To: Kei Koizumi, Deputy Director of Policy, Office of Science and Technology
From: Daniel Bullock, Senior Staff Officer

Modern AI's impacts are profound but manageable

Executive summary

Contrary to sensationalist accounts, modern AI is a substantial, but not unprecedented advancement. Successful management of AI's potential will require: (1) leveraging insights from dual-use technologies, (2) developing governmental interfaces and expertise, (3) conscientious assessment and implementation of incentive systems.

Background

Current AI: The term "artificial intelligence" refers to a broad category of advanced computational methods for processing and leveraging information. Currently, these methods are sufficiently complex so as to prevent their direct production (e.g. coding or programming). Instead, they are created via a process of "training" or "learning", using large, task-relevant datasets. This is reflected in the newfound prominence of "machine" or "deep" learning.

Current Policy: The National AI Initiative Act of 2020 called for the establishment of the National Artificial Intelligence Advisory Committee (NAIAC), which is still under formation. The Office of Science and Technology Policy (OSTP) and National Institute of Standards and Technology (NIST), among others, have sought input and made initial recommendations relating to topics like privacy, safety, bias, and security. Comprehensive policy or regulation is still lacking. Many proposed policies (e.g. the "AI Bill of Rights") feature requirements for AI systems that are not imposed on existing, human-led decision-making processes.

Assessment

Current concerns about AI tend to reflect conventional, even pre-technological categories of societal harm. For example: biased systems, arms races, and privacy concerns all exist independent of AI. Instead, AI's notability can potentially be traced to two characteristics: (1) its method of production and (2) its extensibility.

Production methods: Modern AI is derived from data, which determines the performance profile of the AI system. Generally, "bigger" data permits more complex or precise behavior, more sensitive data permits more potentially grievous applications, and incomplete, misused, or biased data leads to flawed or brittle systems. Thus, the economic and security value of a dataset is roughly proportional to: size x sensitivity x completeness.

Extensibility: AI can be applied to automate nearly any information processing or usage task. Its cost of implementation typically reflects initial development and ongoing hardware/software maintenance and operation expenses. These post-development costs are proportional to the scope of the system, and for small to medium systems can be negligible, thereby facilitating rapid and/or potentially effortless implementation. By automating tasks, AI can substantially reduce logistical demands for the operator, per unit outcome/production.

Recommendations

Incentive systems: Assessment of current incentive structures and calculated implementation of novel incentive regimes is advised. These could include targeted regulatory systems, taxes or penalties, and research funding. Such incentive structures should effectively discourage undesirable outcomes (e.g. robust "bug-bounty" systems to identify and fix flaws; accountability systems to address negligent uses) and/or promote desirable outcomes (e.g. continued advancement of general methods; targeted development of beneficent applications).

Bolstering governmental AI expertise and capabilities: Government AI specialist recruitment should be expanded aggressively. Private sector opportunities for AI specialists are expected to persist as sources of competition for governmental efforts to expand specialist resources. Minimally, ongoing collaborations between private, academic, and governmental organizations should be fostered to bridge this shortfall.

Leverage previous insight: Best practices from existing "dual-use" domains should be adapted and expanded. A great deal of insight on managing disruptive technologies has been developed working with nuclear, communication, and other cyber technologies. For example, establishing hazard metrics (as outlined in **Production methods**) for precursor materials (i.e. data) is advised. Formation of specific AI/data regulation agencies (modeled after FCC or FTC, for example) should be undertaken.