UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

AI for IoT

Tarek Abdelzaher, *University of Illinois at Urbana Champaign*

https://courses.grainger.illinois.edu/CS537/sp2026/AIoT/

Logistics

- **Instructor:** Tarek Abdelzaher (Office: SC 4126)
- Office Hours: Wednesdays 2-3pm on Zoom (Join Zoom Meeting), unless requested otherwise.
- Email: zaher@illinois.edu

A Little About Me

- Ph.D. in QoS Adaptation in Real-Time Systems, Department of Computer Science, University of Michigan, 1999.
- 1999-2005: Assistant Professor, Department of Computer Science, University of Virginia.
- 2005-now: Professor, Department of Computer Science, University of Illinois at Urbana Champaign
- Research Interests: Edge AI, Internet of Things (IoT), Embedded Systems, Real-time Computing, Cyber-physical Systems, Social Sensing

Where and When

Class Times

Tue/Thu, 2:00-3:15pm (SC 0216)

Grading

- ***Class Project:*** 60% of the grade is on a group project that produces a novel AI-based tool for IoT applications (done in groups of 3-4 students)
  - 5% midterm project presentation
  - 5% final project presentation
  - 50% on the final project report
- ***Paper Presentations:*** 10% of the grade is on paper presentations (done in groups of 3-4 students). Resubmissions allowed.
- ***Homework:*** 20% of the grade (8 short homeworks, mostly involving paper critique or AI-assisted search for information/evidence to support or refute a point of view).
  - Example 1: *From an energy-perspective is it more economical to train an AI model specific to the given IoT application, or leverage knowledge distillation from an all-purpose LLM? Give a 3-point argument for your answer.*
  - Example 2: Can foundation models for IoT learn reasoning skills without having a text-based language channel? Support your answer with arguments from prior work.
- ***Class Participation:*** 10% of the grade. Includes active contributions to class discussion.

AI Policy

AI (e.g., Gemini, ChatGPT, and similar LLMs and LLM-based tools) is a great resource that you are encouraged to use in order to be more productive. While the AI can help you find content, make slides, and get advice/feedback on ideas, you are ultimately responsible for the correctness of your (AI-assisted) content.

Homework Debates

You are allowed to use AI assistance in answering homework. However, most homework questions will accept multiple points of view. Your group will need to take one side or another. Next class, a debate will be organized between groups that chose different sides. Group members are expected to argue for their answer interactively during the debate. An informal exit poll will be conducted on which side argued more effectively.

The point: It is OK to use AI assistance in formulating your view, but ultimately you need to understand the topic enough to pick a side and be able to argue convincingly for it.

Project Timeline

- Group requests due 1/27. All groups will be formed by 1/29. (You will have the same group for homework and project.)
- Project title, abstract, and member list: Due February 5th
  - A periodic 30-min meeting scheduled (every other week) to discuss
- 2-page project proposal: Due February 26th
- Midterm project presentations (elevator talks): March 10th and 12th
- Final project presentations: April 28th and 30th
- Final project report due the week of finals. The report follows the standard technical paper format. Successful projects should result in a conference-quality paper.

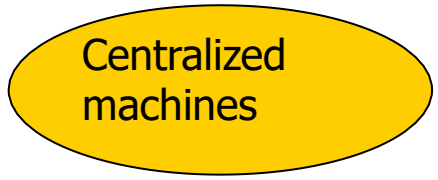
Schedule

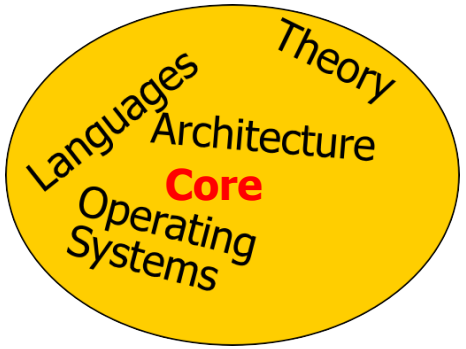
- See Website:

https://courses.grainger.illinois.edu/CS537/sp2026/AIoT/secure/schedule.html

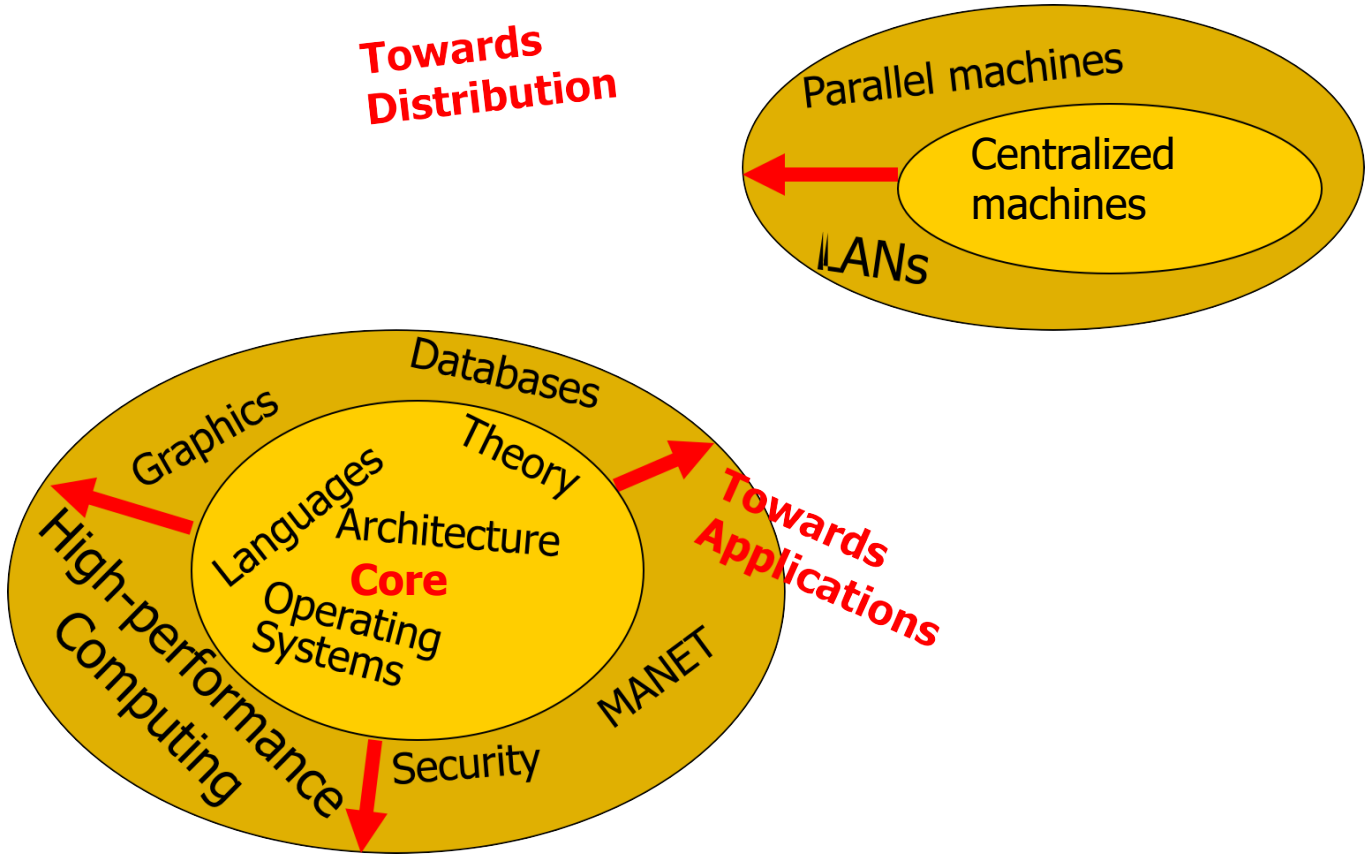
A History of Computer Science Research

The beginning:

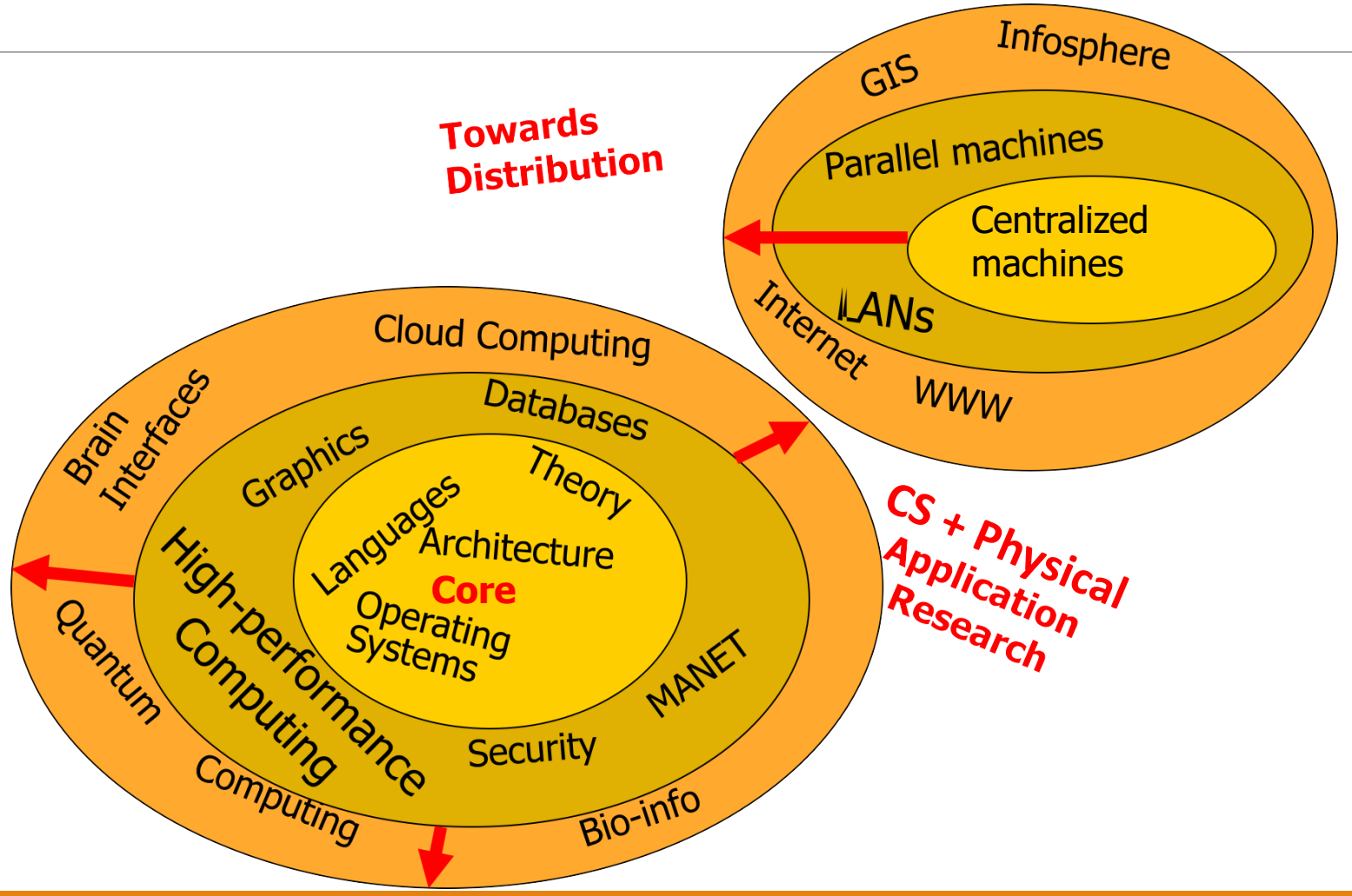




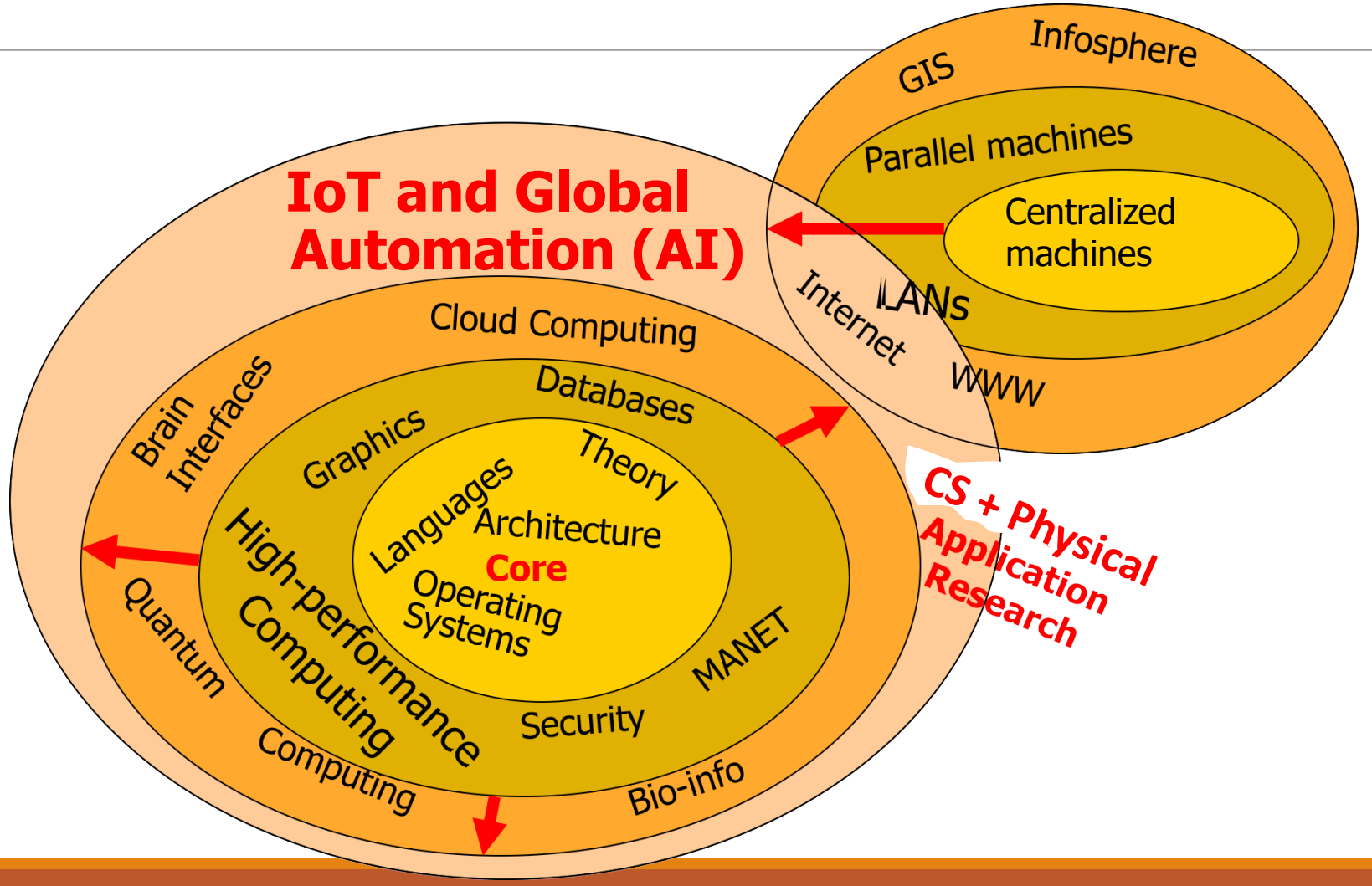
A History of Computer Science Research



A History of Computer Science Research



A History of Computer Science Research



The Context: A Cyber-Physical Perspective

AI is exacerbating the “digital bottlenecks” that computer science as a discipline needs to mitigate

The Context: A Cyber-Physical Perspective

AI is exacerbating the "digital bottlenecks" that computer science as a discipline needs to mitigate

Embedded/CPS and edge computing research, at its core, is about handling resource bottlenecks.

The Context: A Cyber-Physical Perspective

AI is exacerbating the "digital bottlenecks" that computer science as a discipline needs to mitigate

Embedded/CPS and edge computing research, at its core, is about handling resource bottlenecks.

The "AI trend" is *here to stay* for a long time…

The Context: A Cyber-Physical Perspective

AI is exacerbating the “digital bottlenecks” that computer science as a discipline needs to mitigate

Why is AI here to stay?

AI is merely the culmination of a much longer trend in science and engineering away from strong reliance on *algorithmic* and *symbolic* knowledge towards new *data-driven approaches*.

The Context: A Cyber-Physical Perspective

AI is exacerbating the "digital bottlenecks" that computer science as a discipline needs to mitigate

Why is AI here to stay?

AI is merely the culmination of a much longer trend in science and engineering away from strong reliance on *algorithmic* and *symbolic* knowledge towards new *data-driven approaches*.

The trend aims to remove the human cognitive capacity bottleneck (i.e., "symbolic understanding") from the critical path to innovation

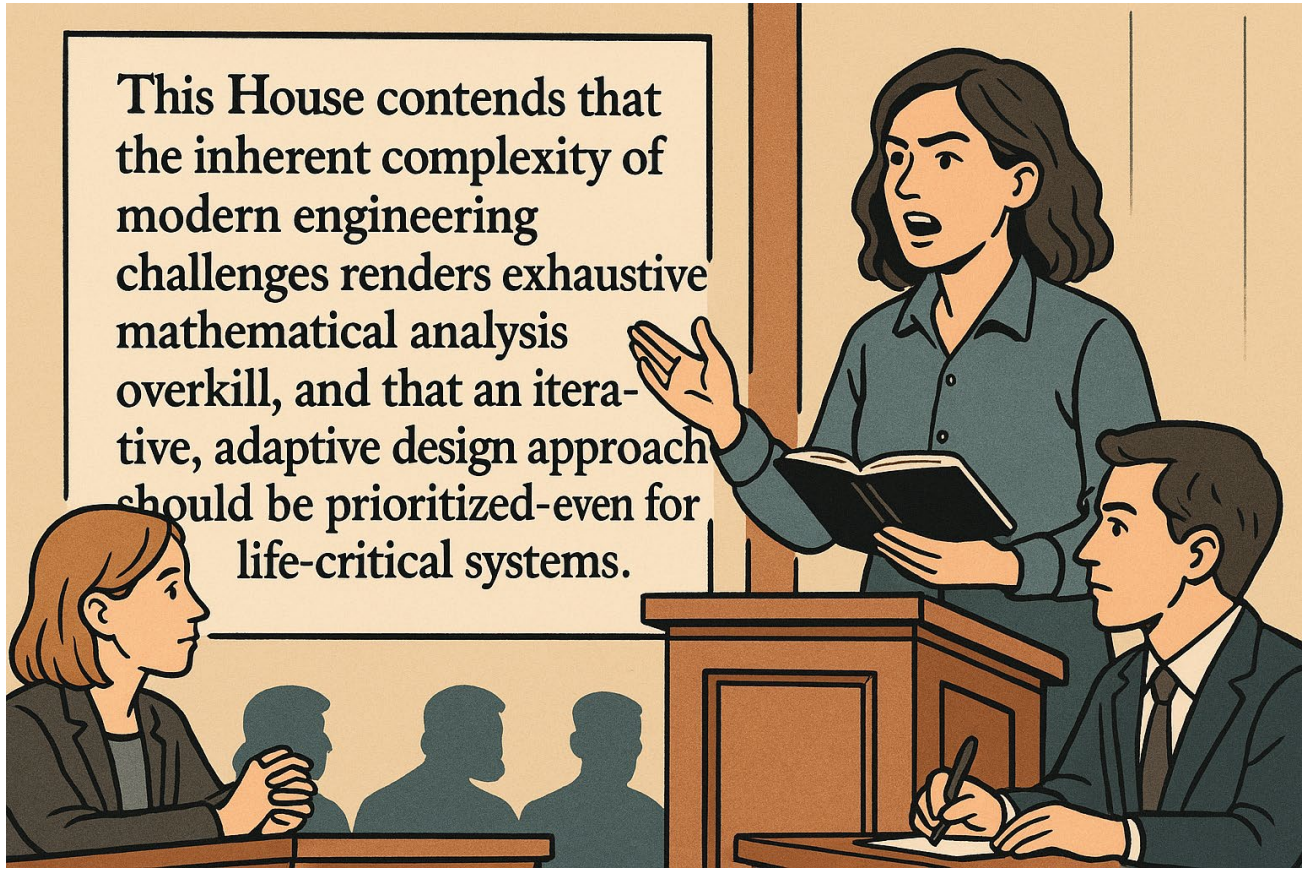
A Recent (Oxford-style) Debate Panel at CPS-IoT Week

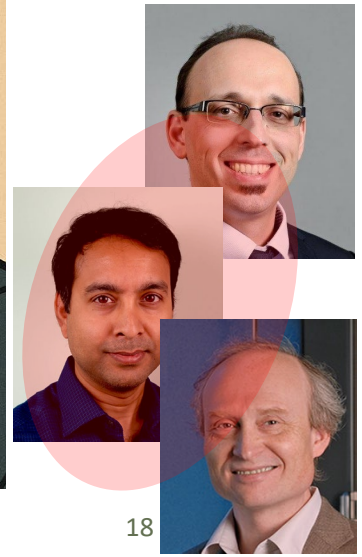
What should come first?

Empirical (i.e., data driven) approaches or theoretical understanding?

CPS-IoT Week 2025

May 6 - 9, 2025 Irvine, USA





Claim: We Can't "Build" and "Understand" Systems at the Same Time

Argument:

As computing capacity grows, it can be exploited to

#1: Build larger systems

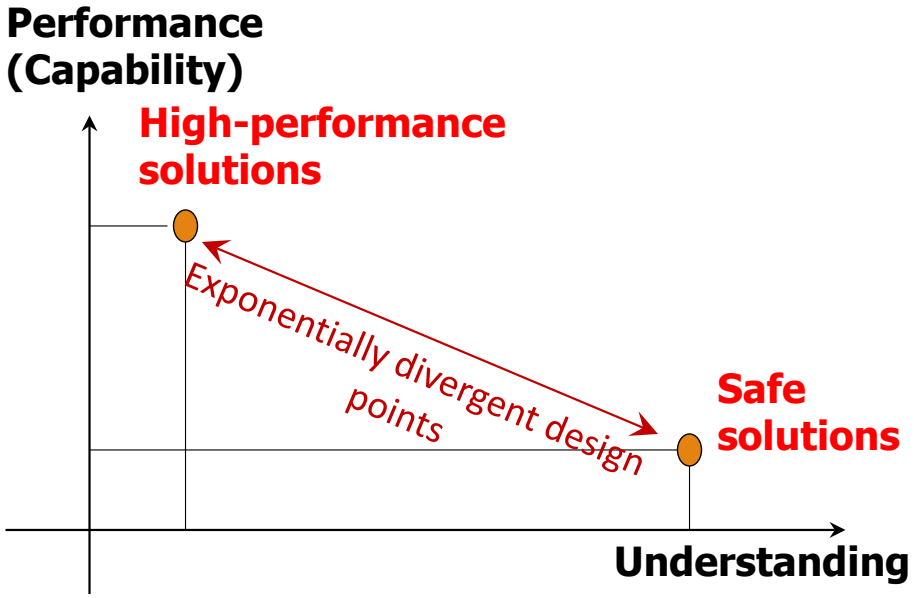
#2: Build better analysis tools to "understand/verify/prove-correct" said systems

Challenge: A linear growth in systems size (components, modules, subsystems), implies an exponential growth in interactive complexity (unless we significantly constrain it) and thus an exponential growth in resources needed to "fully understand" it.

System growth is *linear* in available capacity
"Understanding" demand is *exponential* in available capacity.

Understanding won't keep up with system capabilities as computing capacity grows!

Implication



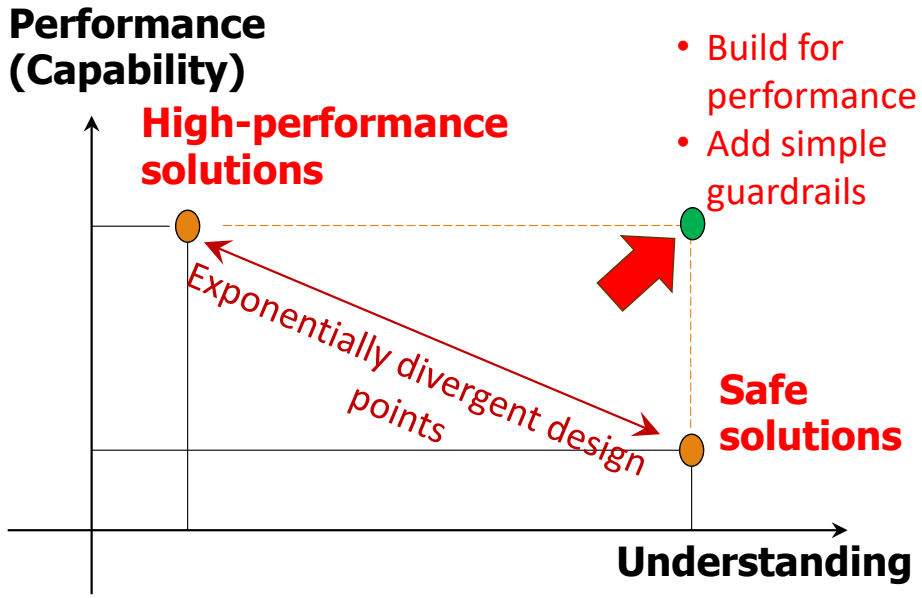
This visualization is credited to Prof. Lui Sha and his work on using simplicity to control complexity.

Sha, Lui. "Using simplicity to control complexity."
IEEE Software 18, no. 4 (2001): 20.

So, What Do We Do?
We Forge Ahead Anyway (With Guardrails)

Simple guardrail
(reduces error cost)

Complex, unpredictable, error-prone intelligence – learns from errors

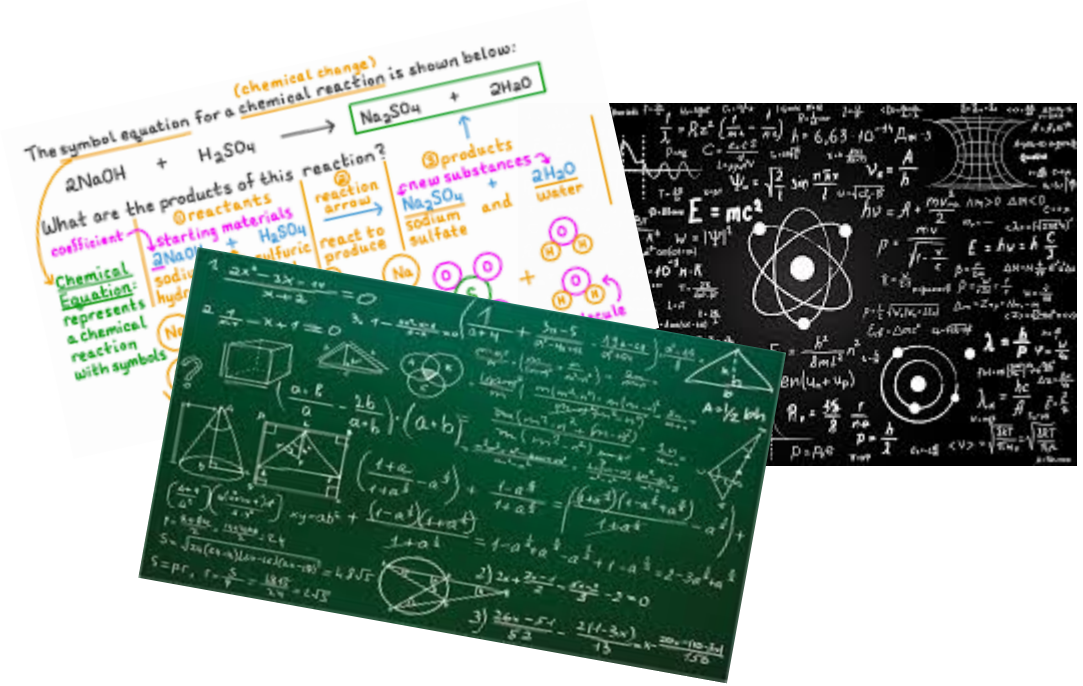


- Build for performance
- Add simple guardrails

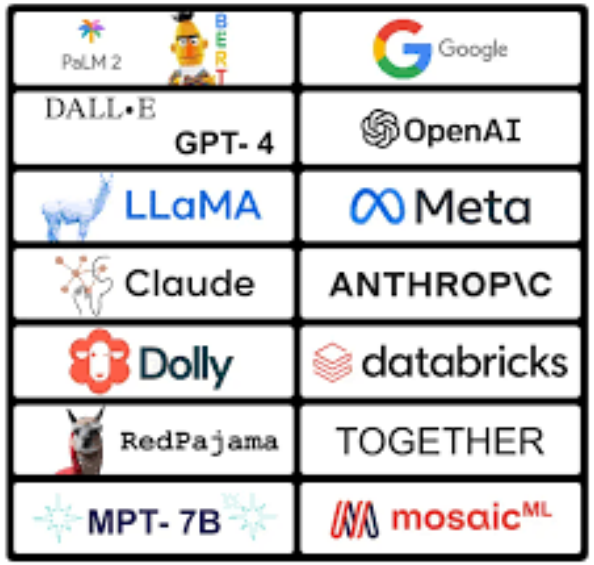
Guardrail examples:

- Collision avoidance overrides navigation
- Commander overrides/approves AI (military) action
- Local PID controller overrides/replaces unreachable remote model-predictive controller
- Digital twin fact-checks/overrides LLM hallucination

Implication: The Era of (Empirically Trained) Foundation Models



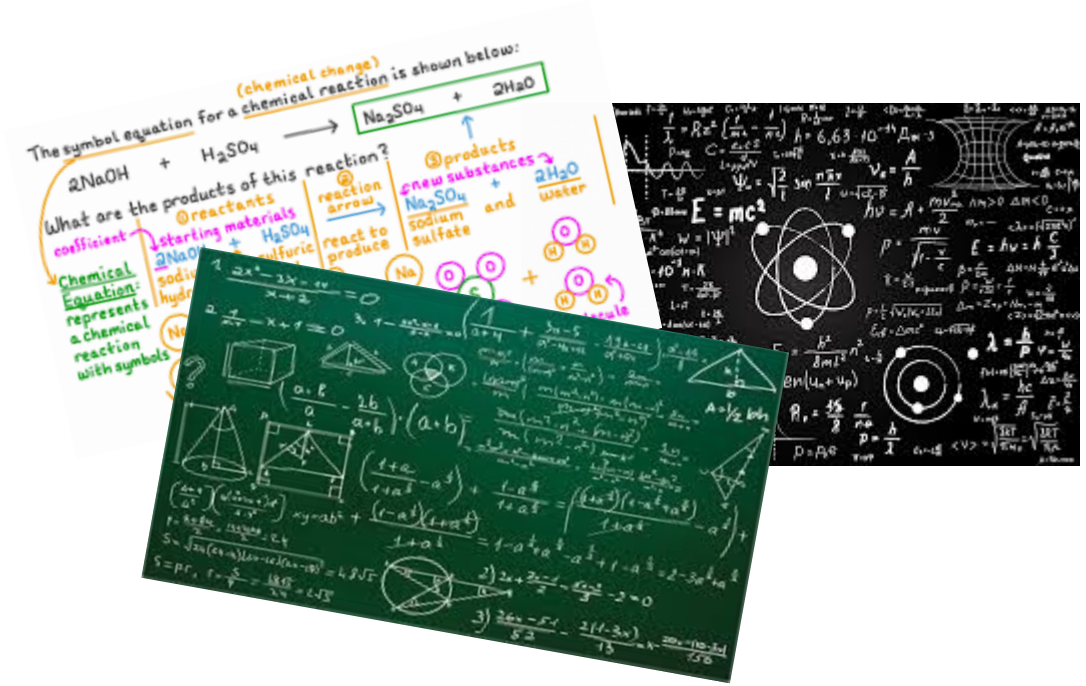
Conventional (scientific) foundation models



Tomorrow's data driven foundation models and LLMs

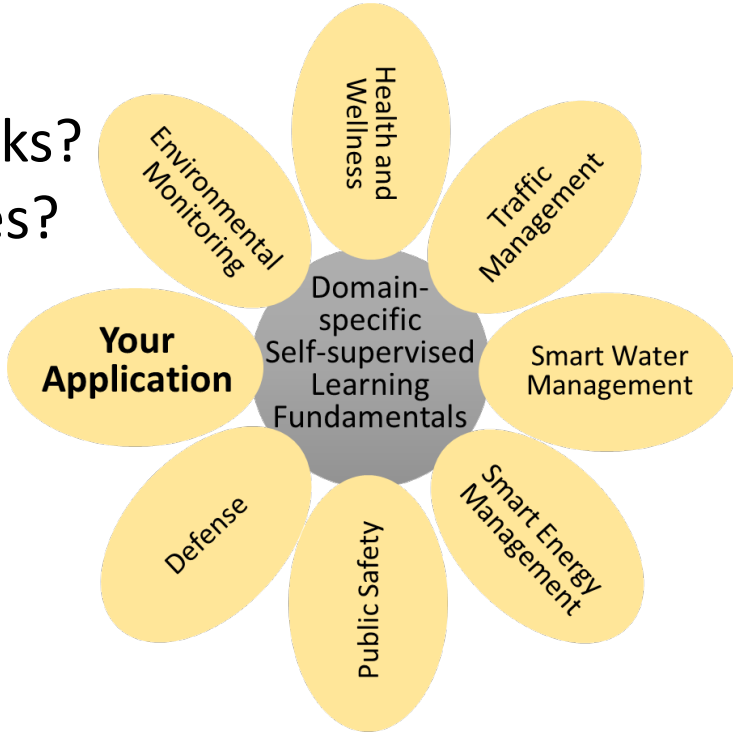
Implication: The Era of (Empirically Trained) *Domain-Specific* Foundation Models

How does it impact CPS/IoT applications?



Conventional (scientific) foundation models

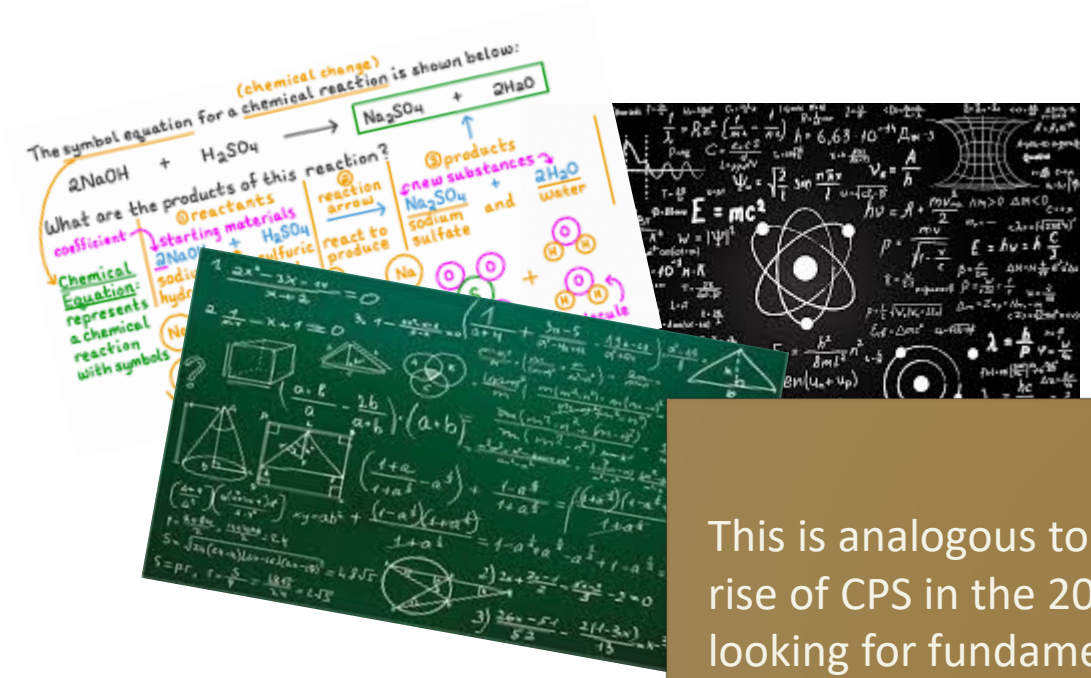
Bottlenecks?
Challenges?



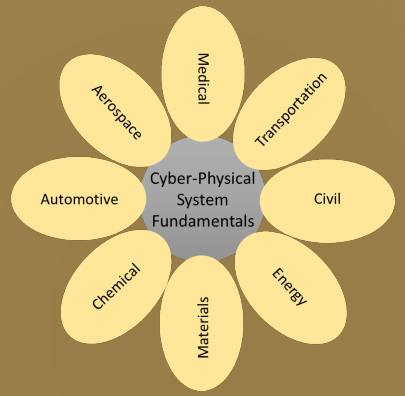
Tomorrow's data driven foundation models and LLMs (for CPS/IoT applications)

Implication: The Era of (Empirically Trained) *Domain-Specific* Foundation Models

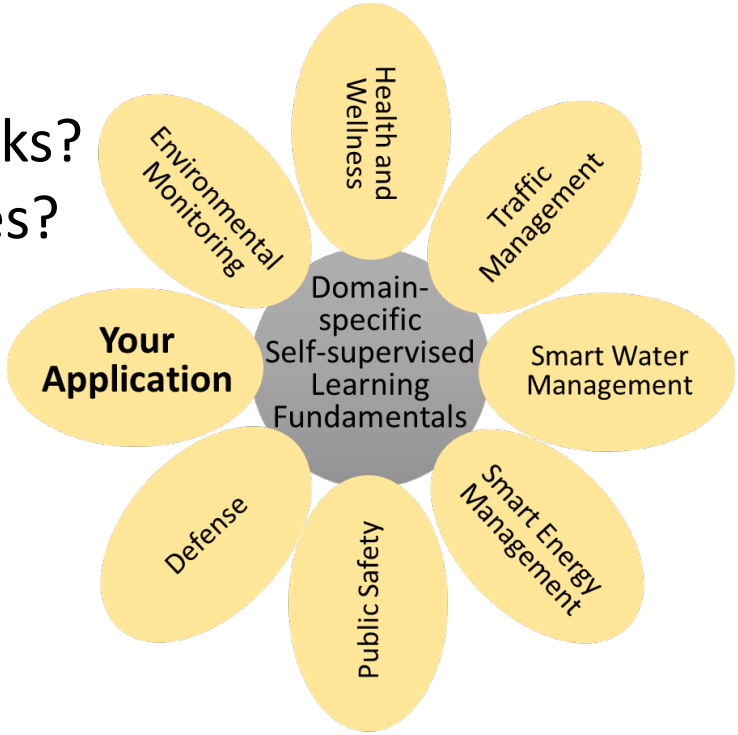
How does it impact CPS/IoT applications?



This is analogous to the rise of CPS in the 2000s, looking for fundamental tenets of the discipline.



Bottlenecks?
Challenges?



Tomorrow's data driven foundation models and LLMs (for CPS/IoT applications)