Jody Miller

Summary of The National Artificial Intelligence R&D Strategic Plan

The National Artificial Intelligence R&D Strategic Plan is a document that provides nine strategies for responsible federal research and development of artificial intelligence with a goal of advancing the technology for the nation's benefit while balancing risks. It is not meant to dictate governance measures, but to outline the chief goals of R&D as it relates to AI and serve as guidance to federal agencies. The plan also outlines metrics that will serve as a basis to evaluate federal agencies' progress toward these strategies moving forward.

Each strategy outlines the objectives as well as the risks within that category: (1) Make long term investments in AI research, (2) develop effective methods for human-AI collaboration, (3) understand and address the ethical, legal, and societal implications, (4) ensure the safety and security of AI, (5) develop shared public datasets and AI training/testing environments, (6) measure/evaluate AI systems through standards and benchmarks, (7) better understand workforce needs, (8) expand public-private partnerships, and (9) international collaboration.

There are many specific objectives within the strategies outlined; however, many revolve around similar points. One theme is obtaining data and creating datasets that are standardized, accurate, free from bias, are representative of and span a wide range of industries and are available to both federal and private national interests including more partnerships between public and private to gain more and better data. Another important point is creating standards and benchmarks to measure AI models for accuracy and performance, which includes ensuring that there is no data poisoning or other malicious activity affecting the model. This then further ties into developing environments that can be used to simulate and test the models and eventually for training users of the technology and developing the workforce.

Another theme that presents itself within each strategy and is its own strategy is the ethical, legal and societal implications, which span things like privacy in data collection, ensuring the AI is used in a way that is not harmful to society, eliminating bias or harmful outcomes, and more. The goal is to build safe and secure AI that feeds the goal of advancement versus causing harm to individuals or society as a whole. It's noted that some of the safety risks are difficult to quantify, which requires further research especially as the datasets become larger, the systems become more complex, and as multiple AI systems become intertwined with each other. Securing these systems then becomes highly important as the disruption they can cause will affect a larger and larger proportion of our daily lives. Finally, the report outlines the collaboration required for the best

possible outcomes with this technology and that includes public and private partnerships as well as international collaboration.

Summary International Standards Organization (ISO)

The International Standards Organization is an international collaboration of experts to "ensure that the products and services you use daily are safe, reliable and of high quality. They also guide businesses in adopting sustainable and ethical practices…" An example that comes to mind is ISO standards for the network layer ensuring that all equipment built to interact with the network layer can communicate the same, otherwise people could have created many different standards and it would be difficult to then further connect disjointed systems. Having international standards that all follow ensures that everything can be interconnected and used on a global scale.

The ISO Strategy 2021-2030 is a long-term plan to evaluate their own work as an organization by measuring its continuous improvement and sustainability priorities against actual progress within this timeframe. The plan outlines four major drivers of change: the economy, technological progress, the environment, and society. The overarching goals are to have ISO standards everywhere by promoting research on standardization, meeting global users needs for standards by monitoring developments across many industries, and making sure all voices are heard.

The ISO also has standards around artificial intelligence (AI) with a technical committee referenced as ISO/IEC JTC 1/SC 42 providing guidance on AI standardization. The committee has published 34 ISO standards with 38 under development. One example of a published standard is ISO/IEC 5259 which focuses on data quality to improve the reliability and accuracy of machine learning models and analytics as well as standardizing data quality assessments. Another standard is ISO/IEC TS 12791:2024 which provides mitigation techniques to prevent and eliminate unwanted bias in classification and regression machine learning tasks. Finally, ISO/IEC 23894:2023 describes ways to integrate risk management into their AI development and usage as well as processes for effective implementation and integration of AI risk management.

The ISO is committed to continuously improving international standards, researching and publishing new standards through specific committees that have a global membership, and keeping abreast of new technologies and user experiences to create and improve standards. A question that comes to mind is how these standards are used and how they are measured at a global scale for their effectiveness and adoption, which ties back into

their 2030 plan and how they intend to ensure value is being provided against their objectives.