Physical AI in Space: Lessons from Formal Methods

Kristin Yvonne Rozier
Iowa State University

November 5, 2024

What Are Formal Methods?

"Formal Methods" are mathematically rigorous techniques for the specification, design, validation, and verification of software and hardware systems.

Intuitively, the system does what you think it should do
and nothing else.

Formal Methods "AI"

Formal Methods	"AI"
• Work in absence of data	• Require lots of data

Formal Methods	"AI"
• Work in absence of data • Can pinpoint over-generalizations and unstated assumptions	• Require lots of data • Can propagate over-generalizations and hide assumptions

Formal Methods

- Work in absence of data
- Can pinpoint over-generalizations and unstated assumptions
- Based on mathematical logic, proceeds in proof steps

"AI"

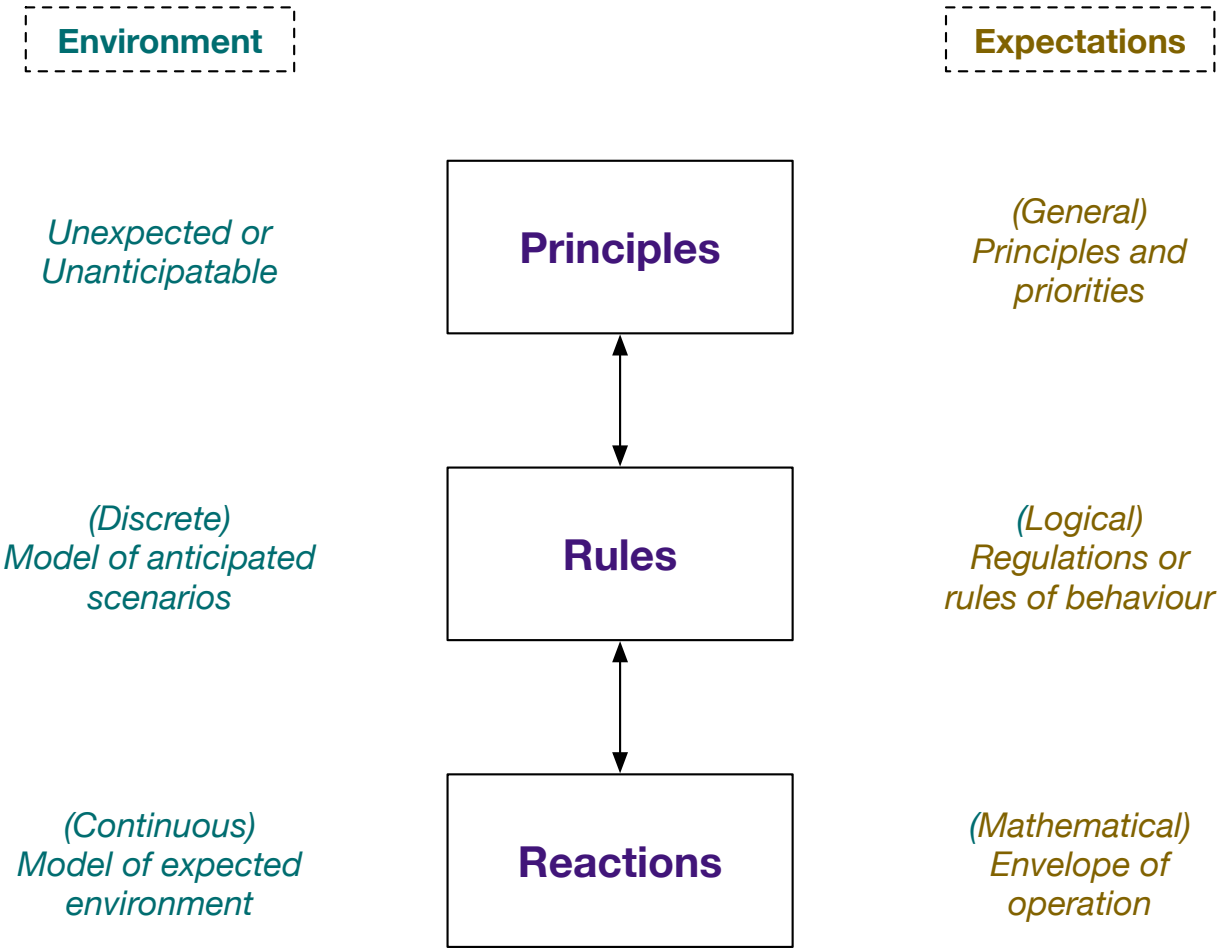
- Require lots of data
- Can propagate over-generalizations and hide assumptions
- Can jump to conclusions not supported by axioms&inference

Levels of Autonomy [1]

- **No autonomy:** Human responsible for all required tasks
- **Low autonomy:** Straightforward (but non-trivial) tasks done entirely autonomously (no human poised to take over operation)
- **Assistance systems:** Human assisted by automated systems, remains in control or must be ready to take back control at any time
- **Partial autonomy:** System operates autonomously; human remains engaged, monitors the operation, and intervenes immediately
- **Conditional autonomy:** Automation in full control of specified tasks; human must still be prepared to intervene upon request
- **High autonomy:** Automation performs all planned functions under certain circumstances; humans can control others
- **Full autonomy:** Automation can perform all its intended tasks on its own; no human intervention required at any time

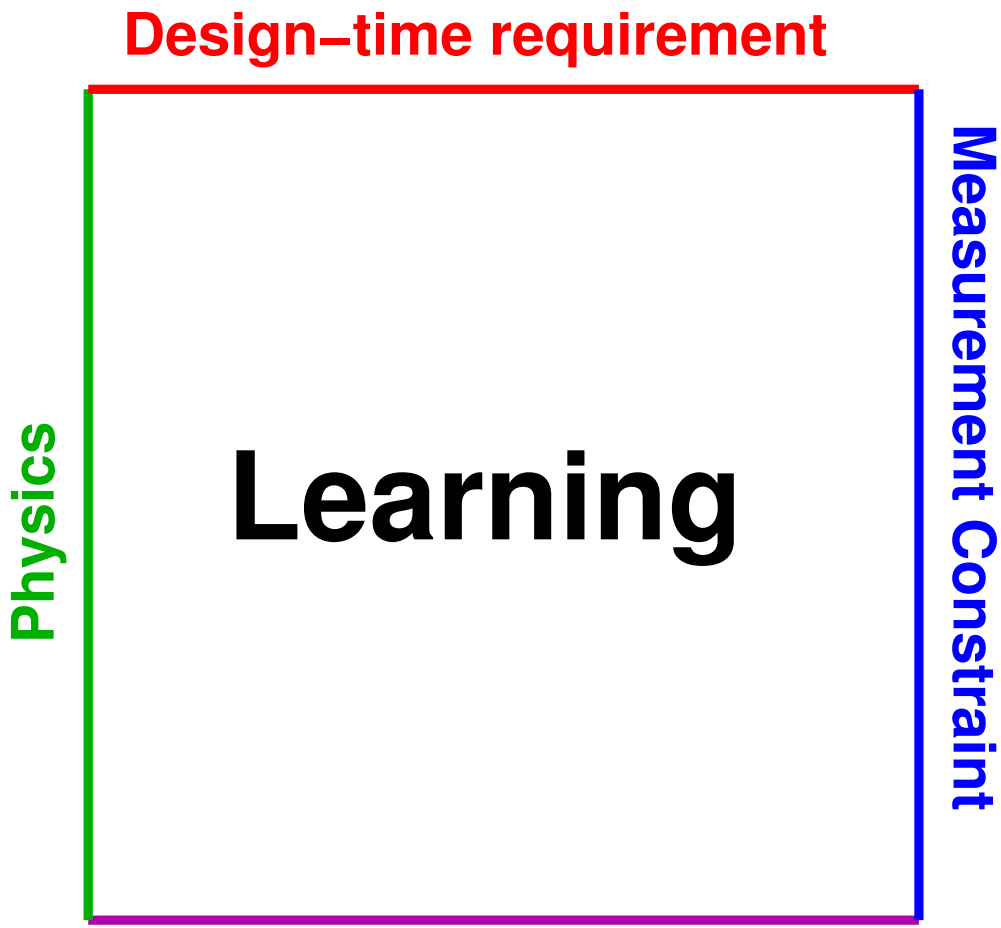
[1] Michael Fisher, Viviana Mascardi, Kristin Yvonne Rozier, Bernd-Holger Schlingloff, Michael Winikoff, Neil Yorke-Smith. "Towards a Framework for Certification of Reliable Autonomous Systems."

A Three-layer Autonomy Framework [2]



[2] Michael Fisher, Viviana Mascardi, Kristin Yvonne Rozier, Bernd-Holger Schlingloff, Michael Winikoff, Neil Yorke-Smith. "Towards a Framework for Certification of Reliable Autonomous Systems."

Learning in a Safety Region

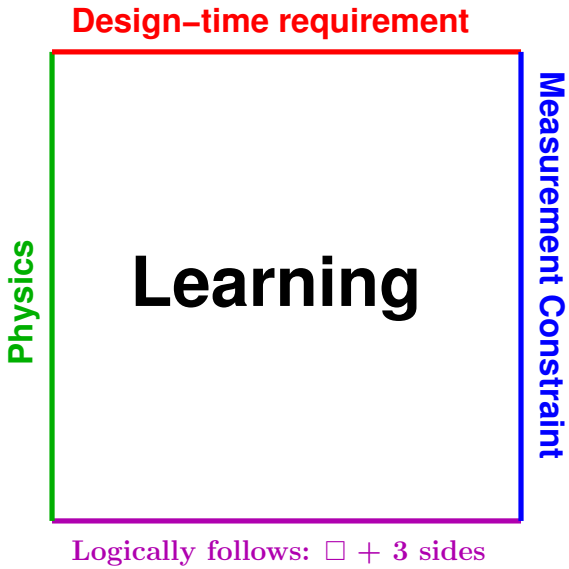


Logically follows: $\square + 3$ sides

Safety Bounds

- Can use logical deduction (e.g., bound by SAT/SMT)
- Can use a priori known bounds (e.g., bounded learning)

- Can we use design-time requirements?
- Can we use technological limits?
  - what we can measure
  - computational complexity
  - what we can verify



Formal Methods can provide bounds for AI

Patience Is a Virtue

(1) A Bike Trail in Iowa

Adding Complexity: A Bike Trail in California

- goat heads
- different obstacles/interruptions
- different terrain/potholes
- ...

Rigorously Building A Library

Before autonomous driving, try **cleaning**:

- bike trails
- sidewalks
- bike lanes
- shoulders
- snow
- parking lots

Lessons from Formal Methods for Physical AI in Space

- Cognizance of **levels of autonomy** and **layers of autonomy**
- Ways to **formally bound learning**
- Building up "trustworthy" autonomy **incrementally, like proofs**