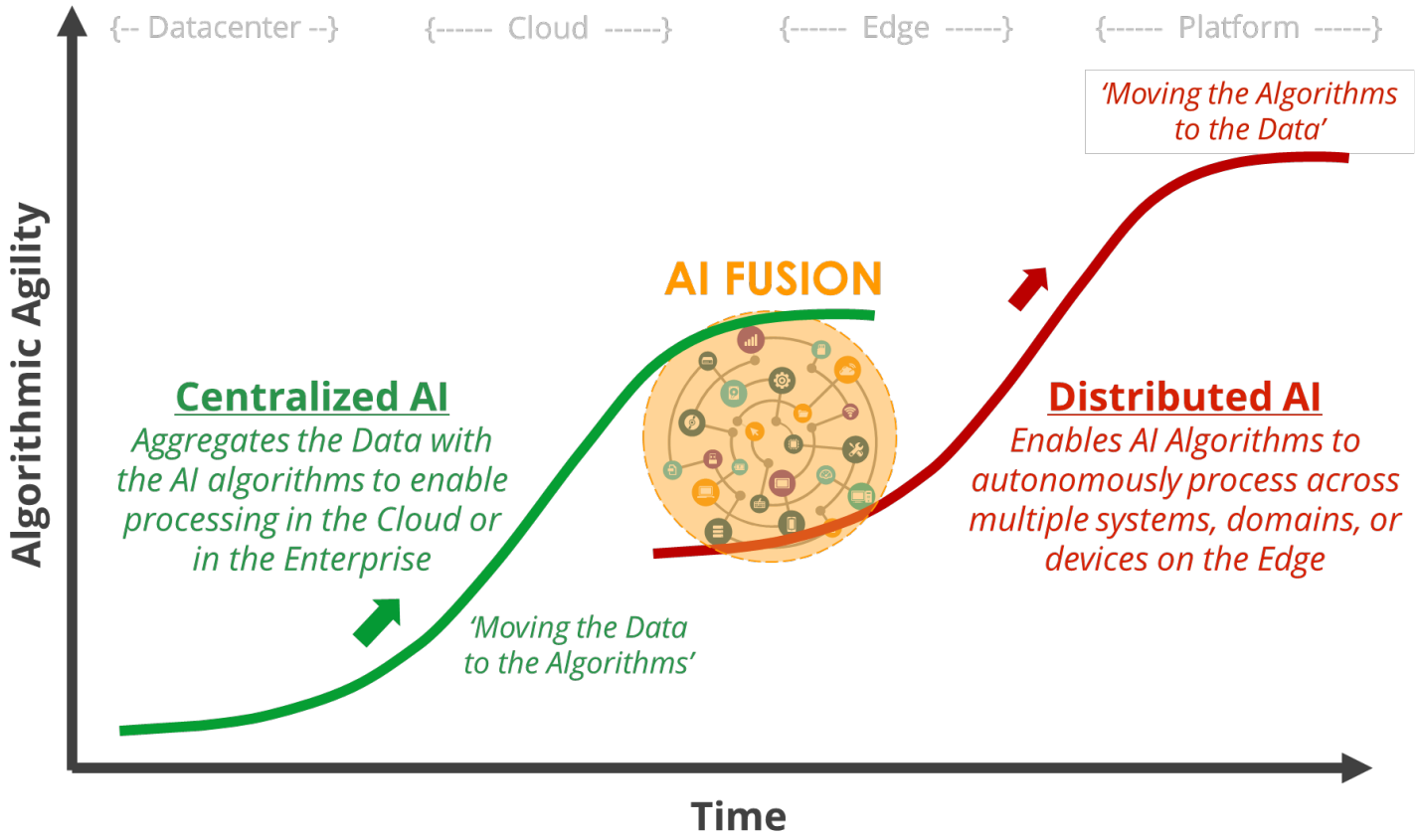
# AI Fusion: Enabling Distributed Artificial Intelligence at the Edge

The pace of innovation in Artificial Intelligence (AI) is completely unprecedented, and it continues to accelerate. With research and advances in key technologies such as machine learning, computational game theory, and autonomy, today's AI is capable of augmenting humans and increasing productivity and efficiency for critical tasks that just two years ago would have been impossible. Carnegie Mellon University has always been at the forefront of researching and enabling new advances and discoveries in computer science and emerging technologies, and our faculty and researchers continue to drive innovations and new breakthroughs in AI, redefining the art-of-the-possible each and every year.

## The AI Fusion Transformation:

Carnegie Mellon envisions a bold, new transformation occurring over the next 5-10 years that will enable AI to evolve from a highly structured and controlled, centralized architecture to a much more adaptive and pervasive, distributed architecture that autonomously fuses AI capability between the enterprise, the edge, and on-platform. We call this transformation AI Fusion. AI Fusion will minimize AI's dependency on aggregating and engineering massive data sets and the need to 'move data to the algorithms' in order for AI algorithms to be processed in the cloud or in a data center. Instead, AI Fusion will leverage algorithmic agility to enable autonomous data discovery and the ability for AI to have 'algorithms move to the data' and process at the edge or on-platform—drastically reducing the need for continuous, high-bandwidth connectivity to transport data and the potential sacrifice of the data's security and privacy in order to enable AI.





**FOR MORE INFORMATION:**

**Michael Mattarock II**

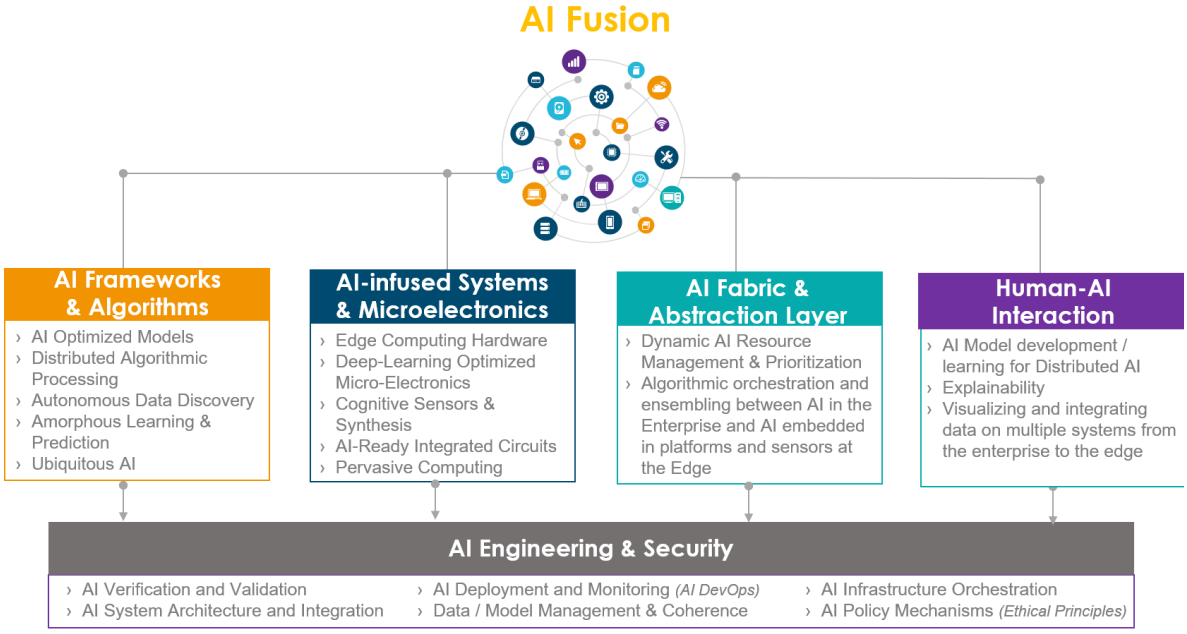
*Executive Director, National Security Research*

O: 412-268-5360 ■ C: 724-413-4731 ■ mattarock@cmu.edu


Carnegie Mellon University

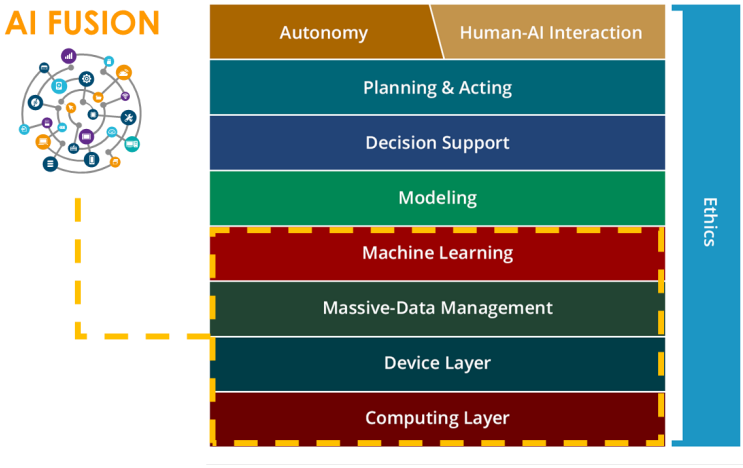
## AI Fusion Integrated Research Thrusts:

Precursors to AI Fusion are already being seen with recent advances in federated learning and microelectronics optimized for neural networks—but to truly unlock the potential of Distributed AI will require integrated research across five critical thrusts as part of a sustained Convergence Research activity:




1) ***AI Frameworks & Algorithms***—Algorithmic agility and distributed processing depend on the development of new AI algorithms that extend autonomous discovery and processing of disparate data beyond the current limits of federated learning, information theory, and meta learning. With these advances, the cloud will serve as an enabler for algorithmic mapping and orchestration between the enterprise and systems and devices at the edge.
2) ***AI-Infused Systems & AI Microelectronics***—Supporting dynamic, autonomous AI processing at the edge and on-platform will require extensive research into microelectronics architectures and connectivity beyond today's focus on 3D architectures and Systems on a Chip (SoC). More importantly, extensive research in the co-design of microelectronics with AI Algorithms & Frameworks and AI Fabric is needed to enable scalable training, inferencing, and prediction at the edge and to support algorithmic multi-threading on a single embedded chip or AI-infused system/sensor on-platform.
3) ***AI Fabric & Abstraction Layer***—The co-design of an AI Fabric is critical to facilitate distributed algorithmic processing and ensembling between the enterprise and the edge. Extensive research into novel mathematical frameworks, based on stochastic analysis and models of distributed systems, is necessary to ensure the performance, scheduling, resource allocation, and security of new AI algorithms—especially with the opportunistic communications associated with military operations in contested environments.
4) ***Human-AI Interaction***— Successful AI systems must be intuitive, usable, and improve the lives of their human users. Key aspects of Human-AI Interaction include: (1) model development and model learning; (2) explainability: the ability for the user to understand why the AI system has made the decision it has; and (3) visualizing and working with data.
5) ***AI Engineering & Security***—As AI applications and deployments exponentially increase, extensive research is needed to establish a new 'AI Engineering' discipline for developing resilient, reliable, and secure AI systems. Simply put, AI Engineering and Security brings 'confidence in capability'—knowing when AI systems are going to work and when to fix them—across the AI Fusion research thrusts, a task made more difficult as we embrace algorithmic agility, distributed processing, and fuse AI capabilities between the enterprise, the edge, and on-platform AI-infused systems operating across multiple domains.

## Transforming the Foundation of the AI Stack




In 2016, Carnegie Mellon created the AI Stack as a technical framework and blueprint to develop and deploy Artificial Intelligence. The premise of the AI Stack is simple—AI isn't just one thing. It's built from technology blocks that work together to enable AI. The AI Stack can also be thought of as a toolbox — each block houses a set of technologies that scientists and researchers can reach for as they work on new projects and initiatives. Each technology block depends on the support of the blocks beneath it, and enhances capability in the blocks above it. In a traditional, Centralized AI architecture, all of the technology blocks would be collocated or combined in the cloud or a single enclave to enable AI. For Distributed AI, AI Fusion will transform the very foundation of the AI Stack by enabling transformational advances in AI Theory, Frameworks, and Algorithms; AI Micro-Electronics and AI-Infused Systems; and an AI Fabric & Abstraction Layer that will fuse the distributed capabilities together to enable dynamic and autonomous AI processing on the edge or embedded on-platform.

Algorithmic agility and distributed processing will enable AI to perceive and learn in real-time by parallelizing these critical AI functions across multiple disparate systems, platforms, sensors, and devices operating at the edge.


Fusing AI capability from the Enterprise to the Edge, and enabling AI to be embedded On-Platform and support parallel and distributed algorithmic processing

Creating AI Frameworks & Algorithms that can autonomously discover and 'Move to the Data' - eliminating the need for massive data aggregation and engineering to enable AI

Driving the co-design and development of AI Microelectronics with AI Algorithms and an immersive AI Fabric that fuses them all together and re-defines the art-of-the-possible in AI

## AI Fusion - Convergence Research across AI and Cyber Physical Systems

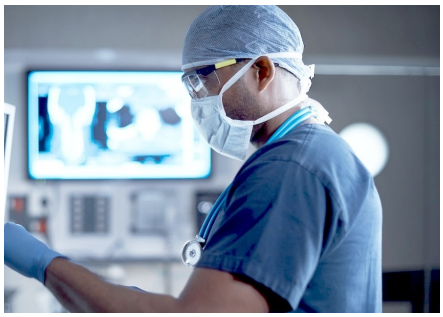
Achieving AI Fusion requires a convergence between the life sciences, physical sciences, computer sciences, and engineering to drive transformational research in AI and Cyber Physical Systems (CPS). A key goal of AI Fusion research is to develop the core system science needed to engineer complex, distributed cyber-physical systems with cognitive capabilities that people can interact with, benefit from, and depend upon across every aspect of their life. By abstracting from the particulars of a specific application or domain, AI Fusion seeks to reveal cross-cutting fundamental scientific and engineering principles that underpin the integration of AI with cyber and physical elements across all application sectors. There is also a convergence of AI Fusion technologies and research thrusts focused on Smart & Connected Communities, the Internet of Things., and Advanced Wireless Networks.

## Impact of AI Fusion and Distributed AI

The growth and evolution of AI has been accelerated due to on-going advances in compute and data storage / processing capability in the enterprise and the cloud—and these ***same*** technology drivers are fueling the transformation towards Distributed AI. AI Fusion will enable machine learning to become more scalable and amorphous between the Enterprise and the edge, and to autonomously discover and move to both known and unknown data sources and process in parallel across multiple, disparate platforms, systems, and sensors in real-time. This will enable unparalleled integration of AI across every sector and aspect of out lives to improve productivity, efficiency, safety, and the quality of life.

### Finance

Enabling AI to enhance accuracy, security, & privacy across banks & institutions



### Healthcare

Enabling AI to optimize societal health & well-being while protecting privacy

### Transportation

Enabling AI to increase safety and efficiency and reduce congestion & emissions

### Manufacturing

Enabling AI to improve production and optimize product design

### Energy

Enabling AI to increase resilience & reliability, and enhance efficiency

### Agriculture

Enabling AI to increase harvest yield & efficiency and the health of crops

### Defense

Enabling AI to enhance situational awareness & multi-domain operations

### Smart Cities

Enabling AI to improve public safety and optimize infrastructure & resources

### Environment

Enabling AI to protect air quality, reduce waste, & enhance sustainability

# AI Fusion Research Thrust Leads

**AI Frameworks & Algorithms:**

Dr. Virginia Smith

smithv@cmu.edu

website

**AI-Infused Systems & AI Microelectronics:**

Dr. Shawn Blanton

rblanton@andrew.cmu.edu

website

**AI Fabric & Abstraction Layer:**

Dr. Srinivasan Seshan

srini@cs.cmu.edu

website

**Human-AI Interaction:**

Dr. Jodi Forlizzi

forlizzi@cs.cmu.edu

website

**AI Engineering & Security:**

Dr. Matt Gaston

megaston@sei.cmu.edu

website