# Synthesis Comments

Ian Foster

Argonne National Laboratory

The University of Chicago

# We need to be clear as to why Acceleration and Automation are needed

- We heard some wonderful presentations on scientific goals, but few clear statements of **why** (and **when** and **where**), accelerating scientific discovery is important, and **how** automation is helpful.

- It might be useful to discuss in three categories:
  - <u>Faster</u>: To solve problems much **faster**
  - <u>Bigger</u>: To solve **larger** and/or **more complex** problems
  - <u>More</u>: To enable **many more people** to solve problems (think calculator vs. machine programming) – democratization
  - Also perhaps <u>repeatable</u> as a fourth dimensions

- I heard few strong statements of why revolutionary changes are needed: where we need **order-of-magnitude improvements**

# What are "automation" and "workflow"?

- Is "workflow" just fancy talk for computer programming?
  - Programming that happens to use a peculiar syntax and tools?
  - "Automated execution of command line programs"
  - "Teach people how to write Python"?
  - All good stuff, but that won't produce transformational advances
- Or is workflow what we say when we want to write programs that accelerate progress by orders of magnitude:
  - Automate things that are **currently being performed entirely or partially by a human** (and hence are slow, unreliable, inaccessible to non-experts)
  - Engage **resources other than computers**: e.g., sensors, scientific instruments ("program the world")
  - **Engage AI methods/agents** to replace need for human judgement

# Workflow tools

We heard about lots of workflow tools: >250 (surely an undercount). But few are used widely, and while e.g., Jupyter, ROs, etc., are very useful, it is not clear that they are changing scientific practice on a large scale. Many are viewed as having inadequate features and quality.

- My zero-order hypotheses:
  a. Current workflow tools are **not solving the real problem**(s), but rather incidental problems
  b. The diversity of tools reflects **perverse incentives**, rather than diversity in requirements
- I suggest that:
  a. We need to identify and work on the **real problems**
  b. We need to focus on **enabling infrastructure** rather than workflow tools

# The real problem: Enabling science to operate orders of magnitude faster

- I.e., **far faster** than can be achieved when humans perform any task manually

- Requires:
  - **AI agents** driving experiments and simulations
  - New **experimental apparatus** that can be controlled by computers and driven at high speed and large scale (low latency as well as high throughput)
  - New **computer systems** that can be controlled by computers to perform simulations, model training, model inference at high speed and large scale
  - Integrated **data and knowledge stores** to inform AI agents

- What technical obstacles must be overcome to realize this goal?

- What sociological, policy, and funding obstacles must be overcome?

# The need for a science platform

- Hypothesis: Bemoaning the lack of quality workflow tools and the need for sustainability misses the point
  - Allocating 2x more $$ or RSEs won't help much
  - We need to transform the economics of the field to enable positive returns to scale for platforms and tools that work

- A **science platform** is the technology that is ubiquitous, that we can take for granted, that provides the substrate on which more powerful tools can be built
  - E.g., Internet, GitHub
  - E.g., InCommon, Globus auth and data infrastructure

- What are the platform capabilities that we need to make ubiquitous so as to "raise all boats" in scientific discovery?