

Framework and Methodology

- **Visual Encoding**

- Next-Generation Encoder
- Native Foundations

- **World Rules**

- Physical Laws
- Human Knowledge
- Understanding→Discovery

- **Interactive Conditions**

- Universal Encoder
- Hierarchy

- **Simulation**

- Joint, Cross-Modal Simulation
- Perception, Reasoning and Imagination

- **Different Designs**

- Fusion and Unification
- 3D Models→Dynamic Simulators

Evaluation and Application

- **Complete Evaluation**

- Physics Judgment Model
- Real-World Complexity Datasets

- **Downstream Tasks**

- As Multifaceted Tools
- Opportunities for Grounded Intelligence

Emerging Frontiers

- **Vision**

- The Upcoming Vision's GPT-3 Moment

- **LLM**

- Leveraging LLM Trajectories for VWMs