Peter Welch
U.S. SENATOR FOR VERMONT

Trustworthy by Design AI Act of 2024
Section-by-Section
Senator Peter Welch

Section 1 – Short Title: This *Act* may be cited as the *"TBD AI Act of 2024"*.

Section 2 – Definitions: Defines AI systems, and abbreviated references for the Director of the National Institute of Standards & Technology (NIST) and covered federal agencies.

Section 3 – Guidelines for Evaluation of Trustworthiness of AI Systems: Tasks NIST with creating AI system assessment guidelines and defines requirements for their creation.

Core Guidelines to Guide AI System Evaluation

- Sets a 1 year timeframe for NIST to develop a set of AI trustworthiness guidelines
- Defines covered components to include:
    - AI models themselves;
    - Data and related training activities, like collection and filtering;
    - Processing and later training activities, like fine tuning and reinforcement learning;
    - Content generated by AI systems;
    - Hardware used by AI systems; and
    - Risks presented by anthropomorphic AI systems.
- Defines elements of trustworthiness that guidelines must cover to include:
    - Validity and reliability
    - Safety
    - Security
    - Resiliency
    - Transparency and accountability
    - Explainability and interpretability
    - Privacy
    - Fairness and bias
    - Other metrics NIST deems appropriate
- Notes that guidelines must specifically consider accuracy/bias risks for protected classes.

- Requires NIST to identify risk management strategies for AI reliance on synthetic content.
- Clarifies that trustworthiness elements should assess all applicable covered components.

Additional requirements to ensure evaluation is comprehensive, appropriate, and adaptable

- Allows NIST to use and compile existing work to meet these requirements.
- Requires all work that meets these requirements be centrally located and publicly available.
- Asks for periodic, minimally annual, reevaluations of the guidelines.
- Ensures that guidelines consider iterative or ongoing evaluation to consider AI system design, development, and deployment.
- Asks NIST to develop guidelines with transparency, cooperation, and collaboration in mind, especially with developers or evaluators of AI, academia, and civil society.
- Requires NIST to report to Congress any expected barriers to implementing and adhering to these guidelines, highlighting any transparency, cooperation, or collaboration issues.

Section 4 – Federal Deployment of AI Systems: Discusses how USG agencies are expected to use these guidelines to ensure trustworthy deployment of AI.

- Covered Use is defined as:
    - AI systems used in any automated decision-making.
    - Does not include any use case that NIST exempts from the guidelines, including systems subject to other national security assessments or edge cases as determined by the Director.
- Establishes an effective date for compliance reporting that goes into effect on the date the guidelines are released.
- Requires existing, covered AI systems deployed before this Act to be evaluated against and meet these guidelines within 2 years of the effective date or cease using the AI system.
- Requires new, covered AI systems to be evaluated against and meet these guidelines prior to deployment.
- Requires heads of USG agencies to publicly report on the evaluation status and compliance details for all covered AI systems.
- Requires that AI systems not compliant or not evaluated within 2 years of the effective date report on noncompliance, to include progress made, justification for delay, and any barriers to compliance.
- Requires a report to Congress within 3 years of enactment from each head of USG agency on each deployment.

- Requires that heads of USG agencies designate Chief AI Officers to implement AI governance and oversight work, including compliance with guidelines.
- Details that CAIO's must be senior (GS-15 or higher) full-time employees.