AI for discovery: a preliminary glimpse

Yan-Qing Ma （Peking University）

yqma@pku.edu.cn

2025 年“微扰量子场论及其应用”前沿讲习班暨前沿研讨会，
山东大学，2025/07/09

Outline

I. Background and motivation

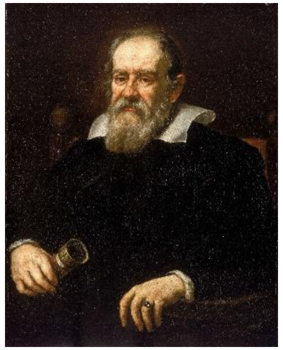
II. AI-Newton: rediscovering classical theories

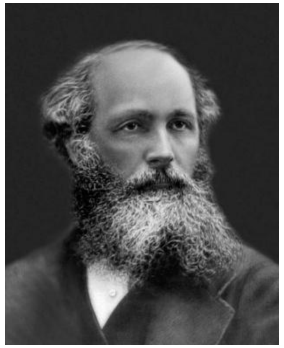
III. Challenges of quantum theories

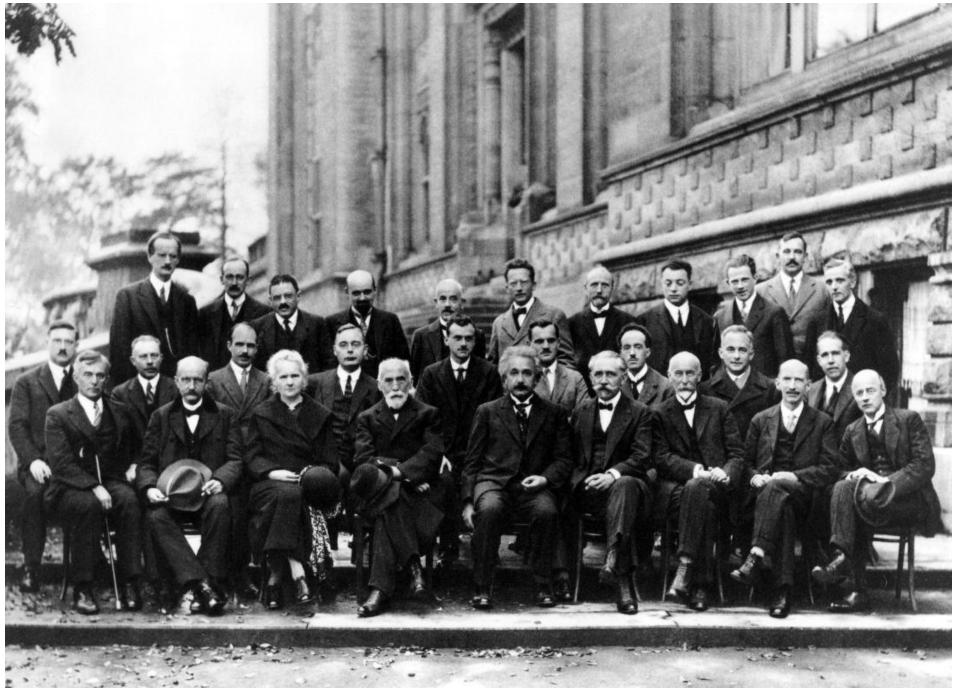
IV. Summary and outlook

Human scientific discovery

- **Fundamental physical laws: human contributions**







Galileo’s laws of motion

Newton’s laws of motion

Maxwell's electromagnetic theory

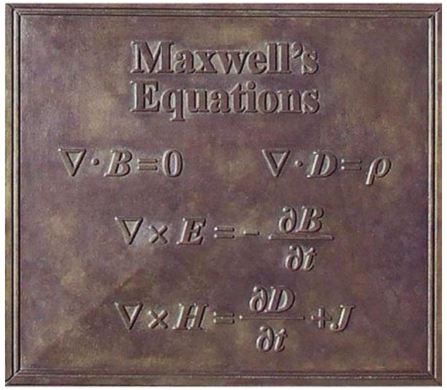
Theory of relativity

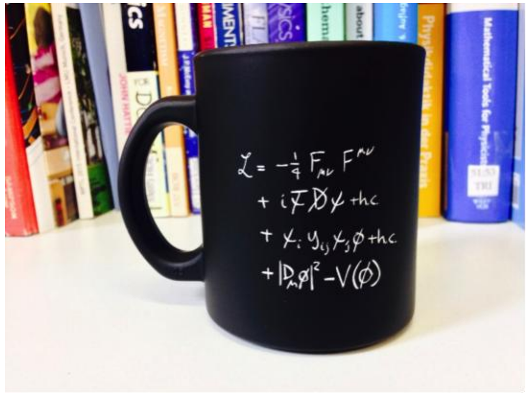
Quantum theory

Reflection

➢ Human exploration of natural laws:

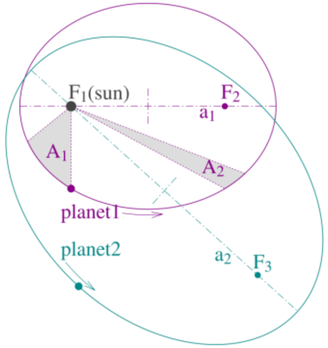
- **Advantages:** interpretability, conciseness, universality

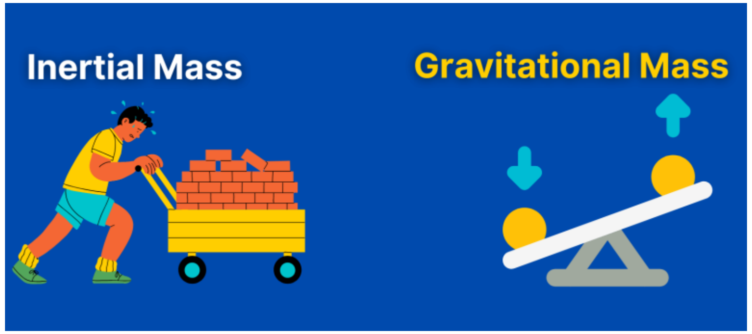


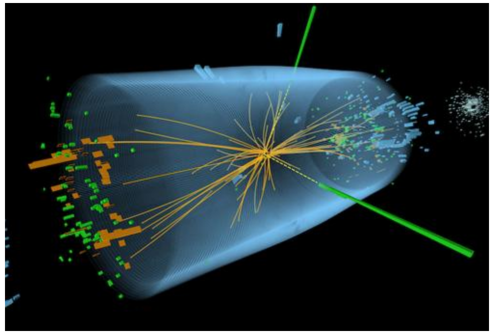


- **Disadvantages:** long period, preconceived notion, insufficient ability to handle complex problems

New paradigm?

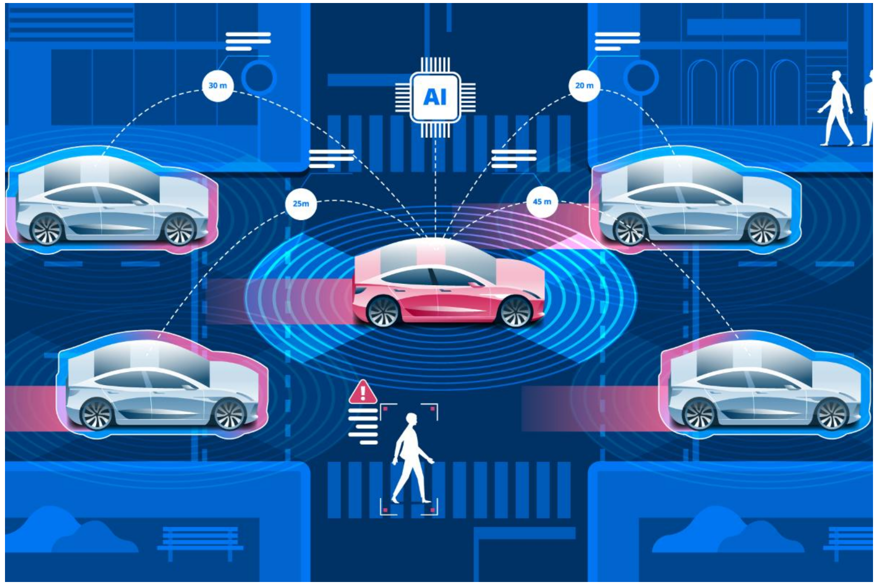






AI-driven scientific discovery

- The power of artificial intelligence (AI):



- AI-driven exploration of natural laws:

Reddy and Shojaee, 2412.11427

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

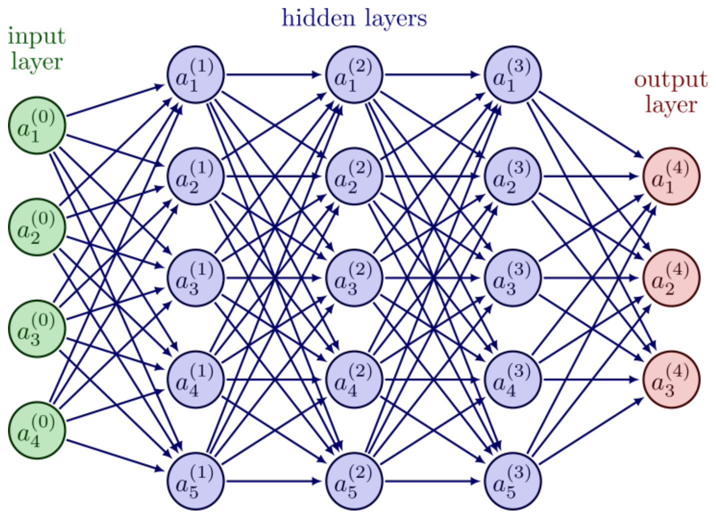
Short period, no preconceived notion, enhanced ability to handle complex problems

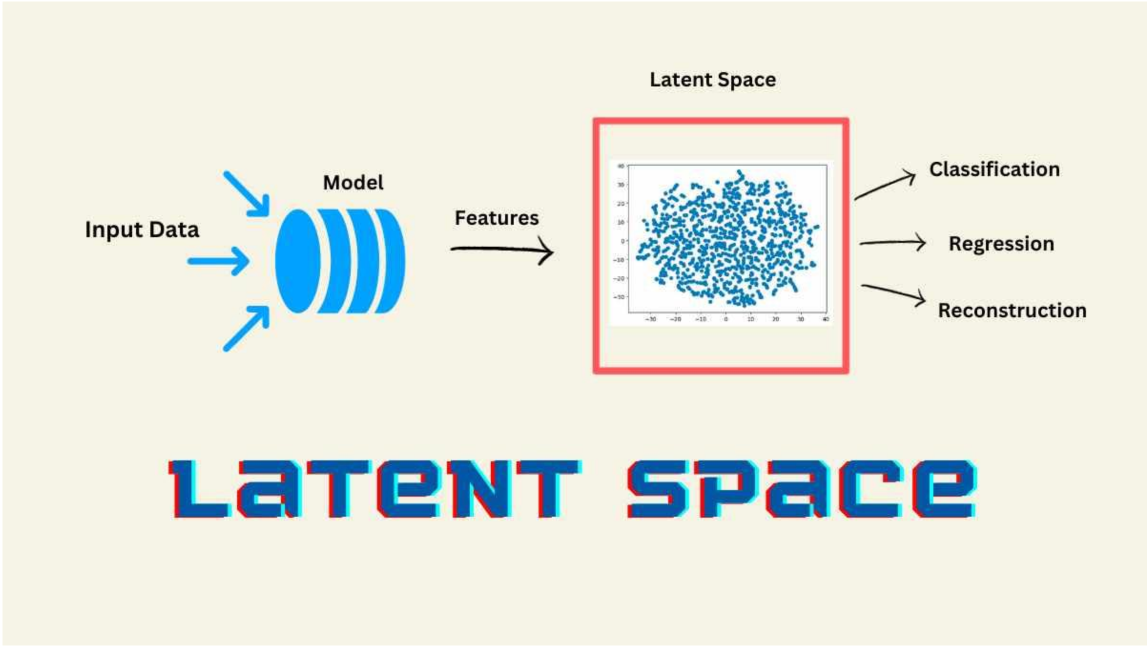
➡ **Continuous and autonomous scientific discovery? Still an open issue!**

Limitations in current methods

➢ NN-based methods:

- Handle complex problems (exceptional pattern recognition capabilities)
- Lack of interpretability (black-box)
- Insufficient extrapolation capability

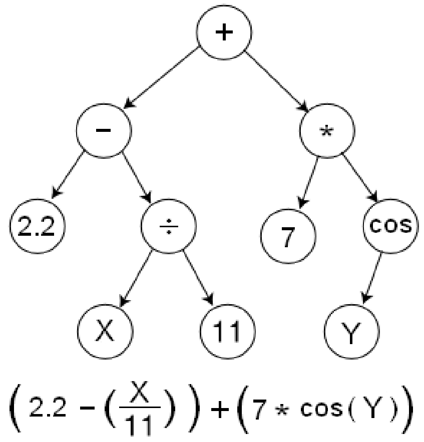


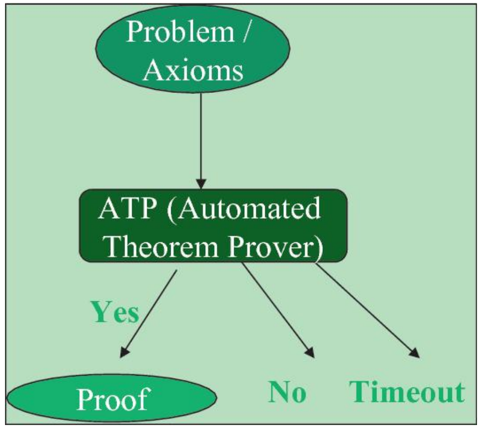


Limitations in current methods

➢ Symbolic methods:

- Good interpretability
- (With LLM) vast interdisciplinary knowledge to guide the search direction
- Limited expressive capability
- Search space explosion
- Limited cross-problem transferability





Core challenges

- **How to represent and manage physical knowledge?**
  - 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
- **How to effectively search for physics knowledge?:**
  - Search space explosion
  - Brute-force search is impractical in practice

Outline

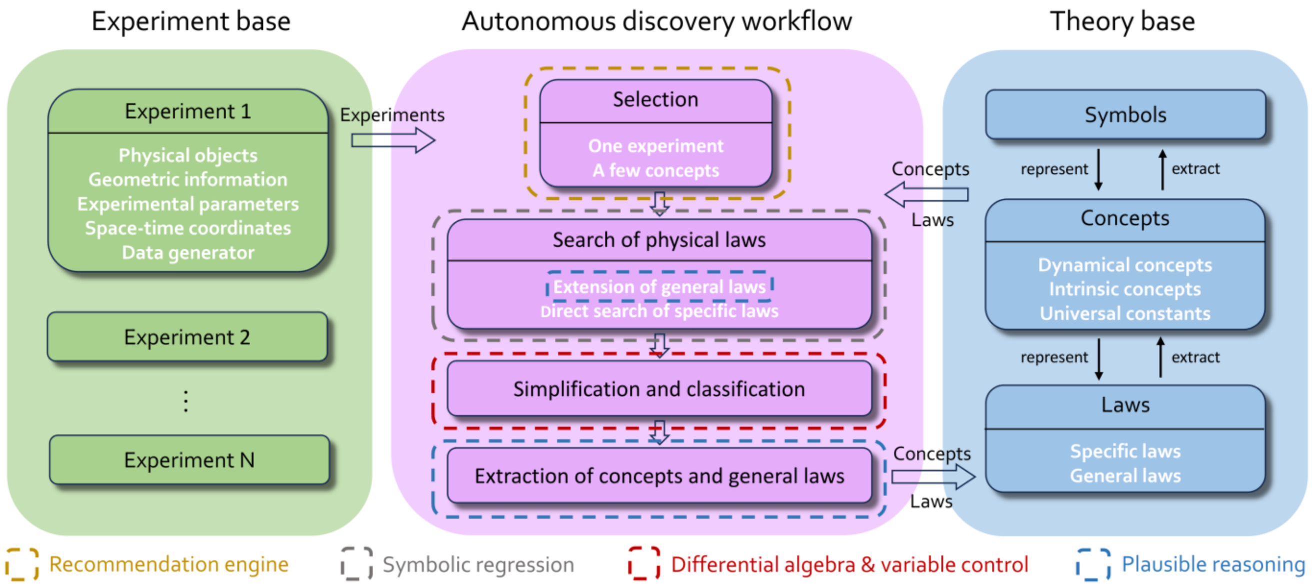
I. Background and motivation

II. AI-Newton: rediscovering classical theories

III. Challenges of quantum theories

IV. Summary and outlook

AI-Newton's architecture



Fang, Jian, Li, YQM, 2504.01538

- **Knowledge base (experiment + theory):**
  - stores and manages structured knowledge
- **Knowledge representation:**
  - employs a physical domain specific language(DSL)

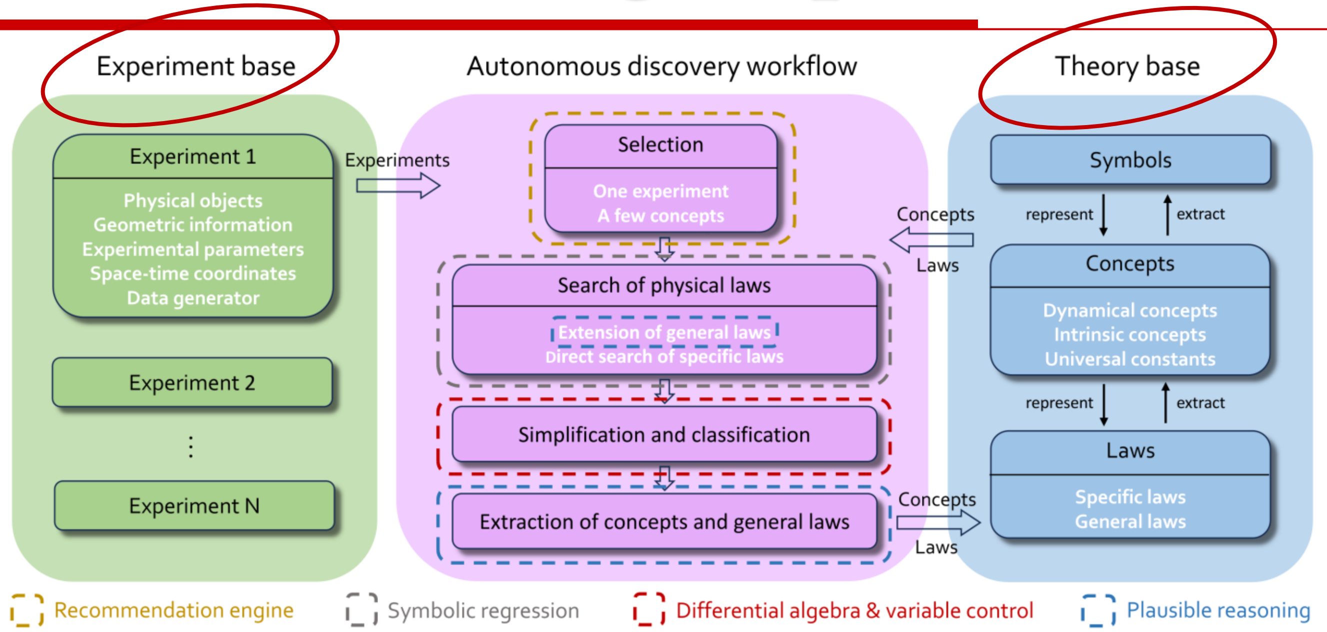
core: physical concepts

- **Autonomous discovery workflow:**
  - continuously explores physical laws,
  - collaboratively updates both general and specific knowledge

core: plausible reasoning

1. Effectively represent knowledge;
2. Reduce search space

Knowledge representation



Fang, Jian, Li, YQM, 2504.01538

- **Functions** involving experiments, physical objects, space-time coordinates, ...

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

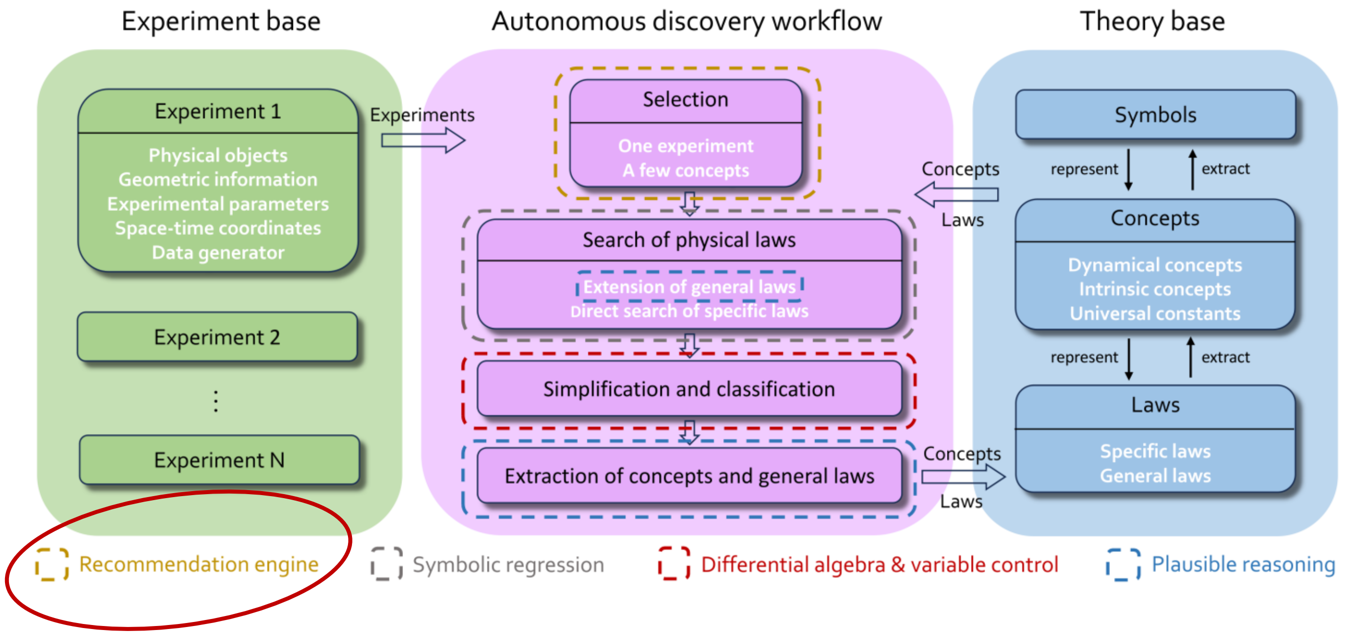
- Far beyond mere formula / NNs, e.g.:
  - Intrinsic concepts, such as mass, numerically depend solely on specific physical objects

    ⟹ Recording their measurements is essential (fixed spring or other exp.)

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

Appropriate knowledge representation is the cornerstone of autonomous discovery!

Recommendation engine



Fang, Jian, Li, YQM, 2504.01538

- **Balance exploitation and exploration:**

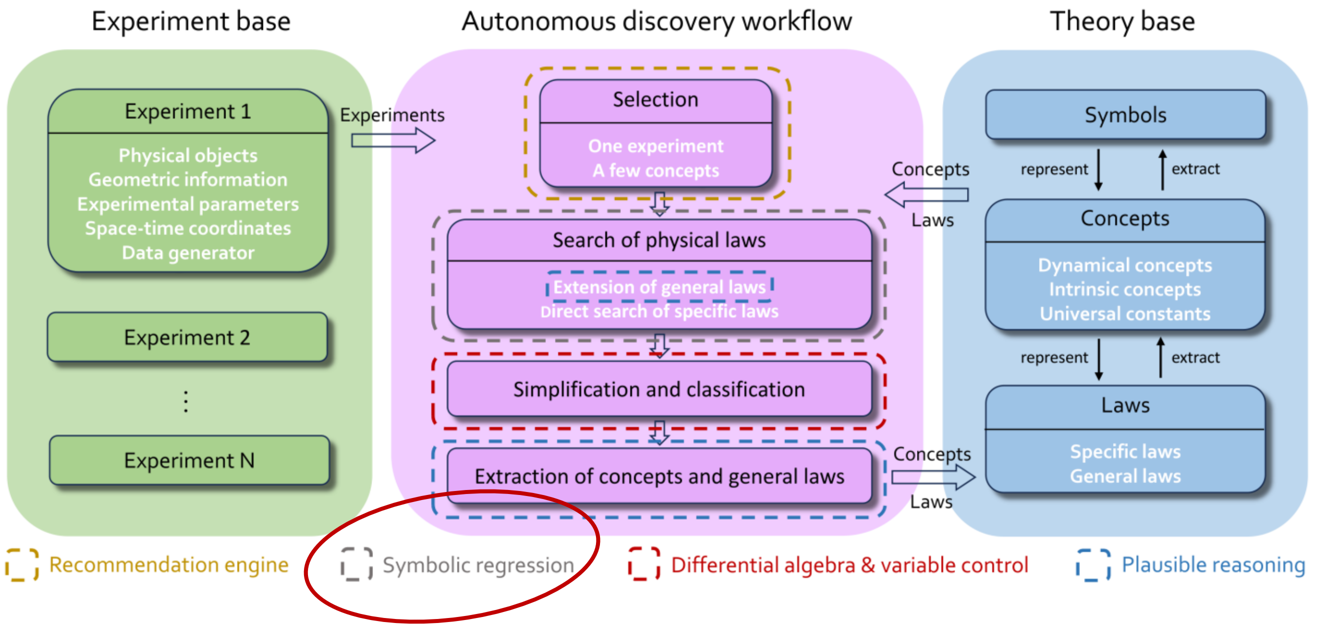
  ⟹ **Value function + dynamically adapted NN**

$$V(k) = \alpha R(k) + \sqrt{\frac{1}{1 + N(k)}}$$

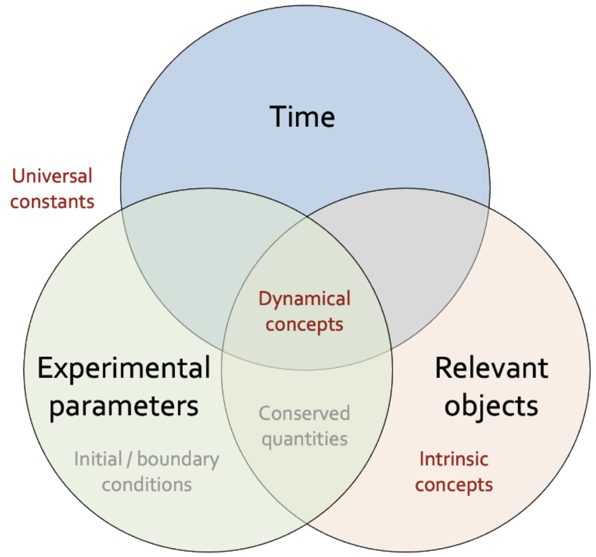
- **Prevent the workflow from grappling with complex experiments too early**

  ⟹ **The era-control strategy**

Symbolic Regression



Fang, Jian, Li, YQM, 2504.01538



- **Optimization objective:**
  - **Traditional regression: parameters**
  - **Symbolic regression: <u>function forms</u> + parameters**
- **Search space explosion**

⟹ **Good concepts and general laws to address this issue,**

(general law)

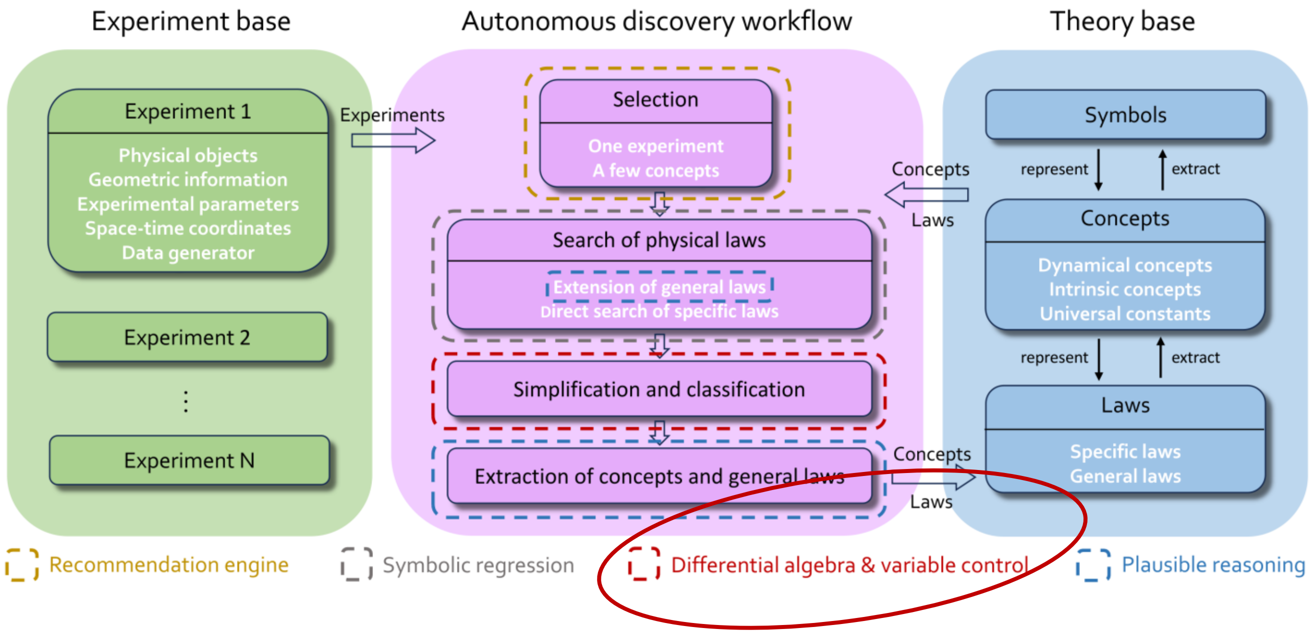
$$\forall i: \text{Ball},\ m_i a_{i,x} + (\nabla_i V_k)_x + (\nabla_i V_g)_x = 0,$$

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

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

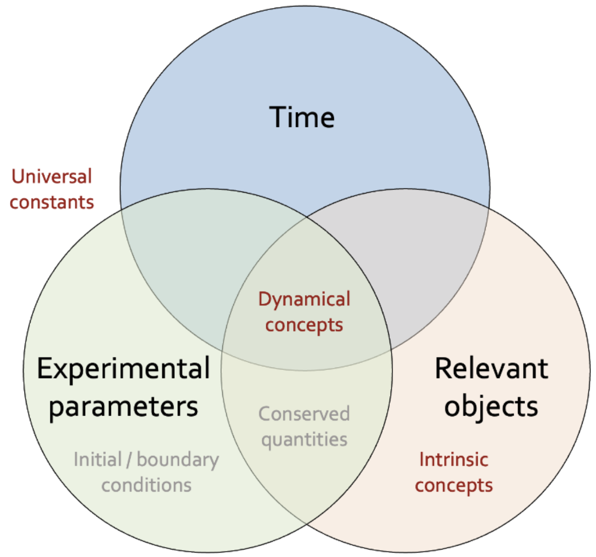
(specific law)

Differential algebra & variable control

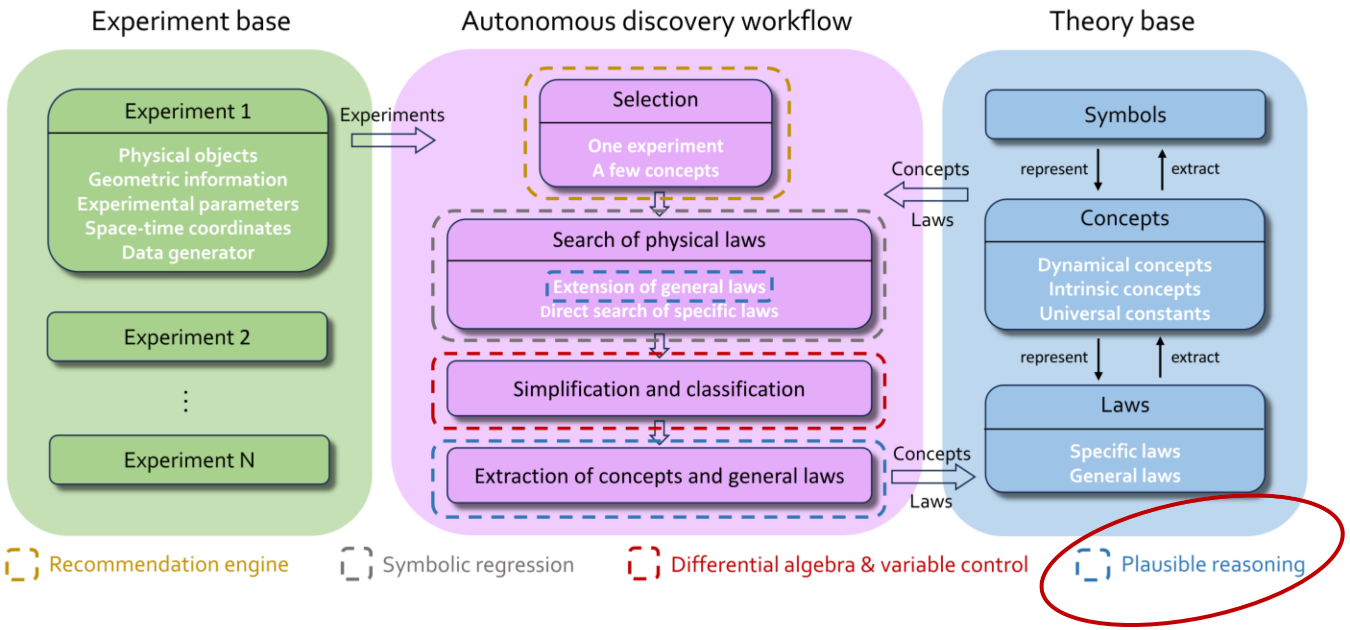


Fang, Jian, Li, YQM, 2504.01538

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



Plausible reasoning



Fang, Jian, Li, YQM, 2504.01538

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

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

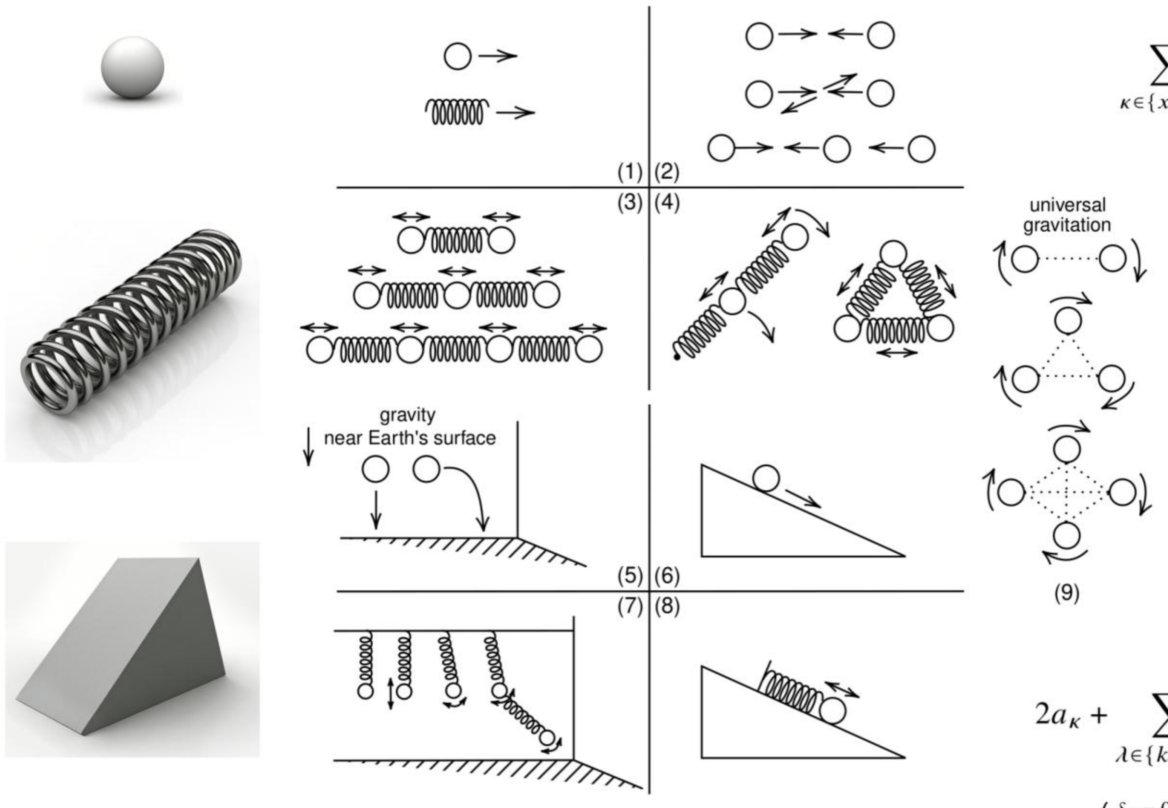
  - **Proposes and extends general laws**, e.g.:

    (in elastic collision) $T = \text{const.}$ $\Longrightarrow$ (Valids in others? Capable of extension?) (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!**

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:

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

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

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

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

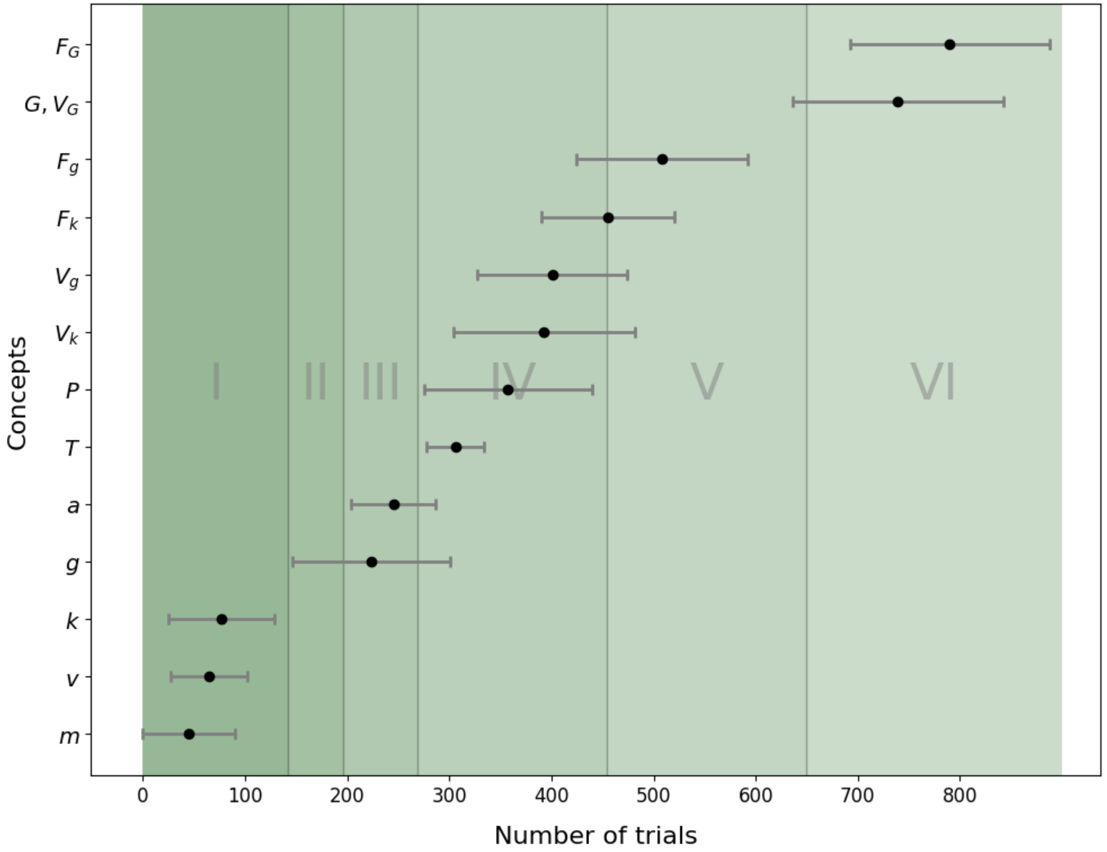
Newton's second law

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

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

Tests and results

- **Statistical analysis of concept discovery timing:**



(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

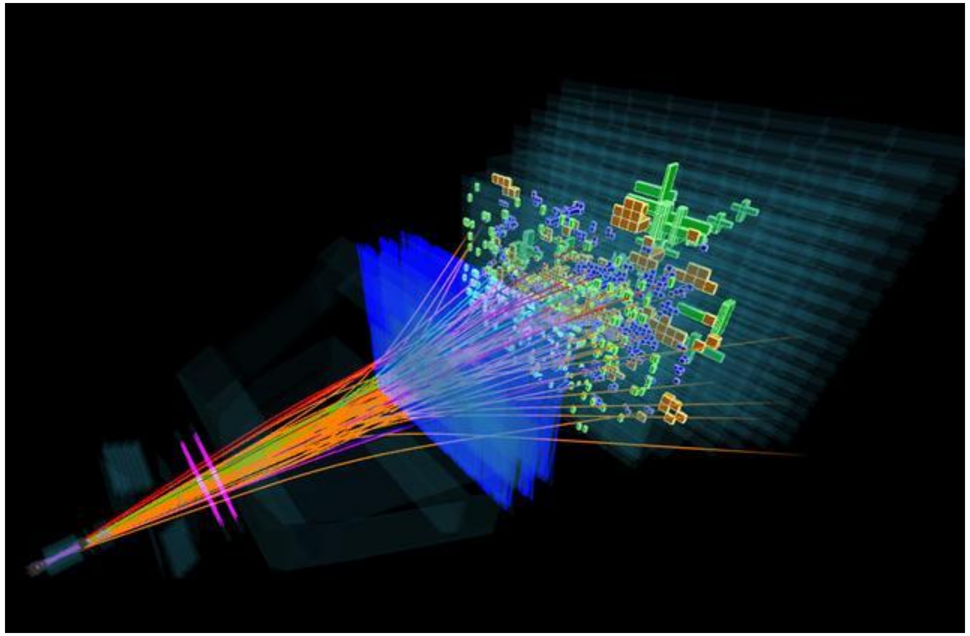
➢ Gap between classical and quantum system

- 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

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

Dynamics of hadronization



- **Evolution of quantum state**

$$i\frac{\partial}{\partial t}|\psi\rangle = \hat{H}|\psi\rangle \;\; \rightarrow \langle\phi|\hat{\mathcal{T}}|\psi\rangle$$

- **Human's method**
  - Construct lower dimensional projection, like FFs, EECs, etc.

- **Is AI a way to understand it as a whole?**
  - How to parameterize Fork space (combining partons and hadrons)?
  - What kind of evolution equtions do we expect?

Still in progress…

A simple example: Learning potential

➤ **Potential model: fundamental of non-relativistic system**

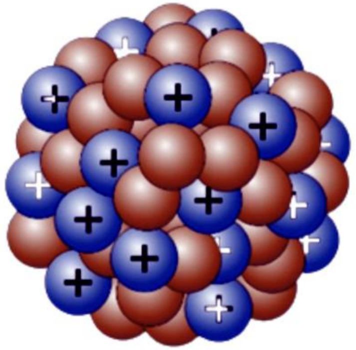
Theory

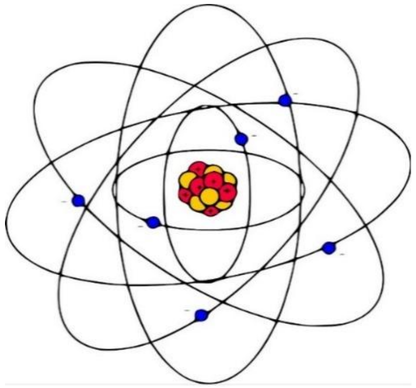
$$\left(-\frac{\nabla^2}{2m}+V\right)\psi_n(\vec{x},s)=E_n\psi_n(\vec{x},s)$$

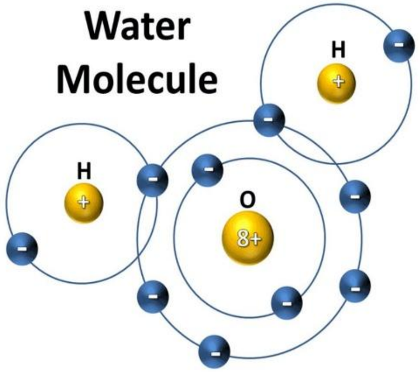
↑ inverse spectrum problem of Schrodinger's equation

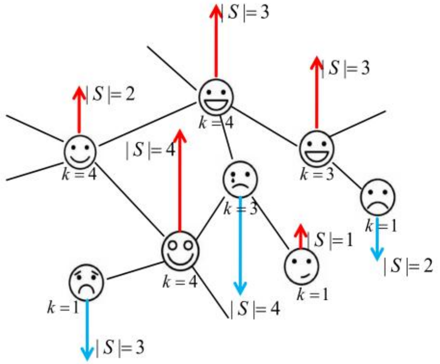
Experiment

energy levels, scattering,…





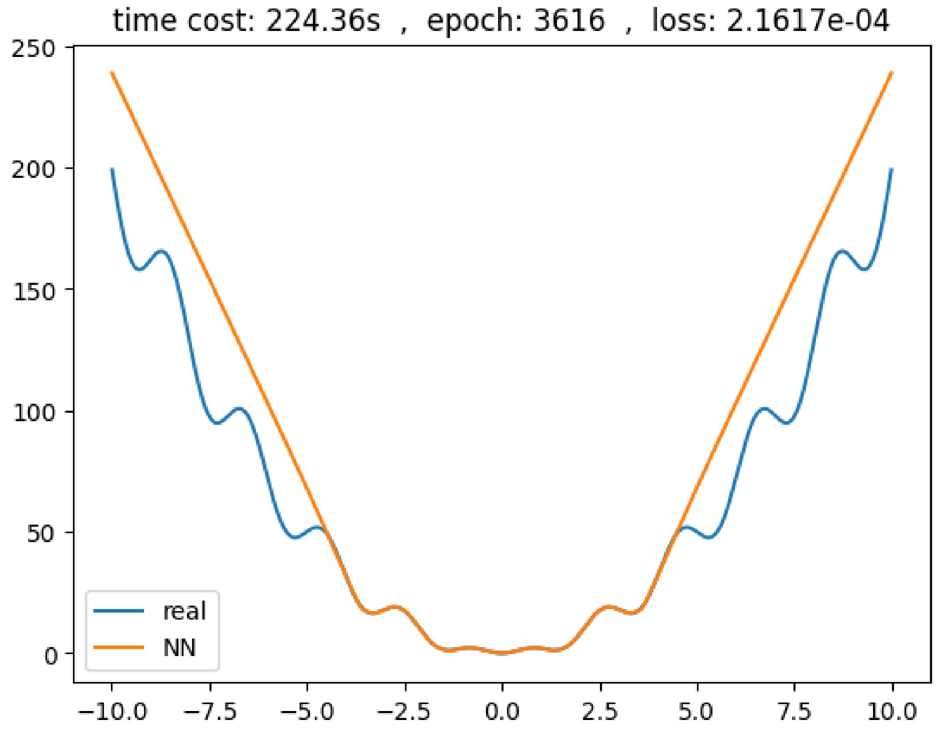




A simple example: Learning potential

➤ **Limited number of energy levels as input**

- Learning from the first a few energy levels
- No prior assumptions beyond symmetry
- Result: stable learning of potential within the effective range



$$V(x) = x\sin(\pi x) + x^2$$

from the first 10 energy levels

Fully data-driven

providing guidance for theoretical models

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, like mechanics of hadronization, though still a substantial amount of work to be accomplished
- AI for scientific discovery: remains in its nascent stage

Thank you !