**Testimony of Dr. William L. Scherlis**
**Professor of Computer Science, Carnegie Mellon University**

**before the**

**Subcommittee on Cybersecurity, Information Technology, and Government Innovation**
**U.S. House of Representatives**

**January 17, 2024**

**"Meeting the Challenges of Developing an AI Workforce"**

Chairwoman Mace, Ranking Member Connolly, Members of the Subcommittee, thank you for the opportunity to participate in this important hearing. My name is William Scherlis, and I am a professor of computer science at Carnegie Mellon University. I am honored to join the witnesses who have testified in the previous AI hearings, as well as my fellow witnesses today, who are advancing vital partnerships, programs, and services to expand and democratize access to AI training and education.

My remarks draw upon more than four decades of experience in research and education in software, cybersecurity, and artificial intelligence (AI). I also have the honor of having served two tours in government at the Defense Advanced Research Projects Agency (DARPA), with a mission to advance innovations in information technology, including AI and cybersecurity, that are pertinent to defense and national security, and with benefits to our economic competitiveness. My remarks also reflect my experience living and raising a family in Pittsburgh, Pennsylvania, a community that has experienced the benefits of research, education, and technology advances in computing, robotics, and AI – supported by strong educational institutions – in economic revitalization and in extending the reach of innovation to all citizens.

Workforce education and training that is focused on AI, both within the government and across all sectors, is vital to the continued competitiveness and prosperity of our nation. In this testimony I discuss some of the salient characteristics of modern AI and its applications, including examples of the extraordinary benefits, but also of the weaknesses and vulnerabilities that can impede successful and safe application of modern AI technologies, particularly to government applications. This highlights the wide range of kinds of AI work – and the wide range of expertise and skills important for the various segments of the AI workforce. With this background, I present some examples from successful education and training programs at Carnegie Mellon. These are concrete illustrations of the diversity of approaches and partnerships that are needed to advance our AI workforce, within the government and broadly across the Nation.

## 1. Introduction

The charge to this hearing focuses on developing an AI workforce that enhances American strength and prosperity. This could not come at a more critical time. We see an accelerating pace of innovation in artificial intelligence and also in development of a growing range of applications. We are already seeing transformative impacts on defense and security, and on economic growth and opportunity in many sectors – and these benefits accrue at both national and regional levels, and include the delivery of diverse cost-effective services directly to our citizens. Anyone who uses a mobile phone, for example, is

affected by AI-enabled developments – which include speech recognition, face recognition, navigation assistance, photo search and categorization, fitness tracking, online shopping recommendations, and so on. AI applications do more than enhance productivity; they create new kinds of value and capability.

It will be an all-nation effort to fully realize the benefits of AI. Importantly, modern AI has weaknesses and vulnerabilities as well as tremendous potential benefits, and so we must develop and operate AI applications in a manner that is safe and responsible. The impacts will be transformative, analogous to the early years of the industrial age more than a century ago. We face an upskilling challenge and imperatives as great as any time in our history. In short, the opportunities are huge, the competition is fierce, and there are risks and challenges to be faced.

I want to start with an expression of support for H.R. 4503, the *AI Training Expansion Act*. We have learned through experience the essential importance of taking explicit steps to sustain technological currency in the federal workforce. This is now especially true for AI literacy and skills. The federal agencies can and should be a pillar of a national AI education and workforce strategy. Such a strategy can enable the affordable development of AI applications that are effective and capable – but also trustworthy and responsible.

In addition to the immediate benefits, a comprehensive training initiative for the federal workforce will act as a force multiplier for deepening AI skills in regions across the nation. In this regard, the *AI Training and Expansion Act* could work in tandem with the *AI Scholarship for Service* program that is authorized in both the *National AI Innovation Act* and the *CHIPs and Science Act* to deepen both technical talent and AI readiness across the federal workforce. Further, an understanding of AI and its capabilities is vital to the federal workforce as it works to advance the economic development of our nation while enhancing our safety and security.

My testimony today offers perspective on AI and on the particular dynamics of this moment in our technological advancement. Recognition of these dynamics, including both positive and negative aspects, is vital to the design and effective implementation of AI education and workforce programs and strategies. This is because we must take care to address the weaknesses, vulnerabilities, and responsibility requirements that come along with the astonishing AI capabilities now being developed.

## 2. AI at Carnegie Mellon

I benefit from a work environment where there is a tremendous diversity of perspectives on AI technology and applications – not everybody agrees. This is a strength of my university, Carnegie Mellon (CMU), which is an R-1 research university with six decades of history of research and education in computing and AI. This strength is evident in the legacy of accomplishment of my colleagues who have been early pioneers of speech recognition, computer vision, and robotics, for example, with research in these areas now having impacts on billions of people.

In responding to the charge to this hearing, I draw on insights from CMU's long experience in developing AI education and training programs. We have been involved in AI since Herb Simon and Allen Newell's formative work in the 1960s in the basement of our business school. PhD students across the university have been completing dissertations on AI-related topics since the 1950s. The thread of AI work related specifically to machine learning advanced to the point that in 1997 we created a research and education center, and in 2006 this center graduated into a full-fledged academic department.

There are more than two dozen AI-related master and PhD programs, and Carnegie Mellon now offers the nation's first undergraduate AI degree. The curricula draw on diverse disciplines including computer science, statistics and data science, mathematics, ethics, and humanities and arts. This multi-disciplinary approach gives graduates a strong technical foundation in AI that will endure despite the rapid pace of AI advancement. We also have efforts in support of K-12, post-secondary education, professional masters programs, and workforce training. This experience also includes the development and delivery of comprehensive AI training and education to enhance the federal workforce. (A number of specific examples are given below.)

A key observation, based on this experience, is that there is a broad and diverse range of AI-related work roles and, consequently, we must necessarily offer a broad range of approaches to AI education and workforce training. At CMU, our education and training offerings include not only our degree programs and courses, but also outreach programs that span the continuum from early education and informal learning to flexible and focused training capabilities to serve the needs of workers at all levels. These levels range, for example, from ML data curators and LLM prompt engineers to program managers and engineering leaders responsible for developing AI-based systems. It is important to include organizational leaders, so they can be more fully aware of the less obvious features and weaknesses of modern AI – and thus be more effective leaders of organizations that will become increasingly AI-reliant.

**Collaboration in research and education**. A comprehensive AI workforce strategy will likely require development of new models of collaboration across educational institutions at all levels and among industry, government, and academia, due to the diversity of needs, the diversity of contributing disciplines, and the fast pace of progress. At CMU, we benefit from a close coupling of research and education. This includes extensive collaborations across departmental boundaries, as well as strong "live-in" partnerships with industry and government. These partnerships help ensure that research is well coupled with the challenges of the world as well as being creative, principled, and long-sighted. All stakeholders benefit when we are able to advance research concepts into prototypes that can be tested in real-world settings. This coupling is also important in ensuring that our curricula are technically current and have sufficient application realism.

**Human-systems interaction**. A strong theme in Carnegie Mellon computing research and education is human-systems interaction. This field draws on computer science, design, AI, and psychology and cognitive science. Work in this area includes understanding and improving how people can interact effectively with AI-based systems, including with robots. One key area of focus is the development of AI-based educational tools, informed by cognitive science, that can provide individualized tutoring at the K-12 level that is dynamically tailored to each student's learning style, pace, and background knowledge. This work builds on development of educational AI-based tools pioneered by Herbert Simon and John Anderson at CMU beginning as early as the 1970s. AI-based tutoring is being advanced both within the university and at spin-offs, and the resulting tools have been used by millions of students – and with measurable outcomes. (The spin-off is Carnegie Learning, founded in 1998.)

## 3. The dynamics of modern AI – what makes this moment different

An effective AI workforce strategy must flow from the technical characteristics and market dynamics of AI. AI is an expansive discipline with a long history. The term "artificial intelligence" was coined in a 1955 proposal for the first technical workshop focused on AI, staged in 1956. Concepts of cognitively capable systems go back much further, including the *Turing Test* proposed by Alan Turing in 1950 as the *imitation game*.

**The scope of AI and its applications**. Early AI models performed most effectively at narrowly defined tasks. AI players for Chess and Go, for example, now surpass the top players, thus manifesting super-human performance. Modern machine learning (ML) and large language models (LLM) capabilities, however, are proving useful in much more broadly scoped and less well-defined tasks. This is one of the reasons why we are seeing such a strong focus on these technologies, with a vast range of applications being developed – and why we are engaged in this urgent consideration of AI workforce.

Because of this, the AI workforce must encompass applications as diverse as health care, financial services, intelligence analysis, life sciences, robotics, manufacturing, and many others. Applications also include support for AI roles in the processes of scientific discovery and the engineering of complex systems.

**Neural network models**. Modern ML and LLMs are based primarily on neural network models, which are inspired by the neurophysiology of the brain. We think of neural network models as recently emerged computational creations, but in fact they draw on a long history of research in statistics, psychology, mathematics, and computer science that goes back to the 1940s. Importantly, ML and LLM models are, at a technical level, members of a large family of predictive statistical models, and are in most cases unavoidably approximate in their outputs.

Modern AI applications are emerging largely due to the confluence of access to extensive data, availability of powerful computing, and the creation of capable and parallelizable ML and LLM algorithms. This enables neural nets to be efficiently trained to support tasks such as recognition and classification with inputs ranging from text and speech to images and signals. As predictive models, they can be turned around, so to speak, and used as generators of text, images, speech, and signals. LLMs, for example, are very good at predicting candidates for the next word in texts, as in "*the dog chased the* ____." They are so good, in fact, that they do not need to be taught grammar – they assimilate it from their training data that sets the patterns that drive the words predicted. By predicting one word after another, they become generators of texts.

It is important to keep in mind that ML and LLM technologies are part of a rich AI ecosystem. Indeed, AI is not a single technology, but rather a bundle of interrelated technologies that span algorithms, massive data, logical planning and reasoning, computational software and hardware, sensors for perception, and actuators for action. Although the focus in this testimony is primarily on ML and LLMs, very different kinds of AI capabilities are also advancing rapidly.

**Modern AI risks, weaknesses, and vulnerabilities.** The democratization of access to advanced AI applications is fostering rapid adoption across government and a multiplicity of industry sectors – as well as by individuals who make use of LLMs. As with all novel systems, there are risks of harms when systems are not properly engineered and tested, when they are used for unintended purposes or domains (for example, domains poorly matched to the training data used), or when they are designed and used for nefarious purposes.

Modern ML and LLMs, however, additionally present weaknesses and vulnerabilities that are particular to the underlying neural network technologies – and their statistical nature. These create potential for additional kinds of harms. Modern AI systems can give wrong answers, LLM "hallucinations" for example, even when the training data are fully accurate. They can be deliberately – and easily – misled to give wrong answers. They can be used to develop deep fakes and operate cyber-attacks. These weaknesses are particular to neural-network models and can be non-obvious and subtle, especially when compared with other kids of computational capabilities. (More examples are below.) This is

another important reason why AI workforce education is so essential – developers and users must understand not only how to realize the benefits, but also how to avoid the pitfalls derived from the particular characteristics of the AI systems.

This is why, if we are to talk about workforce, we must first talk about the work. The work includes not only fielding and using new applications, but also recognizing and mitigating the weaknesses and vulnerabilities that are associated with neural network models such as modern ML and LLMs. Mitigation actions encompass development, fielding, and operations for AI applications.

## 4. ML and LLM developers and users need to be aware of weaknesses, vulnerabilities, and limitations of current capabilities.

The fundamentally statistical nature of the neural network models of ML and LLMs means that adopters of these technologies must be alert to the weaknesses and vulnerabilities that are generally associated with these kinds of models. Many of the weaknesses are well known, such as potential bias in data and privacy exposures. Other weaknesses, perhaps less well known but amply documented, relate to the potential for adversarial actions on neural network models. There are a number of well-known examples where image inputs are given that include certain features that are not readily apparent to human eyes but that have the effect of misdirecting machine learning nets.

**Risks.** It is often argued that with any new technology there can be "long tail" risks and harms that are rarely encountered but that may be highly consequential. Unfortunately, however, many of the most significant ML and LLM weaknesses are not rare and unusual, but rather easily triggered, even to the point where they can be the subject of undergraduate homework assignments in introductory AI courses.

It is therefore a key challenge in AI research not just to enhance accuracy of results, but also to reduce vulnerability to adversarial attacks. There are differing views among researchers regarding the extent to which many of the weaknesses (sampled below) are intrinsic, as well as what steps might be taken to mitigate those weaknesses. AI risk analysis is further complicated by potential trade-offs – some researchers argue, for example, that there are technical tradeoffs between accuracy and attack resistance.

The bottom line is that system developers must be mindful of weaknesses and mitigations as they craft new applications.

**AI Engineering.** Generally speaking, ML and LLM capabilities are almost always employed as parts of larger systems and organizational workflows. This context is important, because adaptations to mitigate the effects of weaknesses can be made not just to the ML and LLM capabilities (and their training data), but also to the design of the systems encompassing these capabilities and to choices regarding how the systems are employed in organizational workflows.

It is therefore important to mention *AI Engineering*, which is the practice of designing, developing, testing, and evaluating systems and workflows that embed AI capabilities. With the right kinds of system designs and operational choices, it becomes more often possible to exploit fundamentally un-trustworthy ML and LLM components in the construction of systems that are both useful and trustworthy for certain critical applications. AI engineering may also encompass use of risk models to inform choices regarding the scope of application use cases – for example restricting operational

contexts to those deemed safe and responsible. Developing and advancing a practice of AI Engineering is a focus of activity at the CMU-affiliated Software Engineering Institute (a DoD FFRDC). AI Engineering is also the focus of a new family of master degree programs in the engineering college at CMU, with individual programs focused on seven different engineering disciplines and an additional program at the CMU Africa campus.

**Test and evaluation.** Testing, evaluating, and certifying modern AI components and AI-based systems remains a major technical challenge and subject of research. This topic is highlighted in the recent Executive Order 14110 on *Safe, Secure, and Trustworthy Development and Use of AI* (November 2023). The EO identifies a critical role for adversarial "red teams" in evaluating AI-based systems. Although red teams are a familiar and established modality in cybersecurity practices, the skill sets for AI red teams will mostly be different. AI red teams must understand data analysis, adversarial AI, AI Engineering (including software and hardware), and considerations that derive from relevant application domains. These domain-focused topics could include, for example, considerations of responsible AI such as privacy protection, ethics, fairness and bias, and other considerations particular to application domains (health care, national security, civil infrastructure operations, etc.). AI red teams will also need to have knowledge of cybersecurity, software, and computing infrastructure.

Concepts of operation for AI red teams are nascent. These concepts of operation will encompass tools, practices, and team-member expertise and skill sets. Because of the rapid pace of change, these concepts may best be framed in ways that will readily support ongoing update in response to changes in the environment, including advances in technical AI, improved understanding of AI-related operational risks, codification of norms and practices for responsible and trustworthy AI, broadening of relevant use cases, and potential adversarial actions.

**Responsible AI.** There is an extensive array of policies, guides, and checklists regarding the safe and responsible use of AI, under rubrics including *Responsible AI*, *Trustworthy AI*, and *Reliable AI*. This is because some of the most vexing challenges in the effective and safe use of modern neural nets and large language models derive from the non-obvious nature of the weaknesses and vulnerabilities that they can manifest in the hands of developers and users who are not fully aware of these pitfalls. As noted above, savvy system developers and sophisticated users also struggle with many of these challenges.

For these reasons, there are now many carefully formulated articulations of Responsible AI principles and guidelines, including from the Department of Defense, major technology firms, system integration firms, and others.

## 5. A sampling of kinds of weaknesses and vulnerabilities.

What follows in this section are some examples of the kinds of issues that AI developers and users face as they work to develop AI-based systems that they hope to rely on. These examples have been identified and developed over the past decade or so by diverse researchers in the AI community, including colleagues at Carnegie Mellon.

Note that this is an illustrative sampling, not intended to be comprehensive. There are a number of inventories of AI-related risks, including an AI risk management framework released by NIST in 2023.

Examples of weaknesses and vulnerabilities often associated with neural-network-based ML and LLMs:

- **Bias and fairness**. When training data are not well aligned with the intended AI use case, models are more likely to produce incorrect outputs. In many contexts, this can lead to adverse societal outcomes, for example, when AI is overly relied upon in supporting decisions regarding granting of credit or doing face recognition for security. This misalignment is often caused when ML networks are trained with data that is opportunistically on hand, even when it is not a good fit (from the perspective of traditional statistical methods) for the questions to be addressed.
- **Accuracy of outputs**. As noted earlier, the probabilistic nature of neural networks can lead to incorrect outputs even when training data are fully accurate. This is intrinsic to statistical models. Additionally, when data are not fully accurate, a small number of poor training cases can disproportionately influence results.
- **Transparency and explainability**. In many cases, the reasons why a neural network reaches a particular output or prediction can be elusive. Computational neural networks have many billions of parameters that are continually adjusted as data is ingested in the learning process, but there may be no "place to look" among these parameters that reveals rationale for a particular conclusion. There is significant research focused on creation of explanations for AI outcomes, and there are some good results. But larger neural network models are in many respects impenetrable – and not just to evaluators but also to their creators.
- **Reliability and predictability.** As noted, models sometimes produce erroneous outputs, even with the highest quality training data. This is primarily due to the statistical nature of the networks. But it can be made worse by poor training methodologies, for example that lead to statistical overfitting and other problems.

Weaknesses and vulnerabilities of ML and LLMs in the presence of adversaries:

- **Spoofing and misdirection.** Adversarial machine learning (AML) involves attacks on neural nets that can be delivered through a variety of pathways. These range from "poisoning" elements of training data to providing seemingly benign inputs that look unremarkable to human perception but that contain elements that misdirect a neural network. Sometimes these elements are almost invisible, appearing to us as low level noise. Other times they can be visible but seemingly benign. For example, tiny patches of color, such colored dots on a person's eyeglass frames, can reliably and repeatably mislead a face recognition net into a false identification – and to an identity chosen by the adversary. Often, the exact mechanism of an AML attack (such as the colors of those dots) may be developed through the use of separate ML-based systems that are used to develop the potential input elements that can mislead the ML network of interest. (Lujo Bauer, CMU)
- **Privacy of training data.** In many cases, such as health care applications, a corpus of training data is developed by extracting elements of data such as from the medical records of a diverse cohort of patients. It might be expected that the many inputs are aggregated and abstracted in ways that might eliminate any possibility of the resulting trained neural net exposing particular data elements from the training data. The technique of *model inversion* demonstrates that this is not always so – it is sometimes possible to extract individual data elements. (Matt Fredrikson, CMU)
- **LLM guardrail bypass**. LLM providers are taking diverse actions to limit adverse uses of the LLMs, for example to prevent them from providing instructions on how to commit credit card fraud. The providers might take actions such as additional training ("fine tuning") that cause the LLMs to deflect these kinds of requests. Researchers have shown that certain kinds of prompts, which appear to contain nonsense texts, can cause the LLM models to bypass those safeguards and provide the fraud instructions, for example. (Zico Kolter and Matt Fredrikson, CMU)

Test and evaluation challenges, in addition to addressing the weaknesses and vulnerabilities identified above:

- **Test coverage.** Software testing in best practice is informed by deep analysis of the software code to identify pathways and decision points – and to ensure that test cases cover most or all of these. For neural nets, it may be intractable to identify "boundaries" among output categories within the full set of possible inputs. A historical example famous in the ML community is reading handwritten five-digit zip codes on postal envelopes (the MNIST dataset). A sloppy "9" could look like a "1" or perhaps a "7." In this limited space it is relatively easy to develop tests to understand these "edge cases." But for more sophisticated modern nets, it may be very difficult or intractable to do this, and consequently to develop useful assessments of "test coverage." Indeed, AML spoofing attacks can build on this difficulty.
- **Specification**. A closely related point is that an ML network is trained according to its data inputs. Unlike other engineered systems being assessed, ML may lack any specification of "correct" or "intended" behavior beyond the corpus of data in its training set. But we nonetheless want it to operate correctly on a full universe of inputs – not just exact matches with the training cases.

Means to defend against adversarial exploitation of ML and LLM capabilities:

- **Multi-modal deep fakes**. We are all familiar with the use of diverse tools, including generative AI, to create deep fakes for images, videos, speech, text, and potentially other modalities. Needless to say, techniques and tools are being actively developed that can detect many of these. And, additionally, many of the generation capabilities are providing "watermarking" capabilities to create signatures for generated outputs. Both creators and detectors of deep fakes are growing in sophistication, so the battle will be ongoing. (The DARPA Semantic Forensics research program tackles some of these challenges.)
- **AML**. As noted above, AI Engineering capabilities and related techniques can assist in mitigating against adversarial ML attacks on critical systems.
- **Social media generation and bots**. It is well understood that ML and LLMs are tremendous force multipliers for generating media postings. Social media firms are also (reportedly) using ML to detect and deflect such attacks. As above, creators and detectors are both growing in sophistication, so the battle will be ongoing.
- **Cyber-attacks**. Also as above, there is an ongoing battle.

## 6. Looking ahead with AI.

My reason in presenting the above sampling is to emphasize that members of the AI workforce need to be constantly aware of the challenges and also to understand the best practice mitigations for those challenges that pertain to their applications. There are ongoing discussions in the technical community regarding how to improve mitigations but also which of the various weaknesses are intrinsic to the current set of technologies and which might be overcome through additional incremental effort.

This means that, in addition to careful application of AI Engineering techniques, we also need to accelerate the development of improved approaches to core AI. There is strong demand for the use of AI in critical applications where verifiable trustworthiness is essential. One class of examples includes faster-than-thought autonomy for critical applications in civil infrastructure operations, emergency response, and national security. This suggests that, in addition to advancing practices to mitigate or

bypass the various weaknesses and vulnerabilities, we also place a priority on advancing research towards AI capabilities that are more likely to be verifiably trustworthy.

Although there is much less mainstream visibility than for ML and LLMs, there is extensive use of AI techniques other than ML and LLMs in applications such as robotics, vehicle autonomy, speech recognition, and others. Some of these techniques also have a long heritage, for example logic-based techniques and knowledge representation techniques. Many of these are now used at massive scale in tech firms, for example to verify correctness of cloud security policies or to provide basic facts in response to search queries.

There are also techniques such as optimization and game theory that harmonize well with neural network designs. AI for applications such as strategic planning with multiple adversaries benefit from the hybridization approach, combining ML with traditional techniques from optimization and game theory, as well as additional techniques. An example of such strategic planning is poker playing – and AI-based systems have prevailed at the poker table in thousands of hands played against leading professional poker players. (Tuomas Sandholm, CMU)

Later AI innovations may be more profound, for example combining the statistical methods (ML and LLMs), which are fundamentally heuristic or advisory, with symbolic logic-based methods that can produce evidence-based argumentation. Statistical methods are excellent at finding possible solutions to many hard problems (for example protein folding), but they sometimes get wrong answers. This motivates us to use symbolic methods to check results. There are diverse perspectives within the AI community in this regard, with some researchers emphasizing continued advancement of pure neural-network models, and others advancing the necessity of hybrid approaches, at least in cases where trustworthiness, safety, and explainability are necessary. Needless to say, both approaches are being explored, with the diversity of perspectives benefiting the overall advance of AI capabilities.

A consequence is that AI developers and users who are stymied by weaknesses in current ML and LLM capabilities may benefit not only from mitigations derived from AI Engineering, but also from new kinds of core AI techniques. (My personal technical opinion is that it is likely incorrect to think that ML and LLMs based purely on neural-network technologies will soon evolve into a state of confident trustworthiness. But I do believe that hybrids of these technologies with other kinds of AI techniques may materially improve prospects for many aspects of trustworthiness.)

## 7. Implications for AI workforce development

It should be clear from the foregoing that there is a wide range of skills, technical background, and depth of training needed to cover the full range of kinds of AI work. Workers focused on data wrangling in an application domain require a base of foundational knowledge, application domain knowledge, and technical skills very different from workers focused on the fine-tuning and prompting of LLMs used to help software developers. Red team members require a range of additional skills such as diverse techniques and approaches for adversarial AI, data tampering, cybersecurity modeling, and the like.

Foundational knowledge will draw in varying degrees on diverse disciplines including computer science, mathematics and statistics, engineering for software and hardware, data science, cloud computing, and psychology. Application focused AI workers need to augment this core AI knowledge with subject-matter expertise in areas of application such as national security, life sciences, finance, manufacturing, and education. The subject-matter expertise must also include an understanding of responsible AI situated in the context of the application. That could include treatment of personally identifiable

information (PII) and information sensitive to business or mission, as well as principles of ethics applied to the creation and use of statistical capabilities of ML and LLMs.

In summary, the goal of HR 4503 – to expand AI education, training, and readiness to the much broader (and rapidly expanding) range of federal employees involved in AI – is vital to the success of the federal government in affordably developing and deploying a broad range of applications.

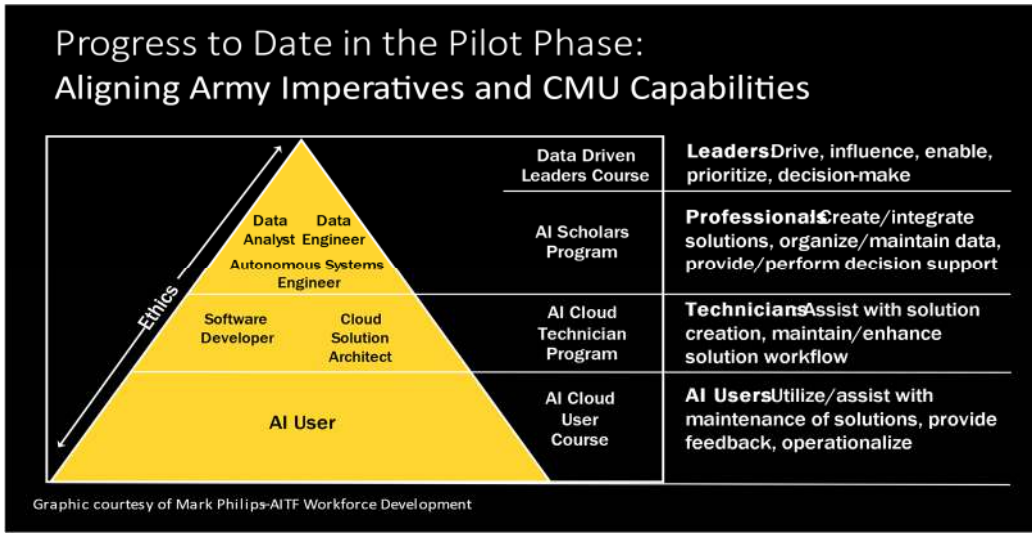
## 8. Building a National AI Workforce and Education Strategy

It is fair to expect that the broadening capability and scope of AI applications is leading us to a future where AI will be a component of nearly every job. For the federal government this will require focused initiatives to build AI readiness among the broad workforce. This includes the men and women who are already playing a critical role in the adoption and use of AI applications to serve core missions and the American public. It also includes many of those who will be front-line operators and users of AI-based applications. Compliance with Executive Order 14110 regarding safe, secure, and trustworthy development and use of AI will additionally require a cadre of employees with deep technical understanding and sophisticated skills who can support the required red teaming initiatives that agencies must implement.

Creation of an AI-ready workforce will require policies and partnerships that can advance a spectrum of AI skills, with tailoring based on characteristics of the application as well as the roles and responsibilities of individual workers. This must include general AI fluency, but also specialized skills (as noted above) related to specific application domains, AI engineering, test and evaluation, and the like. Additionally, it is important for there to be an informed working awareness of AI capabilities at executive, leadership, managerial, and supervisory levels, as noted in the proposed legislation. This is important because AI technology continues to advance – and the landscape of weaknesses and vulnerabilities will shift accordingly.

**An example of a partnership model.** For larger organizations, there is an imperative to adopt education and training strategies that can be rapidly scaled. This warrants broad partnerships and a range of delivery models. The Carnegie Mellon partnership with the Army Futures Command (AFC) and Army AI Integration Center (AI2C) demonstrates one model of building a flexible but comprehensive approach to building AI readiness across the organization.

The pyramid diagram below (from AI2C) illustrates the strategy to build capabilities ranging from general fluency, to technical readiness to support the implementation of AI applications within units, to building an actionable understanding of AI at executive levels where decisions are made on the deployment of capabilities. The strategy combines certificate, formal degree, and executive education programs. For example, a separate targeted executive course has been developed and delivered for Army acquisition staff.




It is a strategy that recognizes that effective adoption of AI capabilities requires the development of an organization wide "AI culture." Similar models are emerging in the private sector. For example, Carnegie Mellon has a partnership with Moderna to craft strategies for ensuring that there is AI readiness across the entire organization.

In developing an AI workforce to advance American strength and prosperity, we must not only realize full potential as effective users of AI, but we must also remain organizationally capable of safely assimilating the fast pace in improvements to AI capability (and trustworthiness) over time. Many of those AI improvements are incremental, for example in the performance and size – initially larger, and (more recently) smaller without loss of function – of modern LLM models. Others will be more transformational in capability and/or trustworthiness.

## 9. Carnegie Mellon programs illustrate diverse approaches to preparing and sustaining the AI Workforce.

To illustrate the diversity of approaches to education and training, I conclude with some examples taken from the many Carnegie Mellon programs that address the challenges of preparing and sustaining the AI workforce. These include the many degree offerings mentioned at the outset. Our non-degree programs enable us to extend our reach beyond our resident student population to diverse members of the expanding AI workforce. The non-degree programs include a number of executive education programs to build AI awareness, understanding, and skills at a variety of levels – from executives and strategists, to managers and practitioners. There are programs aimed at career tracks including government, business, and technical. And, we have programs aimed at aspiring AI minds at the pre-college level. Here are several examples:

**Government training.** One of the lessons of our experience is that education and training initiatives are enhanced when we include *AI design workshops* that engage workers within an organizational unit in the development of a particular AI application aligned with their mission. This hands-on approach (led by Rayid Ghani, CMU) has assisted agencies at all levels of government to design AI and machine learning applications that address mission-critical operational issues. Projects have included the development of applications ranging from reducing recidivism rates for individuals in need of mental

health services to better targeting potential recipients of social services to improving the delivery of services to at-risk students. The value of these design workshops is to deepen a working understanding of the potential of AI towards building broadly useful applications as well as an understanding of the imperative to address the potential for applications to create bias and other harms.

**AI in public schools – CS Academy.** Recognizing the need to accelerate AI education and provide resources to school districts and teachers, Carnegie Mellon launched the *Computer Science Academy* in 2018. CS Academy provides free curricula that enable students to learn basic programming, engage in exploring programming applications and even secure low-cost courses that provide Carnegie Mellon course credit. Critically, the program also provides support to teachers – who come from a variety of backgrounds – and it can be adapted to support students at different grade levels. CS Academy was initiated in response to a request from teachers in the Pittsburgh Public Schools. It is offered at no cost (other than for college-level credit), and it has scaled rapidly. CS Academy is now in use in all 50 states by more than 7,000 teachers with more than 380,000 students including, for example, more than 17,000 students in Virginia and more than 4,400 in South Carolina who have taken advantage of the offerings.

**National Science Foundation – AI for K-12.** With support from Congress, the National Science Foundation has been building capabilities for K-12 AI education that can be scaled across the nation. The curriculum is intended to enable school districts to adapt the delivery to local needs and education approaches. Carnegie Mellon helped lead the development of this curriculum and is assisting in piloting an elective for middle school students in Georgia. The current effort (led by David Touretzky, CMU) is engaging over 1,000 students in nine different school districts in the state.

**Community colleges.** Collaboration of universities and community colleges is also vital to meeting the workforce challenge. Community colleges have broad reach, and can operate online, in-person, and hybrid methods of learning. The *Social and Interactive Learning* – "Sail()" – platform is designed to enable community colleges to rapidly build the capacity to deliver certificate level training. Over 40 community college systems across the nation utilize this CMU-led platform to provide certificates in IT careers. In many cases collaboration with industry can enable these certificate programs to be tailored to specific local career and job opportunities.

**Outreach.** We have also developed a number of outreach programs centered on AI. *CMU CS Pathways* works with communities and organizations to develop programs and initiatives that create a more equitable and inclusive journey to computer science opportunities. The model is predicated on a recognition of the need to improve access for students who are traditionally under-resourced and often underestimated. The goal is to provide a model to greatly increase the STEM workforce and opportunities to participate in it.

An example of direct engagement with workers – for both innovation and training – is a partnership of Carnegie Mellon with the Technology Institute of the AFL-CIO, with support from NSF and U.S. DOT programs. This partnership is engaging hotel and transit workers in bringing their perspectives to the design and development of relevant AI technologies. The goal is to more effectively engage workers and provide insights towards the development of better training programs. This kind of stakeholder involvement is essential to success in the design of complex human-system interactions. The initiatives have engaged universities across the nation and are attracting interest from the companies that develop AI tools.

**Workforce transformation**. As noted earlier, modern AI capabilities are not only amplifying productivity, but also causing beneficial and often-disruptive change in how work is performed. The Block Center (mentioned above) launched a workforce supply chain initiative that utilizes AI and ML to better identify career pathways and specific training interventions that can enable workers to pursue new careers. By building models that examine the specific tasks undertaken by different occupations, the team was able to identify, for example, that radiologist technicians have a set of skills that, with highly targeted training, could enable them to be qualified to work in chip manufacturing – an area facing potential workforce shortages with impact on U.S. competiveness. This workforce supply chain initiative could provide metropolitan regions across the nation with a tool to enable local industry, government, and workforce leaders to partner with workers to adapt more rapidly and effectively meet the specific needs of their region.