HEALTH SECTOR AI COMMITMENTS

As health system and payer organizations engaged in the procurement, development, and use of large-scale machine learning models that can perform a wide variety of tasks (aka "frontier models") in healthcare, we commit to vigorously pursuing these technologies' once in a generation benefits while mitigating their risks and protecting patient's protected health information. The voluntary commitments our organizations are signing onto reflect a series of actions that underscore three principles that must be fundamental to the future of AI: safety, security, and trust.

COMMITMENTS:

1. We commit to vigorously developing AI solutions to optimize healthcare delivery and payment by advancing health equity, expanding access, making healthcare more affordable, improving outcomes through more coordinated care, improving patient experience, and reducing clinician burnout.
2. We will work with our peers and partners to ensure outcomes are aligned with fair, appropriate, valid, effective, and safe (FAVES) AI principles.
3. We will deploy trust mechanisms that inform users if content is largely AI-generated and not reviewed or edited by a human.
4. We will adhere to a risk management framework that includes comprehensive tracking of applications powered by frontier models and an accounting for potential harms and steps to mitigate them.
5. We will research, investigate, and develop swiftly but will do so responsibly.

We commit to taking these actions to optimize the use of predictive and other frontier model uses in health care. These baseline requirements for transparency aim to improve the trustworthiness of content generated by frontier models and support their widespread use in healthcare. Furthermore, we intend to revisit and revise these commitments at reasonable intervals as we learn more from their use.

ADDITIONAL DETAILS:

I) AI Development

We believe that AI is a once in a generation opportunity to accelerate improvements to the healthcare system, as noted in the Biden Administration's call to action for frontier models to work towards early cancer detection and prevention.

II) Honoring FAVES AI principles

We will work with our peers and partners to increase transparency aligned with the FAVES principles as we procure, develop and use "frontier models" in healthcare operations to ensure that the end-to-end process and outcomes frontier models support are:

- **Fair:** Outcomes of model do not exhibit prejudice or favoritism toward an individual or group based on their inherent or acquired characteristics.
- **Appropriate:** Model and process outputs are well matched to produce results appropriate for specific contexts and populations to which they are applied.
- **Valid:** Model and process outputs have been shown to estimate targeted values accurately and as expected in both internal and external data.
- **Effective:** Outcomes of model have demonstrated benefit in real-world conditions.
- **Safe:** Outcomes of model are free from any known unacceptable risks and for which the probable benefits outweigh any probable risk.

These principles were originally proposed in the Office of the National Coordinator for Health IT's (ONC) draft rule for transparency of AI-tools embedded in certified electronic health record systems. We believe these principles should be applied, where feasible, to all outcomes of frontier models being used to support clinical decision making.

III) Transparency in Development and Use of AI Technologies

We believe that the healthcare industry should be transparent about its use of frontier models in order to build consumer and clinician trust and to accelerate development of technologies that will optimize healthcare delivery. We commit to informing users if content is largely or exclusively AI-generated, unless such content is edited or closely reviewed by a human before being shared with end users. We also commit to sharing relevant best practices with industry and the public to accelerate responsible AI development in the health sector.

IV) Adhering to a risk management framework

Effective risk management of outcomes is essential in work pertaining to AI in healthcare. The Federal government has provided reference material (e.g., NIST AI Risk Management Framework, White House Blueprint for an AI Bill of Rights) for us to build upon. The National Association of Insurance Commissioners (NAIC) is developing a risk-based framework to be used by insurers. Each of us employ our own specific approach to risk management. However, we commit to adopting and adhering to a risk management framework covering the following practices:

1. **Risk Analysis** – We will analyze potential risks and adverse impacts associated with outcomes from all frontier models, including without limitation commercial, self-developed, or open-source derived frontier models. Such analysis will consider the particular use case (e.g., clinical or administrative) and the likelihood and degree of adverse outcomes.

2. **Risk Mitigation** – We will implement appropriate risk mitigation practices based on the level of risk identified in the Risk Analysis. In addition to testing use cases before deployment and ongoing monitoring of outcomes, we will consider human review of outputs as needed.

3. **Governance** – We will establish policies and implement controls for applications of frontier models, including how data are acquired, managed, and used. Our Governance practices shall include:
    i. Maintaining a list of all applications using frontier models; and
    ii. Setting an effective framework for risk management and governance, with defined roles and responsibilities for approving use of frontier models and AI applications.

V) Responsible Development and Innovation

We believe that one of the most powerful features of frontier models is their ability to democratize access to AI and thereby tap into the collective creativity of healthcare experts and practitioners. To realize this benefit, a new approach to the adoption and deployment of disruptive technology is needed, one that centers on:

1. **Rapid prototyping with guardrails** – We will leverage non-production environments, test data, and/or internally facing applications whenever practicable to prototype new use cases. We will meet our existing privacy requirements for protected health information such as HIPAA.

2. **Outcome monitoring with feedback loop**s – In addition to software development testing, we commit to ongoing monitoring of the outcomes enabled by use of these models. Such monitoring will include evaluating quality and accuracy by, for example, harnessing existing quality assurance programs or building human-in-the loop feedback mechanisms into the solution itself.

3. **Responsible use of open source models** – One of the most vibrant sources of innovation in the development and improvement of frontier models has been the open source community, and in the long-term society is best served by the competition engendered by open source AI initiatives. When using these models, where appropriate, we commit to mitigating risks associated with their use.

4. **Training** – We believe that one of the most powerful aspects of frontier models is their ability to make AI more affordable and user-friendly, allowing AI to be used and deployed by a wider audience. We will expect each person developing or using frontier models to undergo training concerning the safe and effective deployment of AI, including the need to avoid bias and maintain data security.