AI for Discovery:
A Preliminary Glimpse

Dong-Shan Jian（Peking University）

jiandongshan@hotmail.com

Quantum Computing and Machine Learning Workshop 2025, Qingdao, 2025/08/22

Based on: Y.L.Fang, D.S.Jian, X.Li, Y.Q. Ma, arXiv: 2504.01538

Outline

I. Background and motivation

II. AI-Newton: rediscovering classical theories

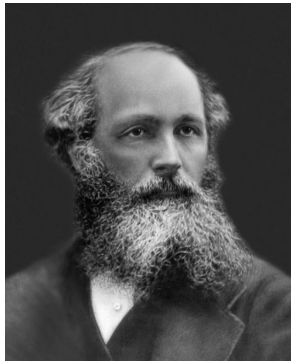
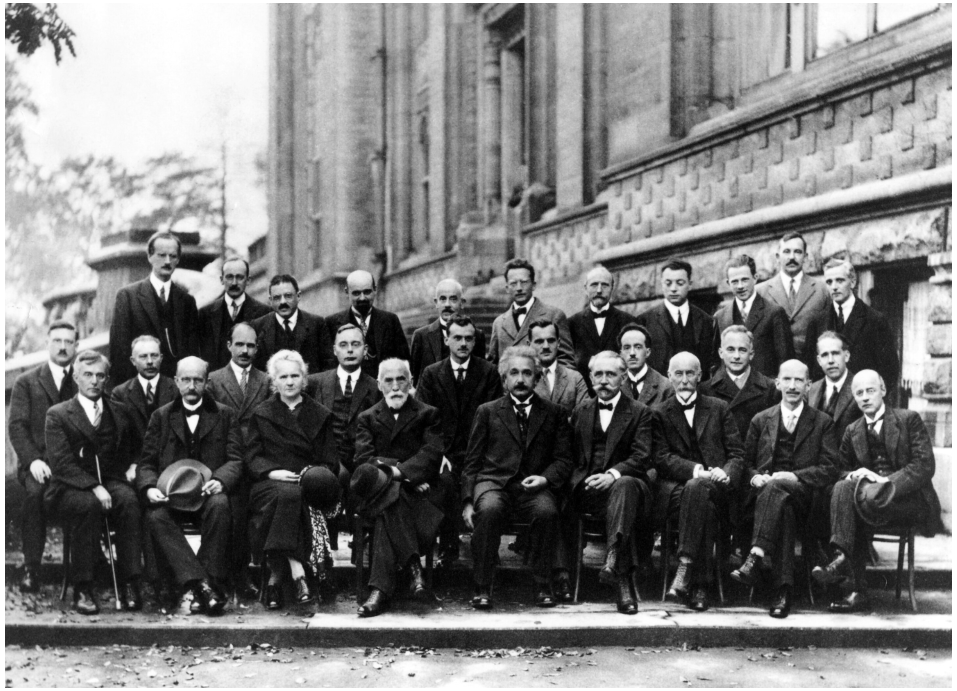
III. Challenges of quantum theories

IV. Summary and outlook

Human scientific discovery

➢ **Fundamental natural laws: human contributions**



Galileo's laws of motion

Newton's laws of motion

Maxwell's electromagnetic theory

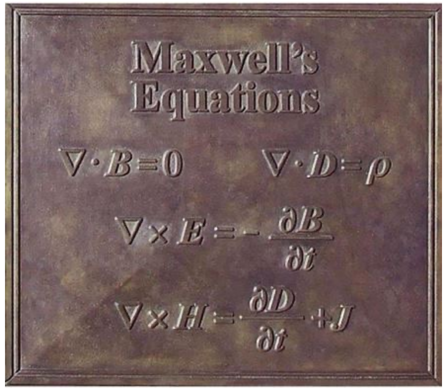
Theory of relativity

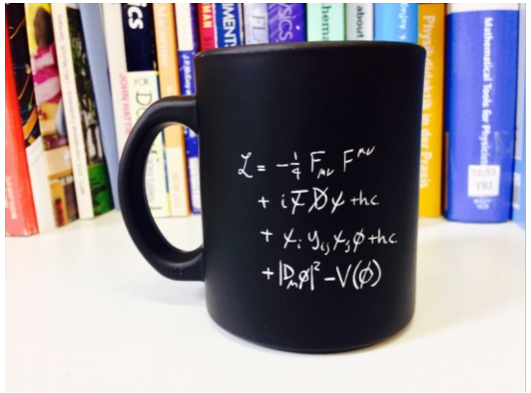
Quantum theory

Reflection

➢ Human exploration of natural laws:

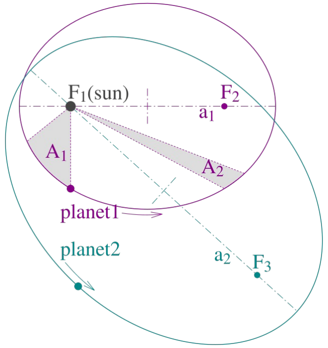
- Advantages: interpretability, conciseness, universality

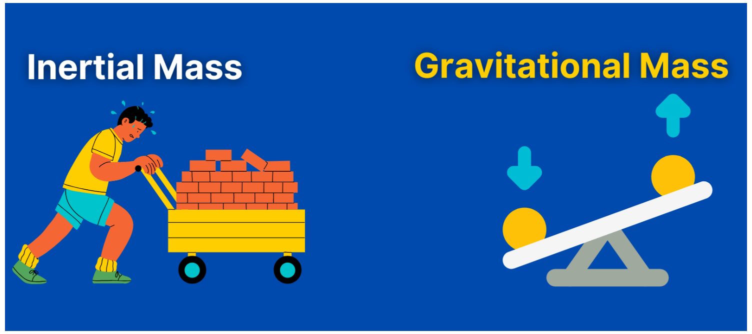


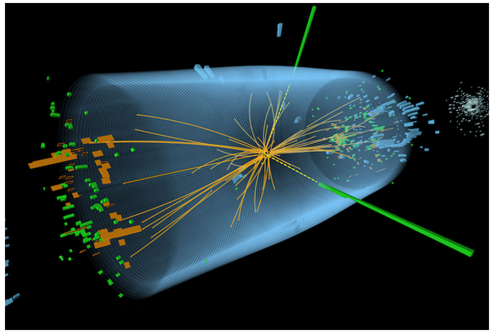


- Disadvantages: long period, preconceived notion,
  insufficient ability to handle complex problems

New paradigm?

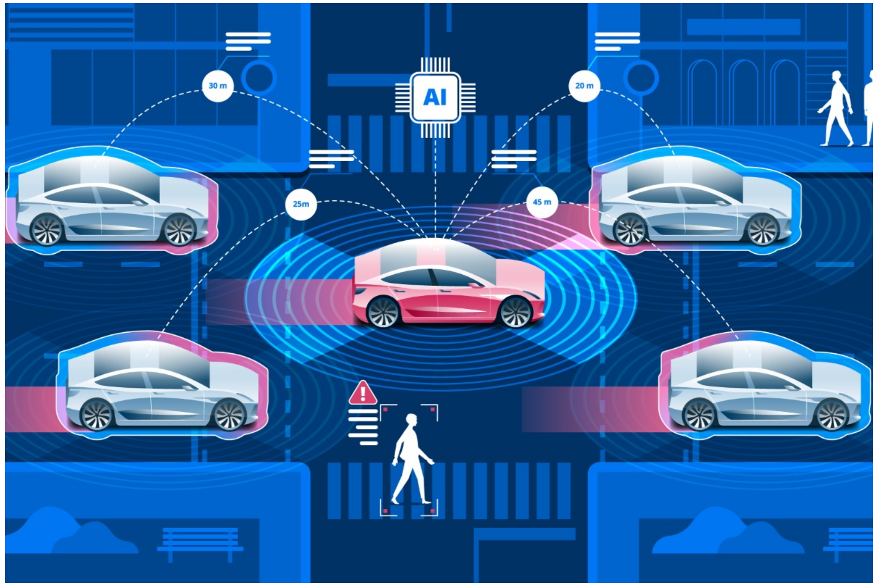






AI-driven scientific discovery

➢ The power of artificial intelligence (AI):



➢ AI-driven scientific discovery:

Focusing on leveraging AI to assist in addressing specific cutting-edge issues

Reddy and Shojaee, 2412.11427

… integrated AI systems capable of performing autonomous long-term scientific research and discovery.

➡ **Continuous and autonomous scientific discovery?**
General theories?

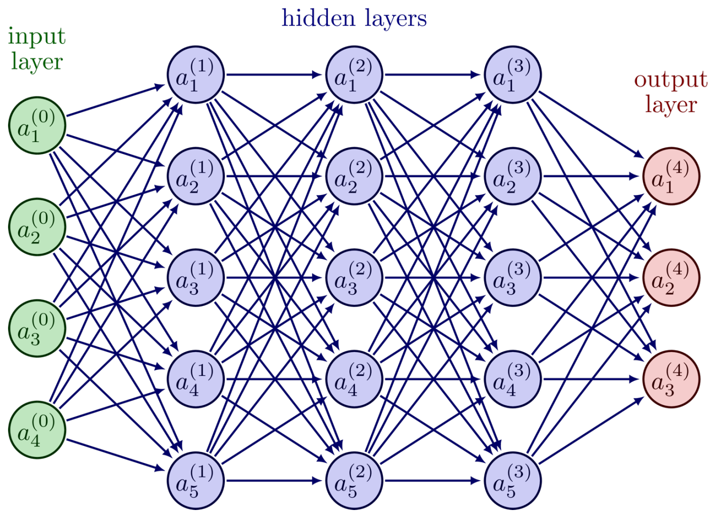
Still an open issue!

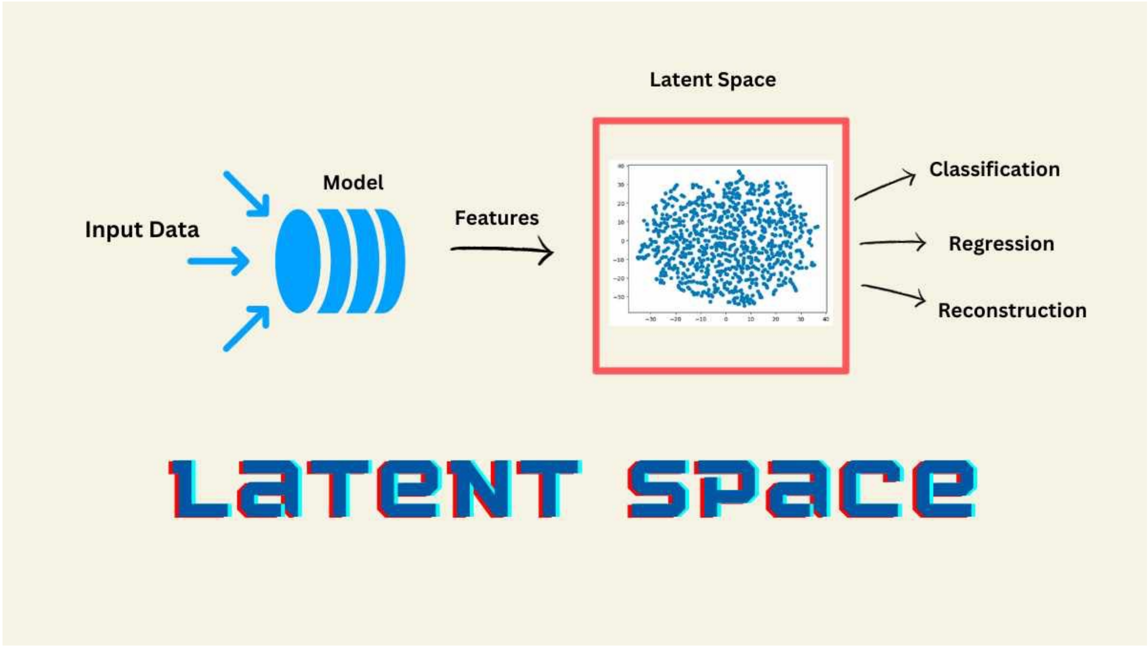
Limitations in current methods

➢ NN-based methods:

- Lack of interpretability (black-box)
- Insufficient extrapolation capability (while critical for science)

NN ≠ AI !!!

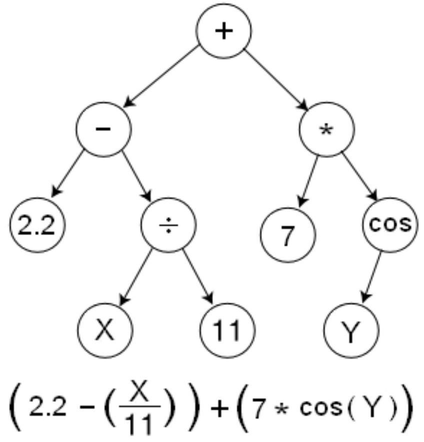


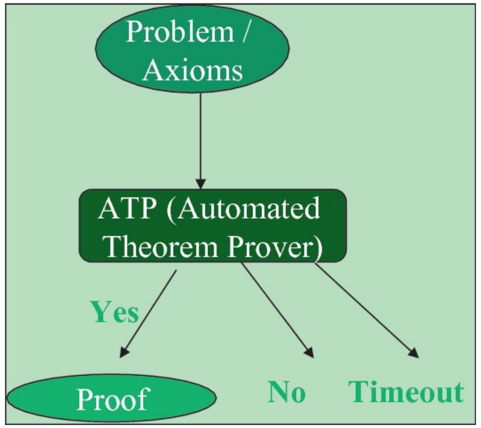


Limitations in current methods

➢ Symbolic methods:

- Limited expressive capability
- Search space explosion
- Limited cross-problem transferability





Core challenges

➢ How to represent and manage physical knowledge?

- Knowledge representation: important both for physics and AI
- Hierarchically structured and integrating multifaceted information
- Mainly functions involving experiments, physical objects, space-time coordinates, etc.
- Far beyond mere mathematical formula or end-to-end NNs
- Extensively researched in mathematics (e.g. Lean), but limited in physics

➢ How to effectively search for physical knowledge?

- Search space explosion and brute-force search is impractical in practice
- Specific ⟶ General ? Driving factor ?

Outline

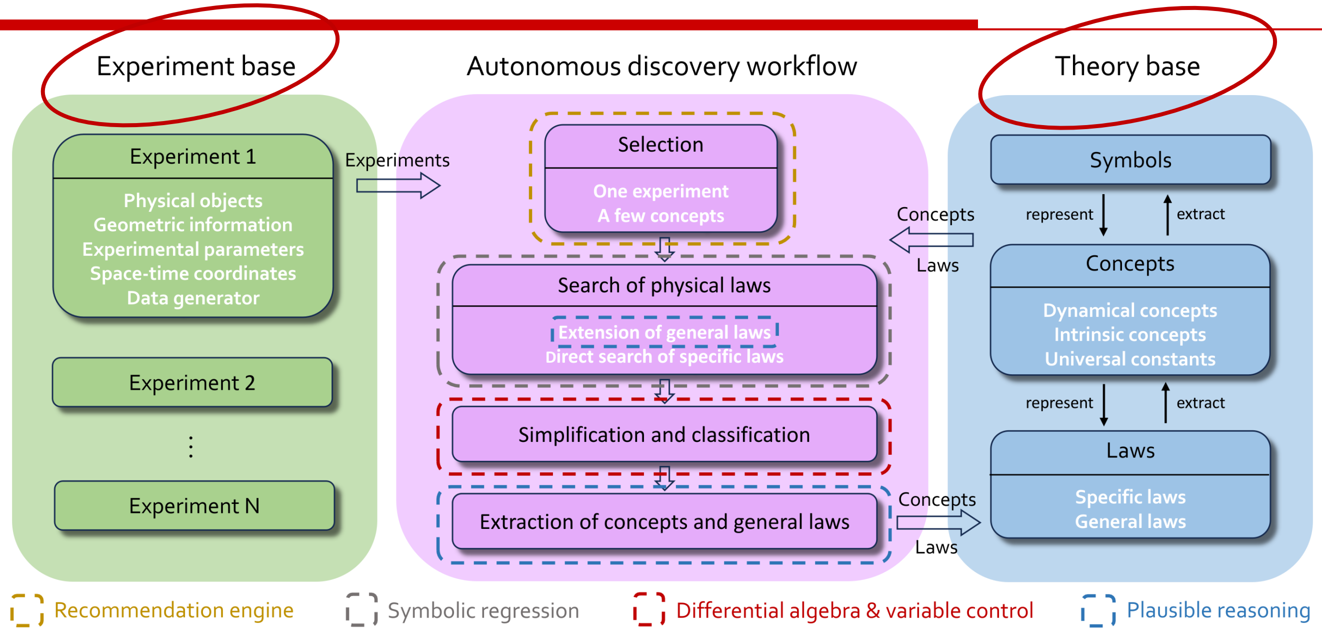
I. Background and motivation

II. AI-Newton: rediscovering classical theories

III. Challenges of quantum theories

IV. Summary and outlook

Knowledge base



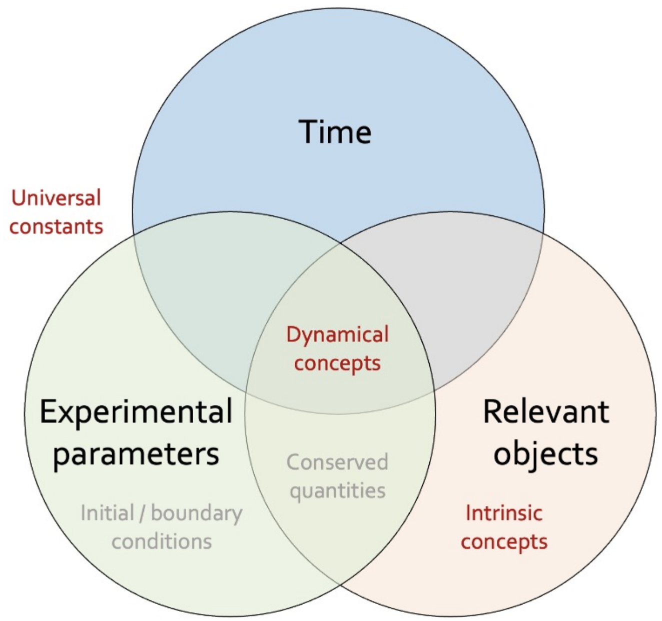
Fang, DSJ, Li, Ma, 2504.01538

➢ Experiment base:

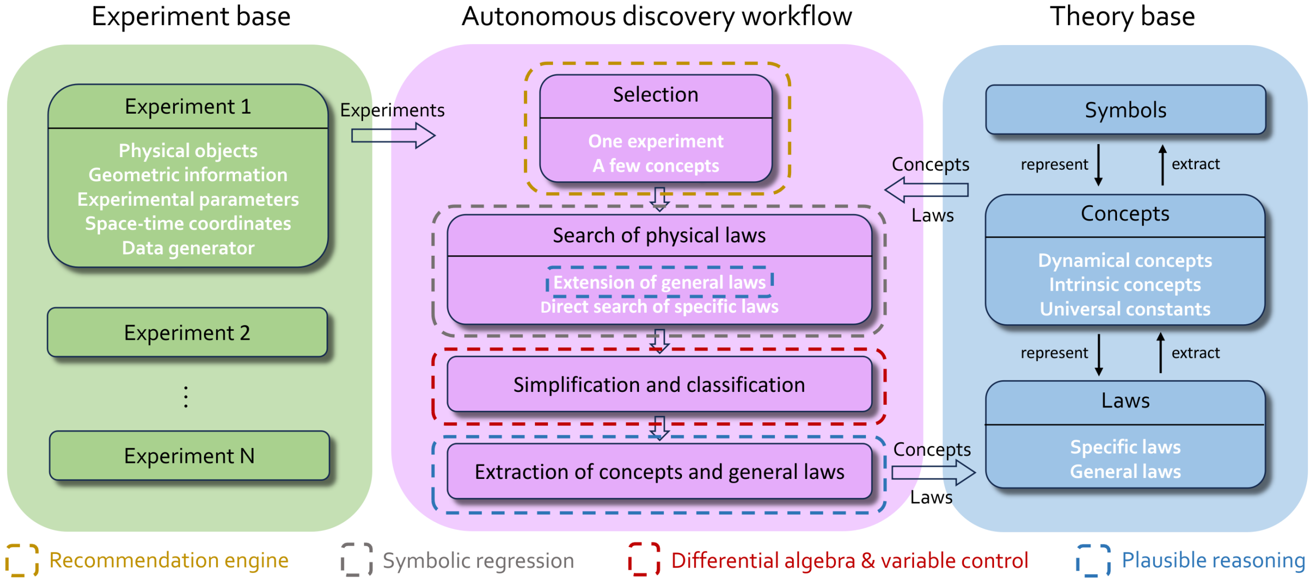
- Numerical simulation ⟹ Data with noise
- Time series of coordinates (no any other concept !!!)

➢ Theory base:

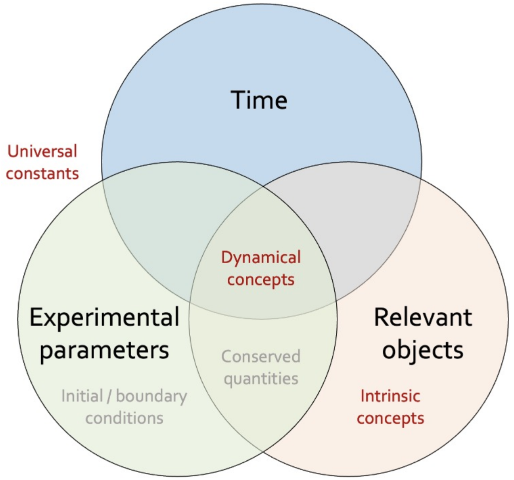
- 3 – layer structure
- Centered at concepts
- Classification based on parameters dependencies



Knowledge representation



Fang, DSJ, Li, Ma, 2504.01538



- **Functions** whose inputs involve experiments, physical objects, space-time coordinates, …

  ⟹ A **physical DSL** for representation / manipulation

  considering generality and self-consistency

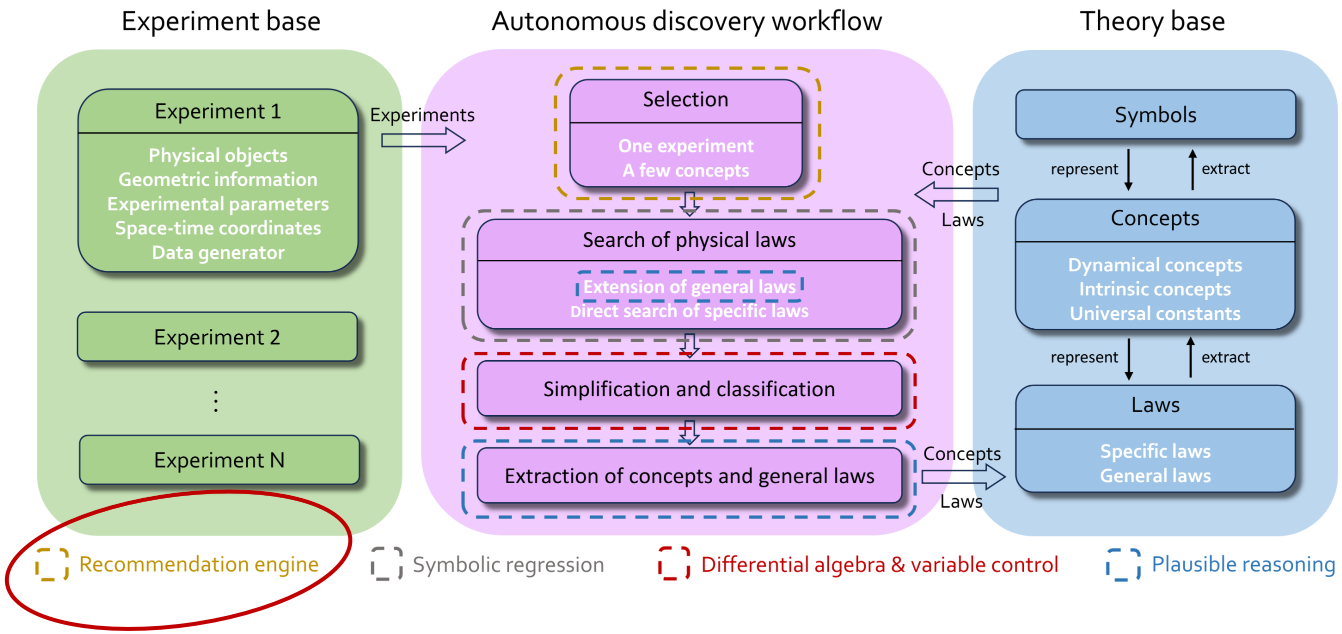
- Far beyond mere formula / NNs, especially when handling multiple problems, e.g.:
  - Intrinsic concepts, such as mass, numerically depend solely on specific physical objects

    ⟹ Recording their measurements is essential

$$C_{02} := \forall i\text{: Ball, Intrinsic}[\text{ExpName}(o_1 \to \boxed{i}\ o_2 \to s), \mathrm{L}[s] - \mathrm{L}_0[s]]$$

Appropriate knowledge representation is the cornerstone of autonomous discovery!

Recommendation engine



Fang, DSJ, Li, Ma, 2504.01538

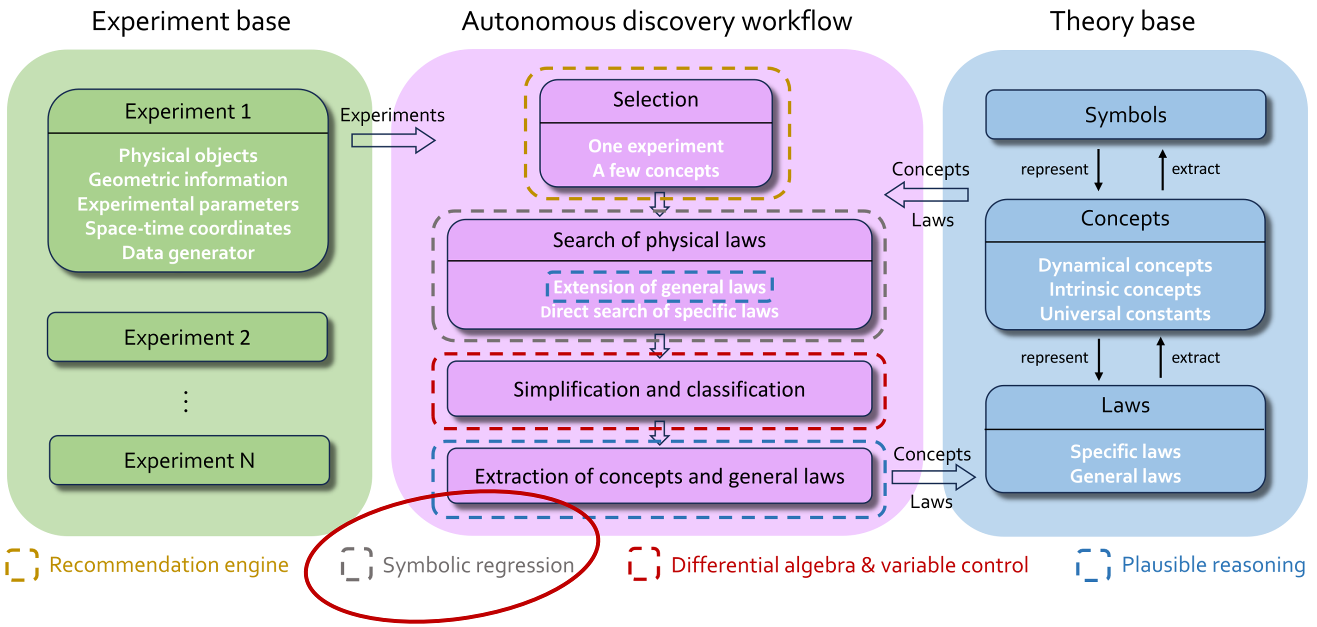
- **Balance exploitation and exploration:** $$V(k) = \alpha R(k) + \sqrt{\frac{1}{1+N(k)}},$$

  ⟹ **Value function (inspired by UCB) + dynamically adapted NN (online learning)**

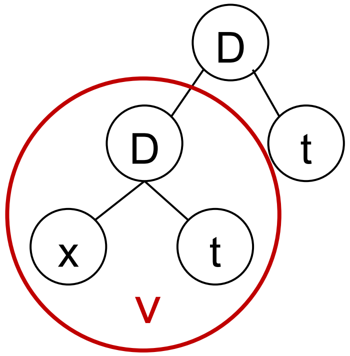
- **Prevent the workflow from grappling with complex experiments too early**

  ⟹ **The era-control strategy**

Symbolic Regression



Fang, DSJ, Li, Ma, 2504.01538



➢ **Optimization objective:**

- **Traditional regression: parameters**
- **Symbolic regression: <u>function forms</u> + parameters**

➢ **Search space explosion**

⇨ **Many other strategies are employed to address this issue, e.g., introduction of <u>new concepts</u> and <u>general laws</u>.**

(general law)

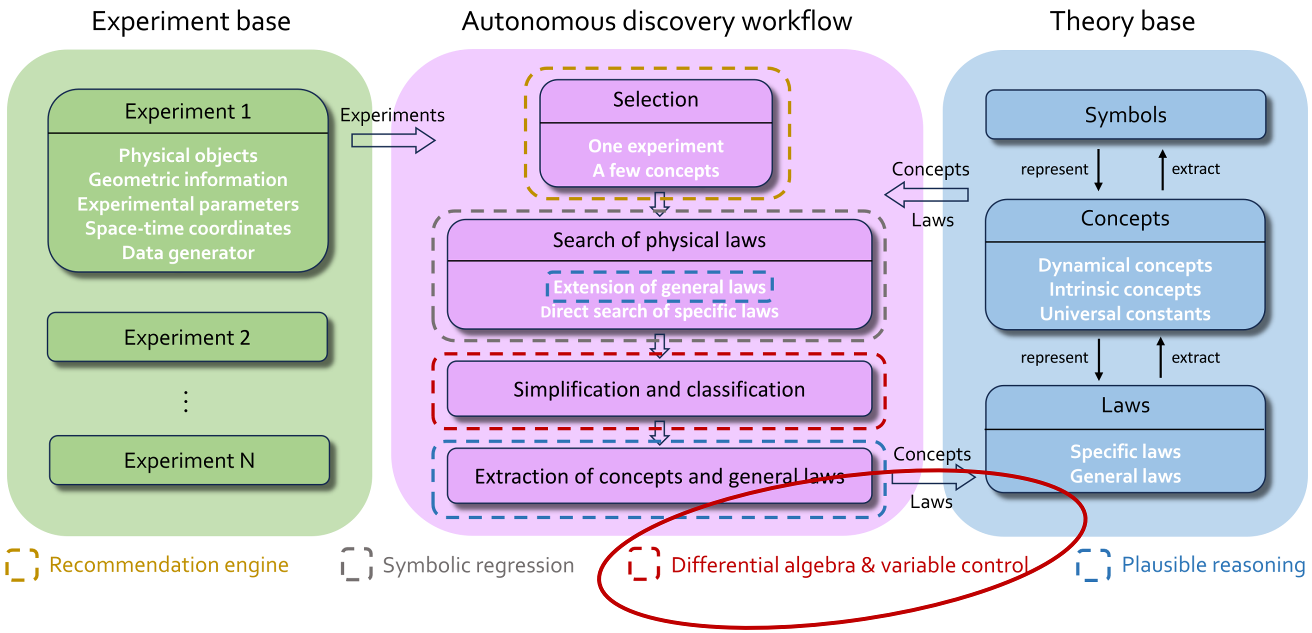
$$\forall i : \text{Ball},\ m_i a_{i,x} + (\nabla_i V_k)_x + (\nabla_i V_g)_x = 0,$$

Specialized for a ball on an inclined plane connected to a fixed end via a spring

$$m a_x - \frac{c_x c_z}{c_x^2 + c_y^2 + c_z^2} mg + \frac{\left(\left(c_y^2 + c_z^2\right) x - c_x \left(c_y y + c_z z\right)\right)}{\left(c_x^2 + c_y^2 + c_z^2\right) L} k \Delta L = 0,$$

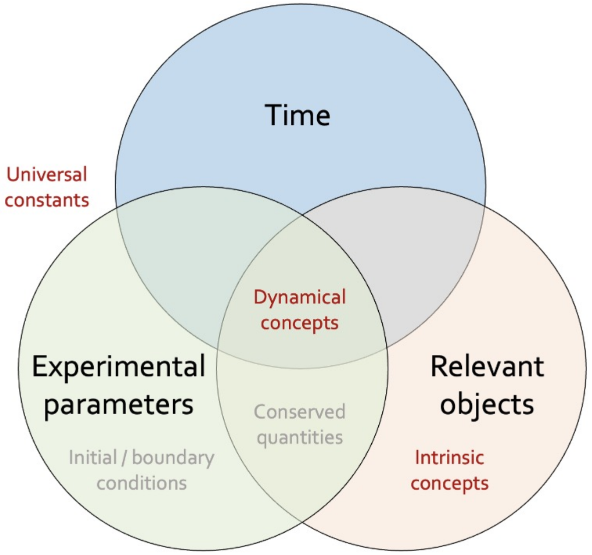
(specific law)

Differential algebra & variable control



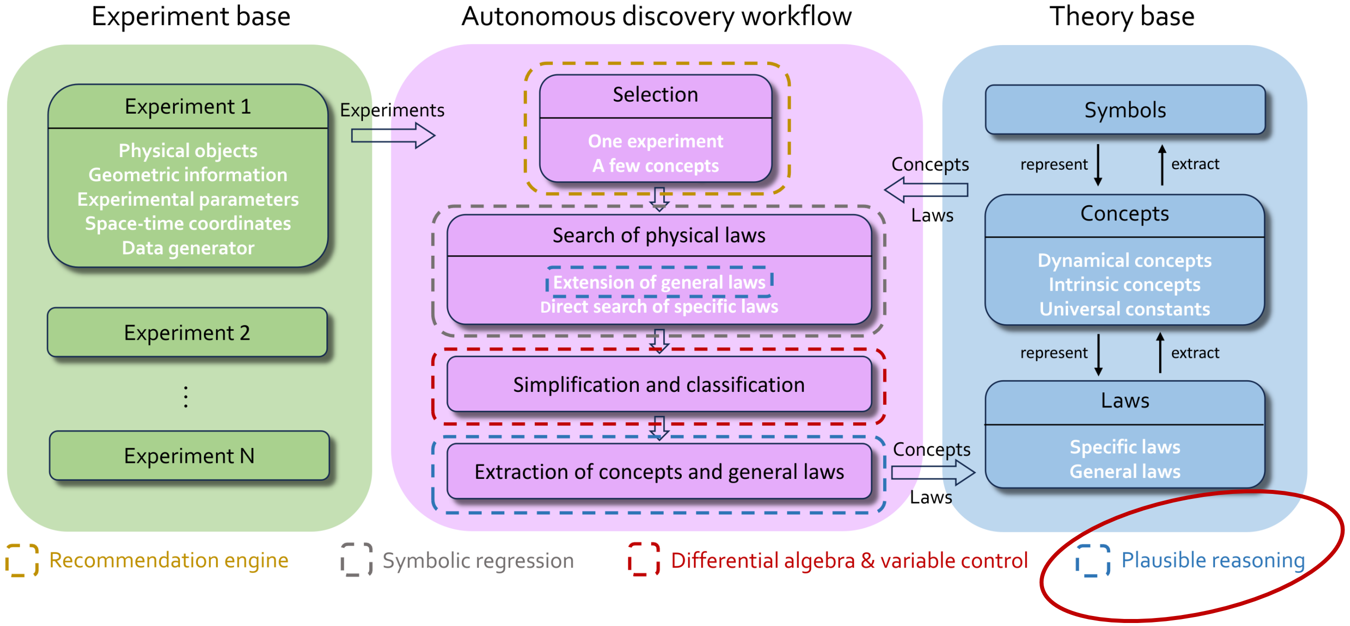
Fang, DSJ, Li, Ma, 2504.01538

- **Differential algebra:**
  - **Rosenfeld Groebner algorithm**
  - **Simplification (reduction of redundant knowledge)**
- **Variable control:**
  - **Classification based on parameter dependencies**



Plausible reasoning

Fang, DSJ, Li, Ma, 2504.01538



- Based on rational inference from partial evidence
- Main functions：
  - **Extracts physical concepts**, e.g.:

    (in uniform linear motion) $\mathrm{d}x[1]/\mathrm{d}t = \text{const.}$ $\xrightarrow{\text{broader utility?}}$ $C_{01} := \forall i\text{: Ball, } \mathrm{d}x[i]/\mathrm{d}t,$ (velocity) $\xrightarrow{\text{try out}}$ ✅

  - **Proposes and extends general laws**, e.g.:

    (in elastic collision) $T = \text{const.}$ $\xrightarrow[\text{Capable of extension?}]{\text{Valids in others?}}$ (in spring systems) $T + V_k = \text{const.}$ (elastic potential) $\Longrightarrow \cdots$

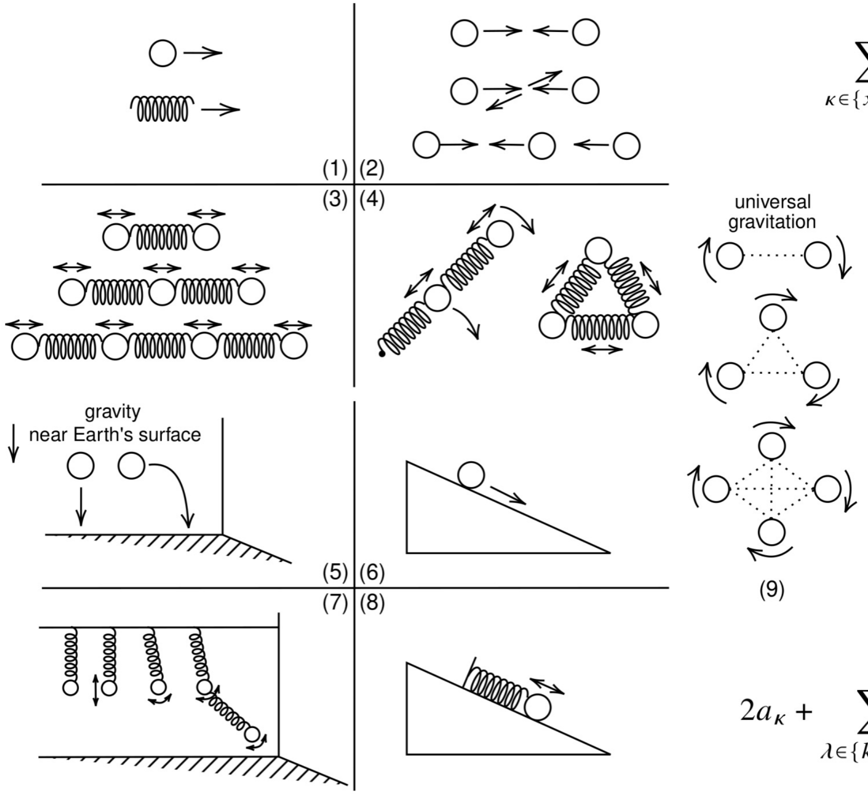
Tests and results

- Based on noisy data, important natural laws are discovered!
- Unsupervised! Without prior physical knowledge!

Fang, DSJ, Li, Ma, 2504.01538

Physical objects

Schematic of experiments



Discovered important general laws

Energy conservation

$$\sum_{\kappa\in\{x,y,z\}} T_\kappa + \sum_{\lambda\in\{k,g,G\}} \delta_\lambda V_\lambda = \text{const.},$$

where T_κ and V_λ are defined as:

$$T_\kappa = \sum_{i\in\text{Particles}} m_i v_{i,\kappa}^2,$$

$$V_k = \sum_{i\in\text{Springs}} k_i (L_i - L_{0,i})^2,$$

$$V_g = \sum_{i\in\text{Particles}} 2 m_i g z_i,$$

$$V_G = \sum_{i,j\in\text{Particles}} 2\left(-\frac{G m_i m_j}{r_{ij}}\right).$$

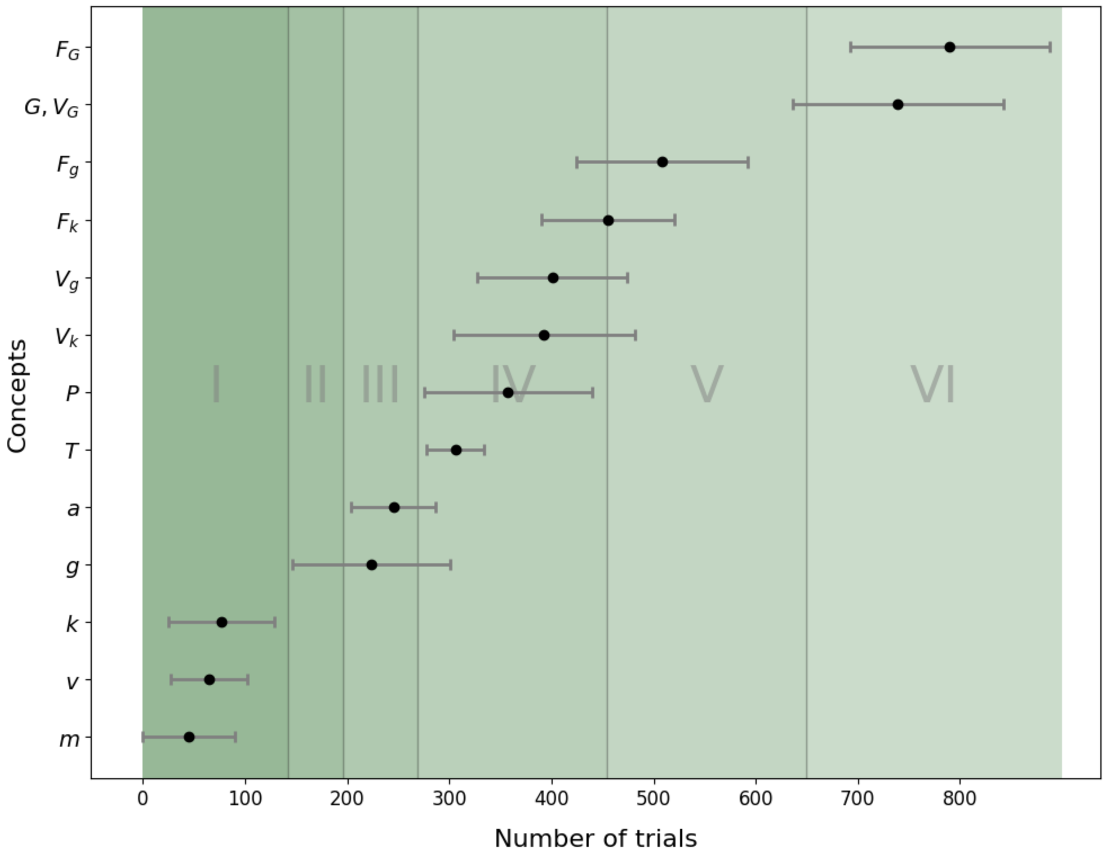
Newton's second law

$$2a_\kappa + \sum_{\lambda\in\{k,g,G\}} \delta_\lambda \left(\frac{1}{m}\frac{\partial V_\lambda}{\partial \kappa}\right) = 0, \quad \kappa \in \{x, y, z\}.$$

($\delta_\lambda = 0$ or 1, determined spontaneously during instantiation as specific laws in experiments)

Tests and results

- Statistical analysis of concept discovery timing:



Fang, DSJ, Li, Ma, 2504.01538

(Roman numerals for era numbering)

- Incremental progression, diversity

Outline

I. Background and motivation

II. AI-Newton: rediscovering classical theories

III. Challenges of quantum theories

IV. Summary and outlook

Laws of quantum physics

➢ Gap between classical and quantum system

- Collapse: No continuous measurement, only “in” and “out” states
- Uncertainty principle: No exact position, only distributions, eigenvalues…
- Nonlocality: Local measurement cannot provide complete information

➢ Key difficulty

- Different knowledge representation, different plausible reasoning strategy
- Need to construct an evolutional (continuous) theory based on discrete data, i.e., only “in” and “out” states
- Is the evolution kernel unique ?

Open issues …

Summary and outlook

- Human scientific discovery necessitates a new research paradigm, AI may help
- AI-Newton: a concept-driven physical law discovery system, no supervision, no prior physical knowledge
- Rediscovered fundamental laws: Newton’s second law, energy conservation, …
- May ultimately contribute to cutting-edge scientific discovery, though still a substantial amount of work to be accomplished
- AI for scientific discovery: remains in its nascent stage

Thank you !