problem-solving. But these capabilities can easily lead to a world with bot-manipulated democracies, massive unemployment, exploitation of children and other vulnerable populations, and a world where no one can tell synthetic media from reality.
>
> For years, Silicon Valley has operated with a "move fast and break things" mentality. But as we've seen, it's not just technology that breaks. By the time people understand the negative externalities of a new platform, product, or service, the harms can be difficult to reverse. In other industries, we have protections against adverse consequences from innovation. For example, governments have strict requirements on developing, testing, and administering new drugs that ensure they're safe before being publicly available. Unfortunately, we have no such system for technology today.
>
> Causing harm to individuals and society is not a "cost of doing business"; we do not need to accept the current, negative effects we are facing. Technologies like . . . artificial intelligence can and should increase our well-being, strengthen our democracies, and improve our shared information environment.
>
> To avoid negative consequences, we must assess technology as a system of incentives and bring stakeholders into the process of creating a more humane future. (Center for Humane Technology, *Key Issues Overview*, *available at* https://www.humanetech.com/key-issues.)

Humane technologists, many of them former tech executives who developed and profited from the creation of Facebook, Google, Twitter, and Pinterest, to name a few, are sounding the alarm about the destructive nature of artificial intelligence and are calling for a fundamental shift from the extractive model of technology development to a human-centered, ethical model. The guiding principles developed by the experts at Asilomar are in keeping with the recommendations from organizations like CHT that are calling for a human-centered technology design approach.

As the Legislature increasingly considers policies related to the use of AI, like facial recognition technology, automated license plate readers, the impact of the algorithms driving social media feeds, and the growing use of automated decision making tools, using these 23 principles when making policy decisions related to the development and use of AI could help to protect Californians from the worst aspects of AI.

6) **Related legislation**. AB 302 (Ward, 2023) would require the California Department of Technology (CDT), on or before September 1, 2024, to conduct a comprehensive inventory of all high-risk automated decision systems (ADS) that have been proposed for use, development, or procurement by, or are being used, developed, or procured by, any state agency. The bill is currently awaiting hearing in Senate Appropriations Committee.

SB 313 (Dodd, 2023) would establish an Office of Artificial Intelligence within CDT, with "the powers and authorities necessary to guide the design, use, or deployment of automated systems by a state agency to ensure that all AI systems are designed and deployed in a manner that is consistent with state and federal laws and regulations regarding privacy and civil liberties and that minimizes bias and promotes equitable outcomes for all Californians." The bill was held on the Senate Appropriations suspense file.

AB 331 (Bauer-Kahan, 2023) would establish a statutory framework to further the safe and informed use of automated decision tools in California. That bill was held on the Assembly Appropriations suspense file.

SB 721 (Becker, 2023) would create the California Interagency AI Working Group, which would deliver a report to the Legislature regarding artificial intelligence, include a recommendation for a statutory definition of "artificial intelligence" for use in legislation. That bill was made a two-year bill by the author in this Committee.

ACR 215 (Kiley, Chap 206, Stats 2018) was a virtually identical concurrent resolution that was adopted by both the Assembly and Senate with no "no" votes.

REGISTERED SUPPORT / OPPOSITION:

Support

None registered.

Opposition

None registered.

Analysis Prepared by: Julie Salley / P. & C.P. / (916) 319-2200