

Brief received

First, algorithmic innovation goes incredibly fast.

Second, we've reached a pivotal moment in Large Language Models. While adding compute power and data has been the go-to for Large Language Models, it's no longer the sole answer: more resources don't yield better results. Ilya Sutskever of OpenAI earlier referred to this as '**the plateau of LLM scaling**'.

Does this mean that the enormous AI-driven demand for compute power will slow down: not at all! We are convinced that it will just lead to an accelerated adoption of AI, resulting in a continued increasing demand for massive compute capability.

In addition, **next generation AI will become much more heterogeneous** across stages like training, post-training and inference. AI models will change continuously.

Today, we are only witnessing the start of these next waves of AI... **agentic AI** focusing on decision-making and **Physical AI** emphasizing embodiment and interaction with the physical world, will enable endless applications in health, automotive and robotics.

This will require the combination of multiple models, each serving a specific purpose and interacting with each other: an AI system that can combine large language models, perception models, and action models, gradually evolving to some form of reasoning.

This evolution of AI towards more advanced reasoning capabilities will undoubtedly continue to drive the need for greater computational power. In this context, while companies like DeepSeek aim to create efficient models that reduce computational requirements, the broader trend towards more complex AI functionalities may counterbalance these efficiency gains, sustaining or even increasing the overall demand for advanced computing infrastructure.

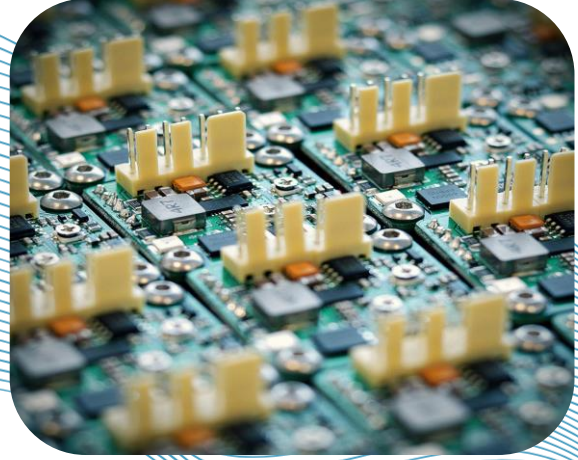
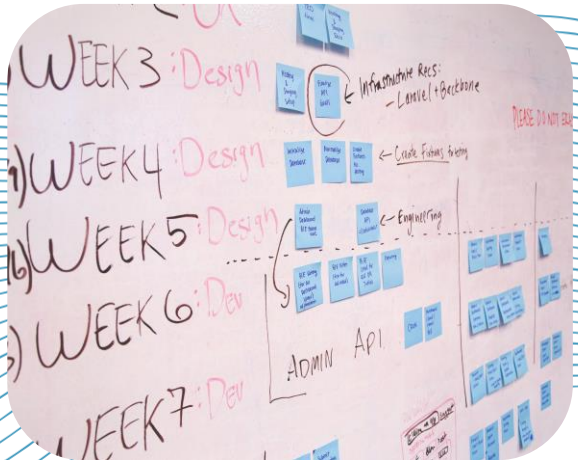


mec

From Scaling to Systems: The Next Phase of AI

By Raluca Dumitrache

AI Innovation Is Moving Exponentially



Algorithmic breakthroughs
accelerating

Faster iteration cycles

Innovation beyond
hardware scaling

Rapid advances in reasoning,
multimodality and efficiency

From months to weeks in model
experimentation

Performance gains increasingly
driven by software innovation

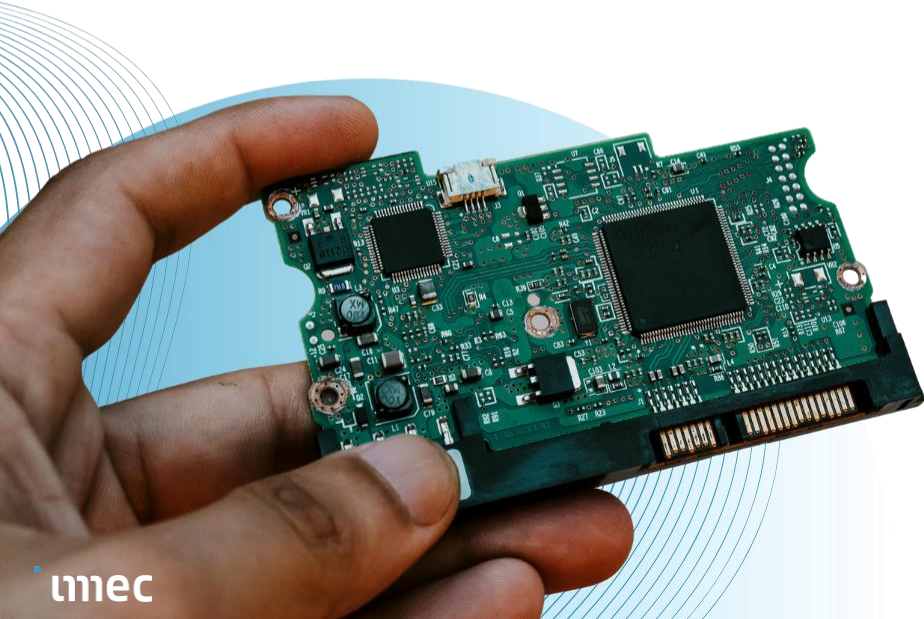
Scaling-driven Progress

- More data
- More compute
- Larger models
- Performance improved through scaling



Architecture-driven Progress

- More resources no longer guarantee better results
- Efficiency and specialization matter
- Multi-model systems emerging
- System-level optimization becomes critical





'the plateau of LLM scaling'

Ilya Sutskever of OpenAI

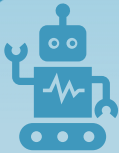
Does this slow compute demand?



No. Demand Is Accelerating

- 01 Accelerated AI adoption across sectors
- 02 AI becoming foundational infrastructure
- 03 Use cases expanding globally and across industries

Adoption Becomes Structural



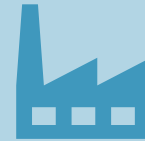
Enterprise Integration

- AI embedded in workflows
- Automation across functions
 - Productivity at scale



Consumer-Scale Deployment

- AI-native products
- Billions of end users
- Continuous usage



Industrial Transformation

- Automotive
- Healthcare
 - Robotics
- Smart manufacturing

From Models to Autonomous Systems

01

Agentic AI

- Autonomous decision-making
- Multi-step reasoning
- Autonomous task execution

02

Physical AI

- Embodiment
- Robotics
- Interaction with the physical world





Complexity Sustains Compute Demand

- 01 Advanced reasoning increases computational intensity
- 02 Multi-model systems raise system complexity
- 03 Expanding functionality offsets efficiency gains



mtec

embracing a better life