A Survey of Robotic Language Grounding:

Tradeoffs between Symbols and Embeddings

Vanya Cohen, **Jason Xinyu Liu**, Raymond Mooney, Stefanie Tellex, David Watkins

IJCAI 2024 Survey Track

August 8th, 2024

IJCAI JEJU 2024 International Joint Conference on Artificial Intelligence Jeju 03.08.24 - 09.08.24

BROWN

TEXAS The University of Texas at Austin

Boston Dynamics AI INSTITUTE

Robotic Language Grounding

Connect linguistic elements in language to the robot’s perception of and actions in the physical world.

Robotic Language Grounding

Connect linguistic elements in language to the robot's perception of and actions in the physical world.

1. What grounding representation to use?

Robotic Language Grounding

Connect linguistic elements in language to the robot's perception of and actions in the physical world.

1. What grounding representation to use?

2. How to ground natural language to the grounding representation of choice?

Robotic Language Grounding

Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

Grounding Language to Symbols

Symbols

- Discrete
- More Structure; More bias
- Unambiguous
- Verifiable
- Interpretable

Grounding Language to Embeddings

Symbols

- Discrete
- More Structure; More bias
- Unambiguous
- Verifiable
- Interpretable

High-dimensional Embeddings

- Continuous
- Less structure; More variance
- Adaptive

Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

Grounding Language to Logic

Grounding Language to Logic

Pros

- Unambiguous semantics
- Verifiable
- Interpretable
- Reduce search space

Grounding Language to Logic

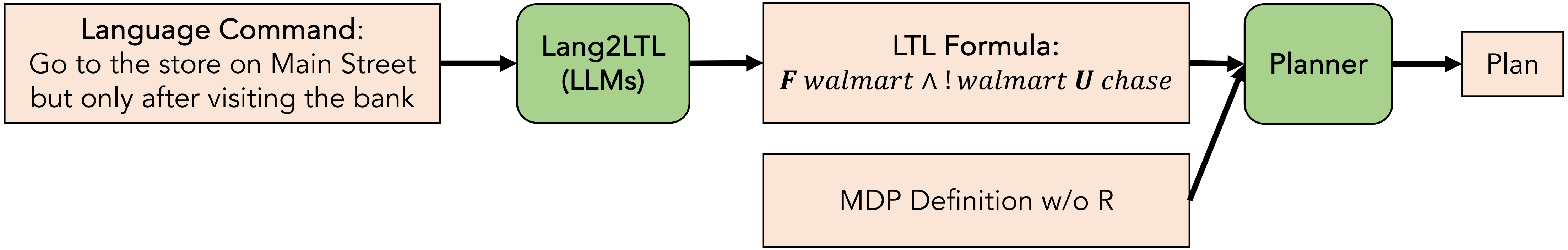
Pros

- Unambiguous semantics
- Verifiable
- Interpretable
- Reduce search space

Cons

- Require manually defined structures
- Difficult to represent low-level control

Grounding Language to Logic: Lang2LTL



Lang2LTL

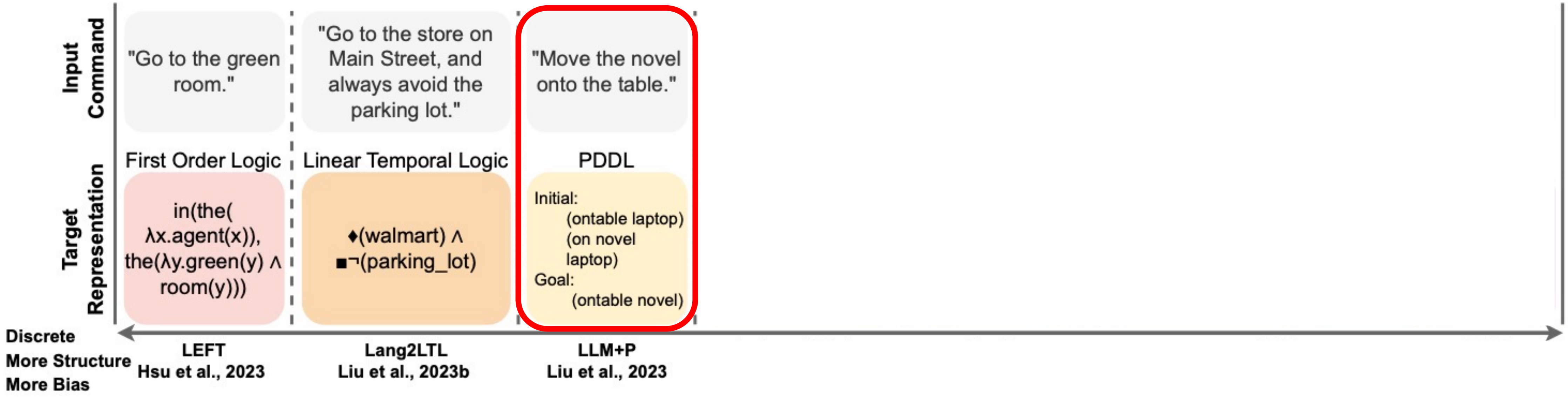
- Natural language navigation command
- Modular system produces a grounded linear temporal logic (LTL) formula
- Given MDP definition
- Planner outputs a trajectory

Grounding Language to Logic

More Papers

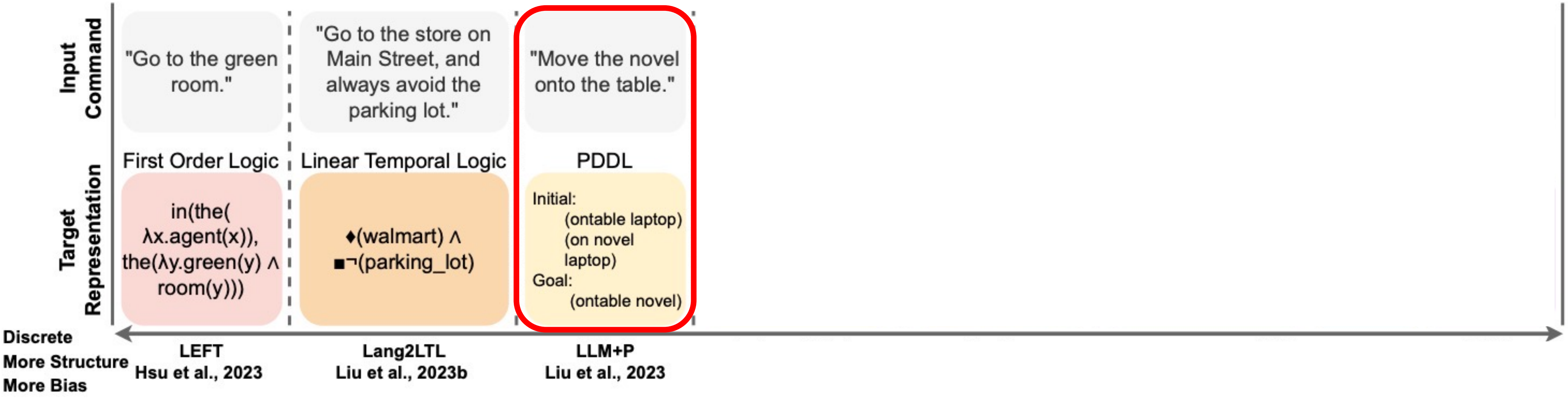
- Lang2LTL-2: Grounding Spatiotemporal Navigation Commands Using Large Language and Vision-Language Models [Liu et al. 2024]
- AutoTAMP: Autoregressive Task and Motion Planning with LLMs as Translators and Checkers [Chen et al. 2024]
- NL2TL: Transforming Natural Languages to Temporal Logics using Large Language Models [Chen et al. 2023]
- NL2LTL: a Python Package for Converting Natural Language (NL) Instructions to Linear Temporal Logic (LTL) Formulas [Fuggitti and Chakraborti 2023]

Grounding Language to PDDL



Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

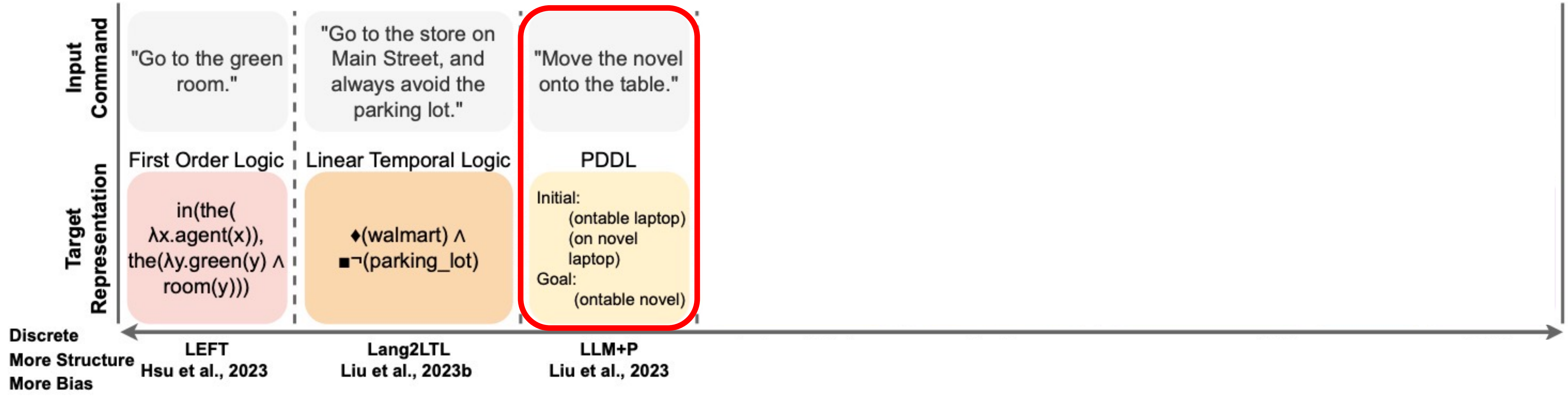
Grounding Language to PDDL



Pros

- Sound
- Complete
- (Often) Optimal

Grounding Language to PDDL



Pros

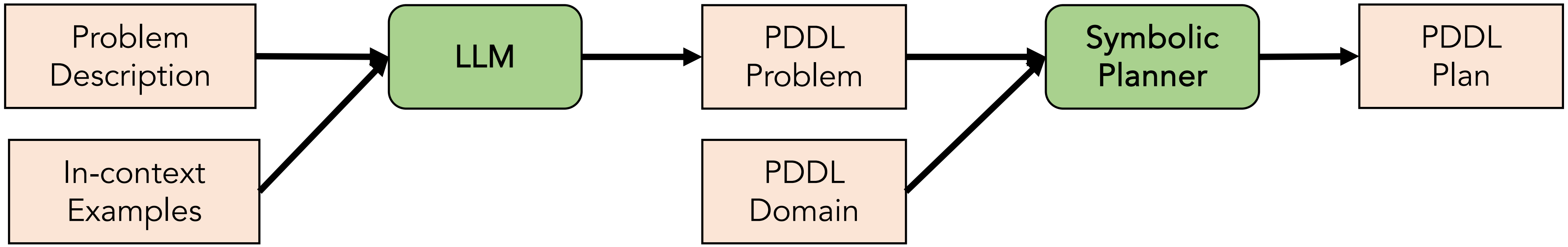
- Sound
- Complete
- (Often) Optimal

Cons

- Require manually defined structures

Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

Grounding Language to PDDL: LLM+P

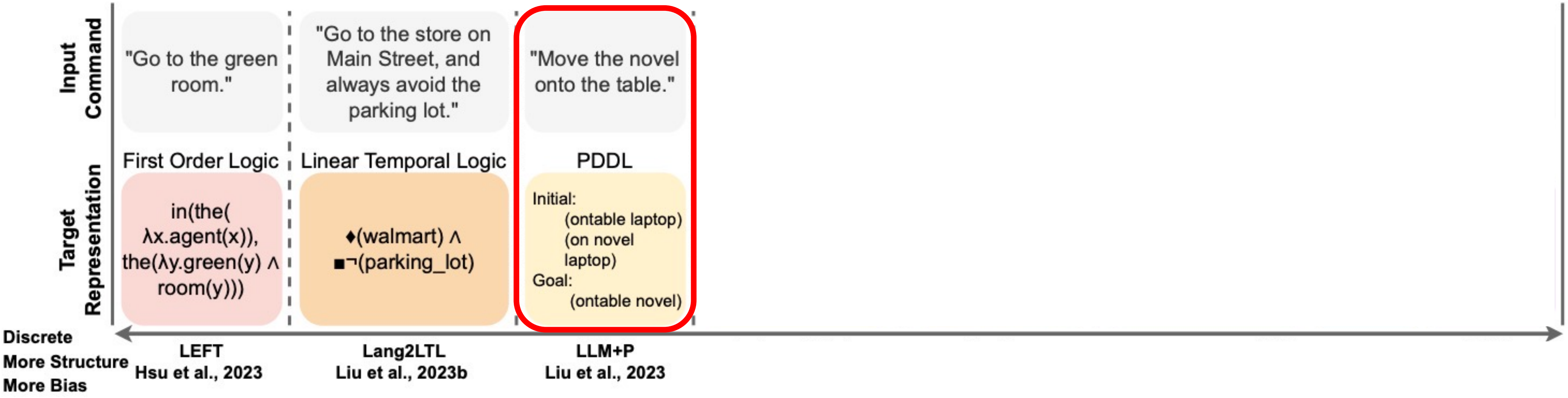


LLM+P

- Natural language description of a planning problem
- LLM translates it to PDDL problem
- Given a PDDL domain description, i.e., action preconditions and effects
- Symbolic planner solves PDDL

Liu et al. Preprint 2023. LLM+P: Empowering Large Language Models with Optimal Planning Proficiency

Grounding Language to PDDL

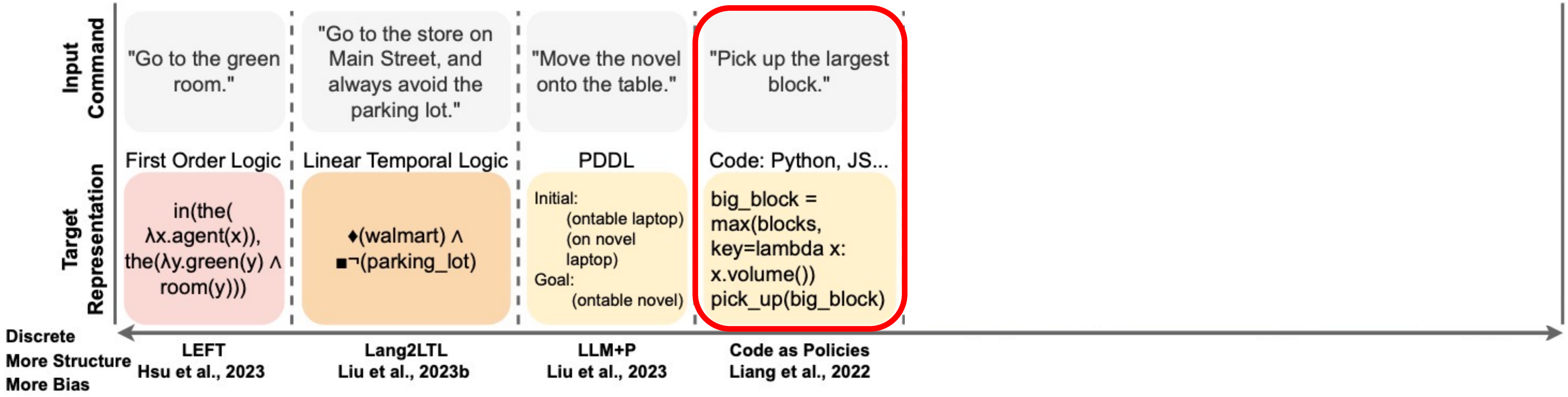


More Papers

- Translating Natural Language to Planning Goals with Large-Language Models [Xie et al. 2023]
- Structured, Flexible, and Robust: Benchmarking and Improving Large Language Models Towards More Human-like Behavior in Out-of-distribution Reasoning Tasks [Collins et al. 23]
- Leveraging Pre-trained Large Language Models to Construct and Utilize World Models for Model-based Task Planning [Guan et al. 2023]
- PlanBench: An Extensible Benchmark for Evaluating Large Language Models on Planning and Reasoning about Change [Valmeekam et al. 2023]
- On the Planning Abilities of Large Language Models : A Critical Investigation [Valmeekam et al. 2023]

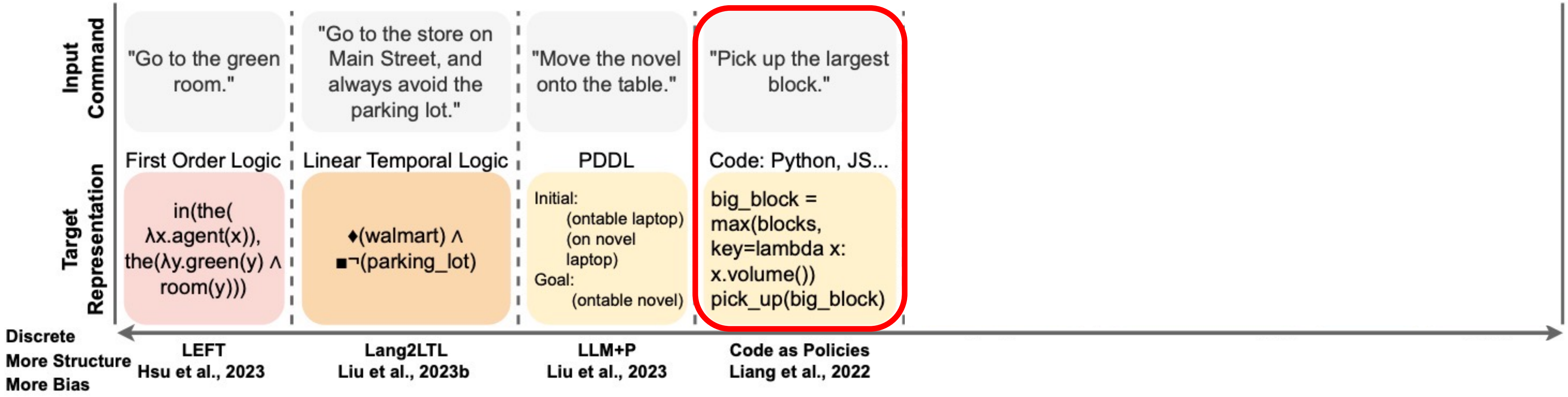
Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

Grounding Language to Code



Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

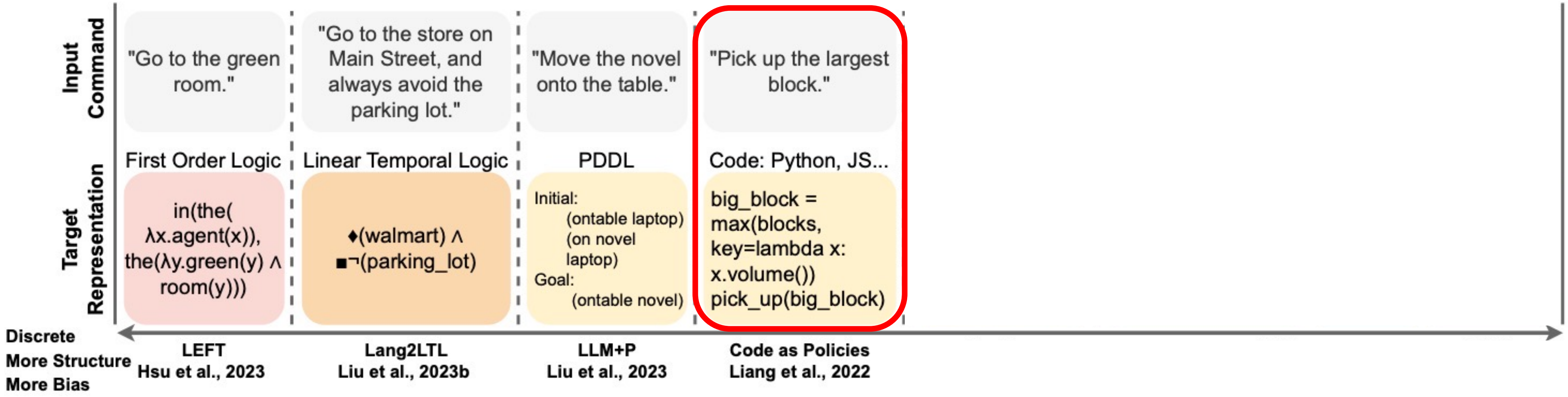
Grounding Language to Code



Pros

- Flexible
- High-level plan and low-level control

Grounding Language to Code



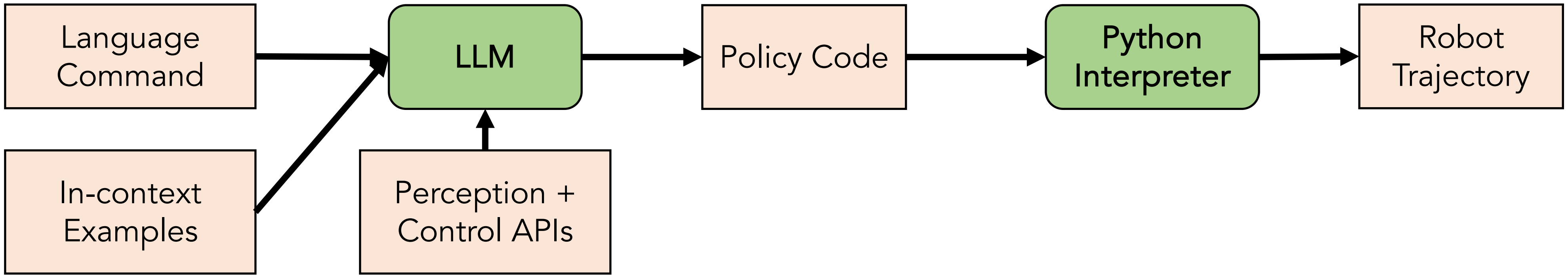
Pros

- Flexible
- High-level plan and low-level control

Cons

- Require predefined perception and control models in specific domains

Grounding Language to Code: Code as Policies

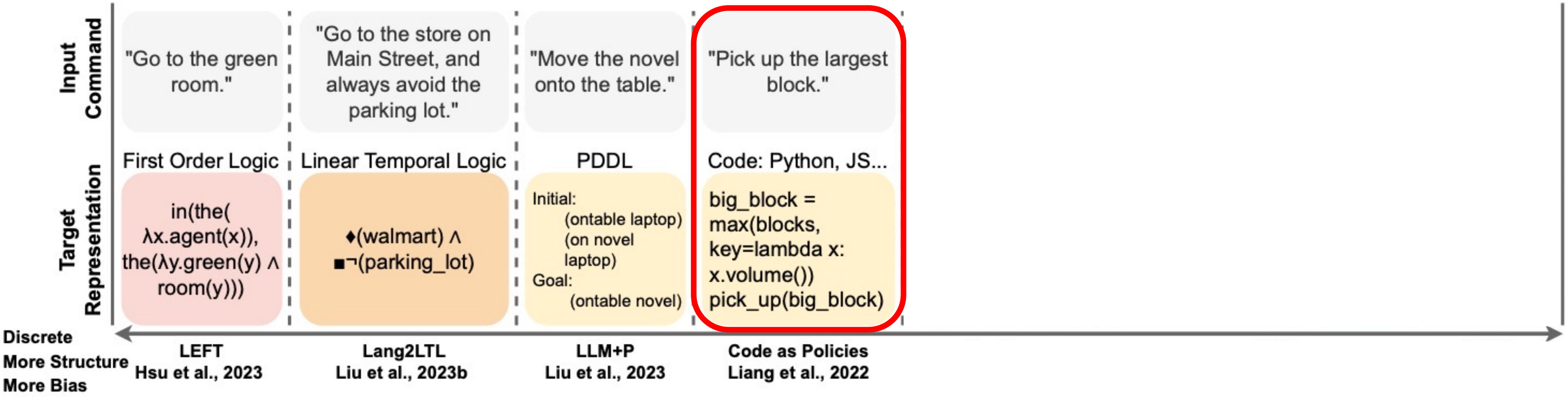


Code as Policies

- Natural language command
- Given predefined perception and control models
- Code-writing LLM outputs executable code

Liang et al. ICRA 2023. Code as Policies: Language Model Programs for Embodied Control

Grounding Language to Code

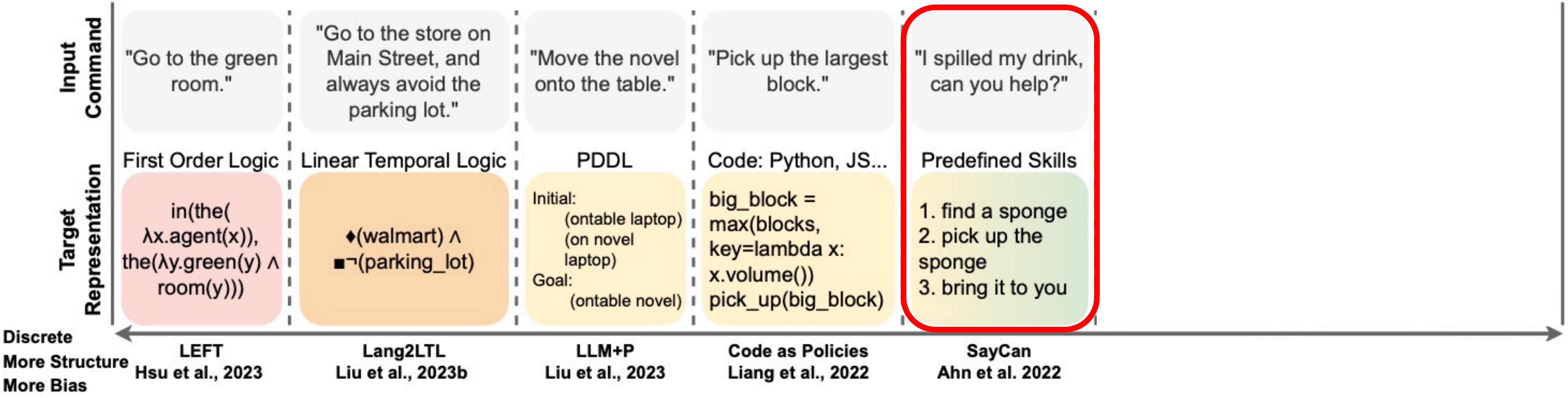


More Papers

- Embodied AI with Two Arms: Zero-shot Learning, Safety and Modularity [Varley et al. 2024]
- ProgPrompt: Generating Situated Robot Task Plans using Large Language Models [Singh et al. 2023]
- Socratic Models: Composing Zero-Shot Multimodal Reasoning with Language [Zeng et al. 2023]
- ITP: Interactive Task Planning with Language Models [Li et al. 2023]
- Voyager: An Open-ended Embodied Agent with Large Language Models [Wang et al. 2023]

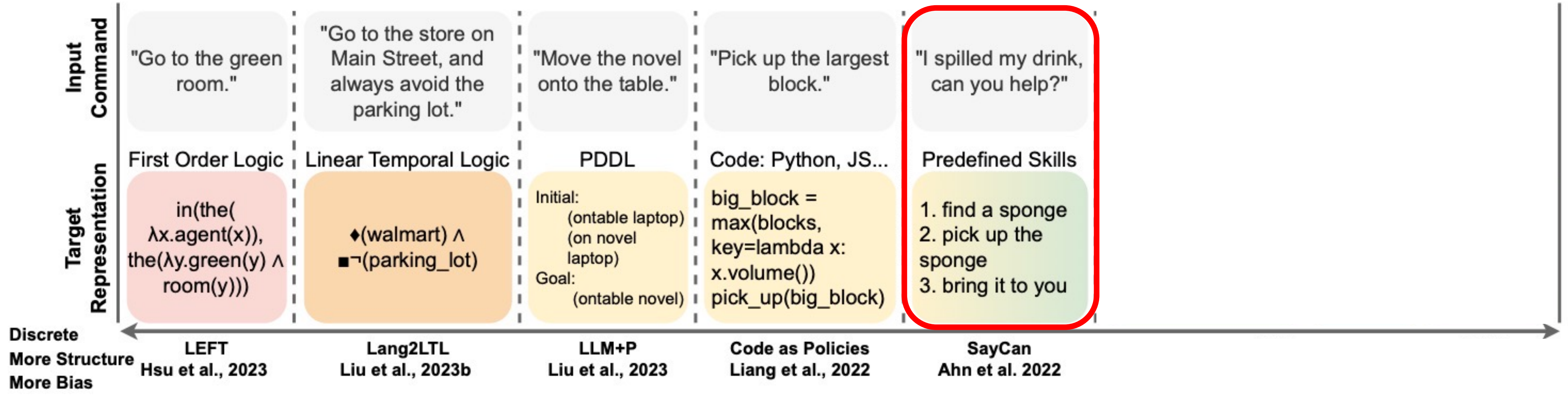
Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

Grounding Language to Predefined Skills



Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

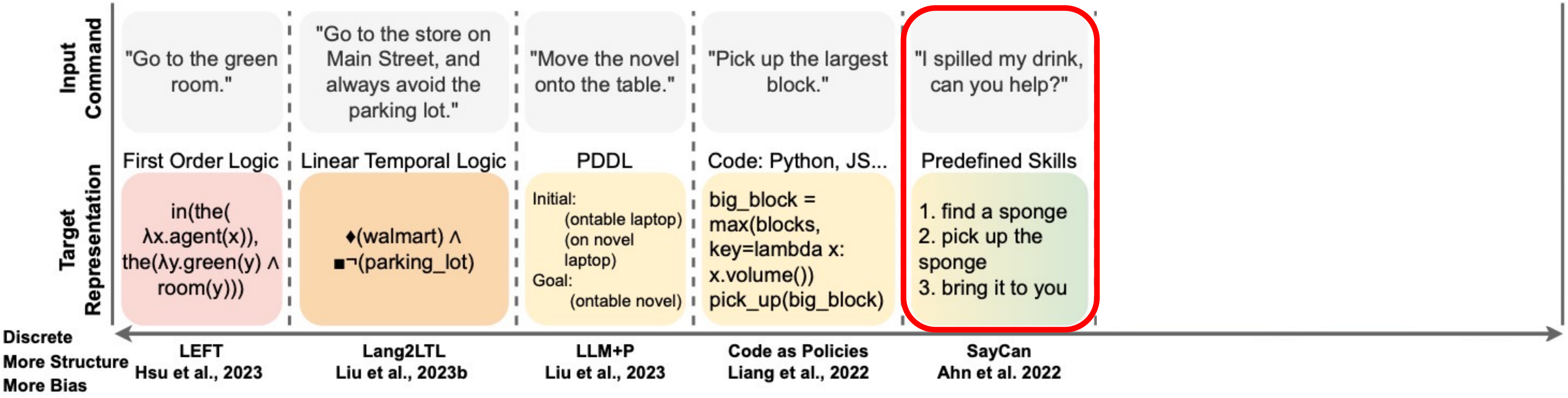
Grounding Language to Predefined Skills



Pros

- Adaptive

Grounding Language to Predefined Skills



Pros

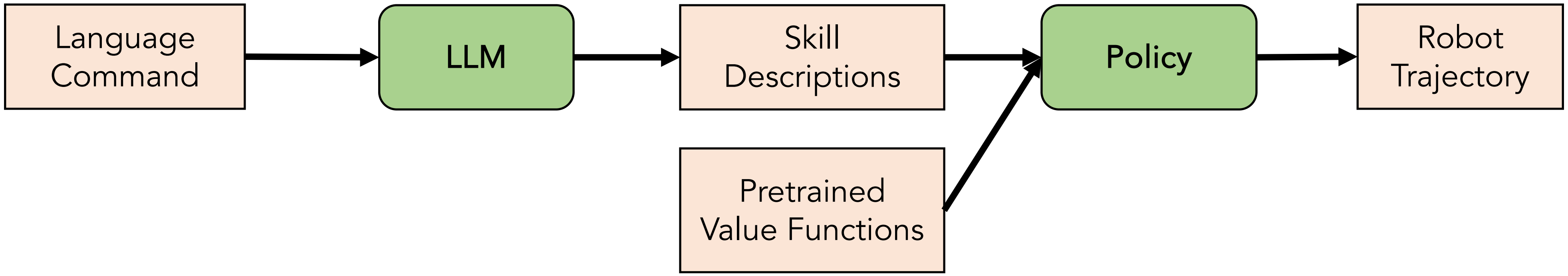
- Adaptive

Cons

- Require predefined skills
- Possibly incorrect plans

Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

Grounding Language to Predefined Skills: SayCan

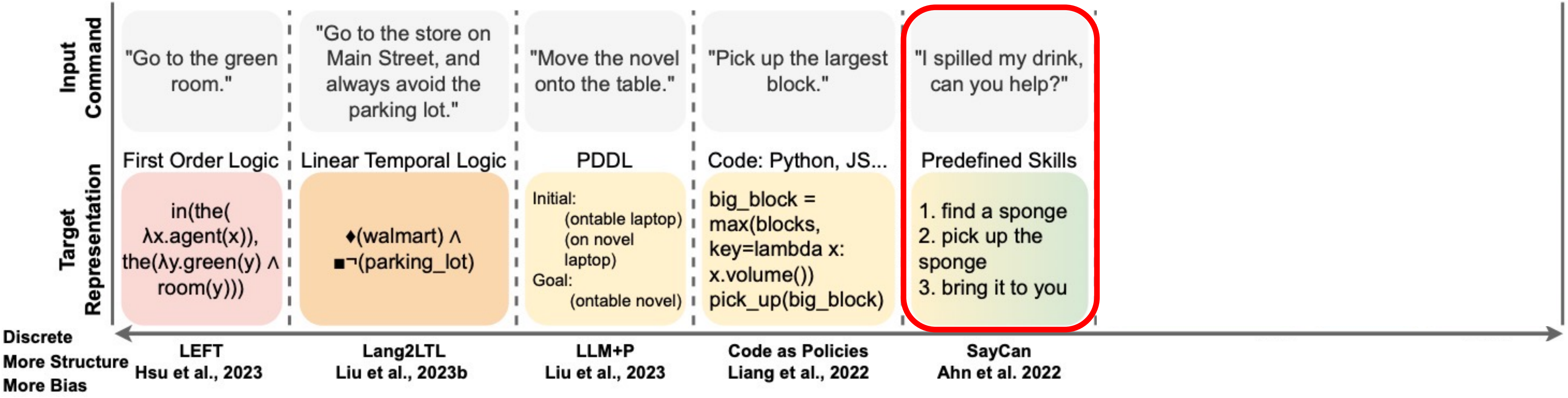


SayCan

- Natural language command
- LLM proposes candidate skills every step
- Pretrained value functions to rank available skills
- Language-conditioned policies execute the top skill

Ahn et al. CoRL 2022. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances

Grounding Language to Predefined Skills

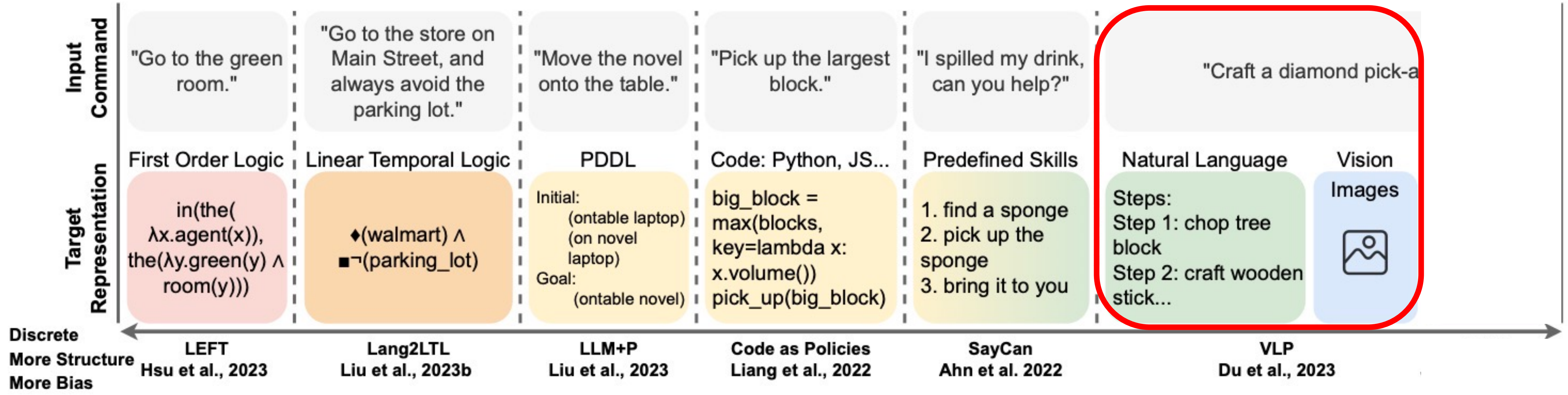


More Papers

- CAPE: Planning with Large Language Models via Corrective Re-prompting [Raman et al. 2024]
- Inner Monologue: Embodied Reasoning through Planning with Language Models [Huang et al. 2022]
- Language Models as Zero-shot Planners: Extracting Actionable Knowledge for Embodied Agent [Huang et al. 2022]

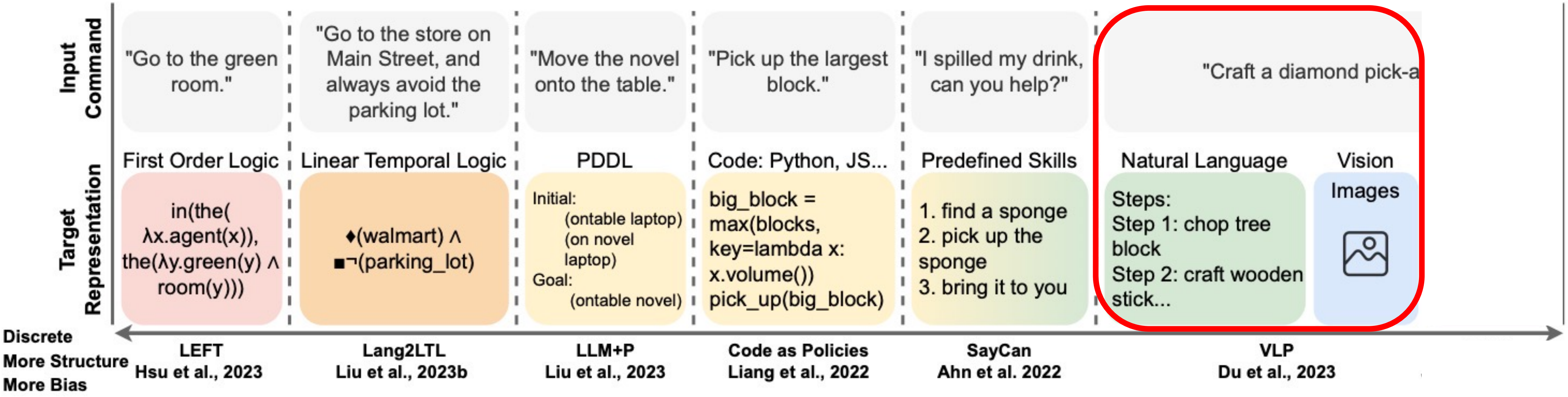
Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

Grounding Language to Subgoals



Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

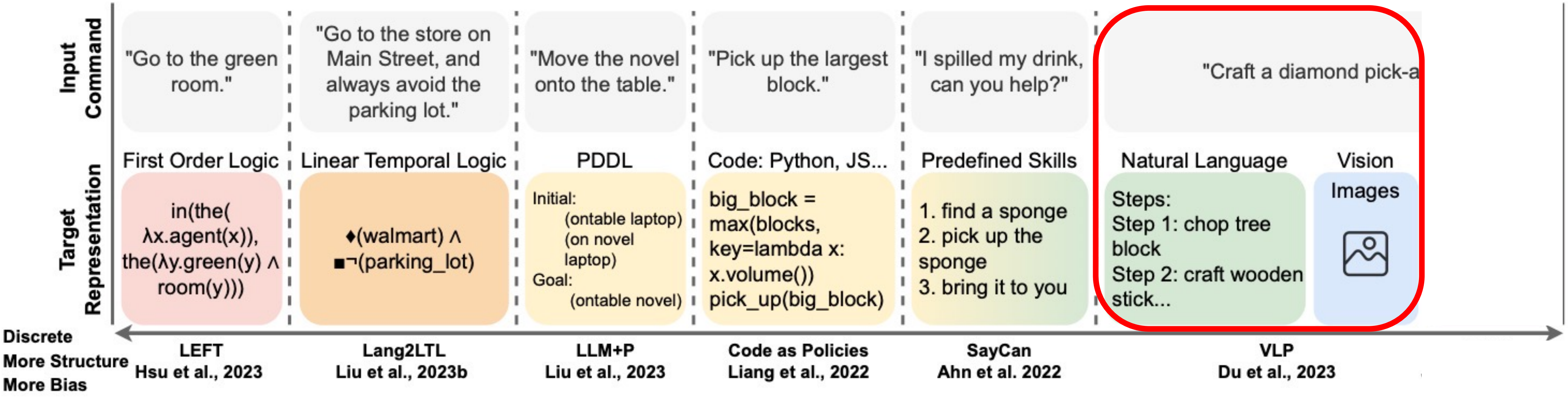
Grounding Language to Subgoals



Pros

- Adaptive

Grounding Language to Subgoals



Pros

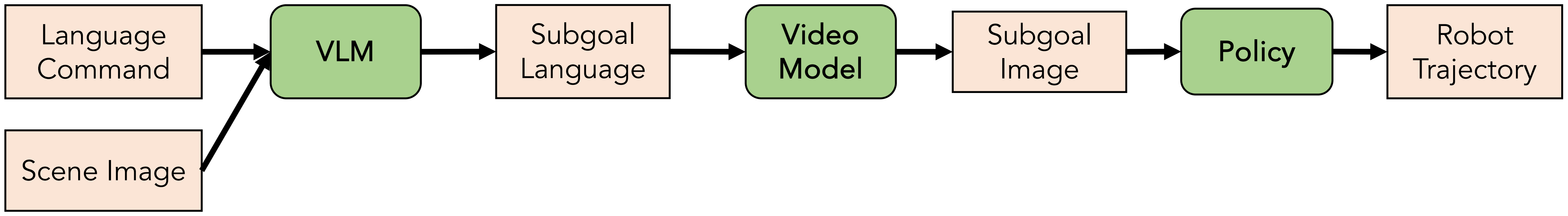
- Adaptive

Cons

- Require predefined skills
- Possibly incorrect plans

Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

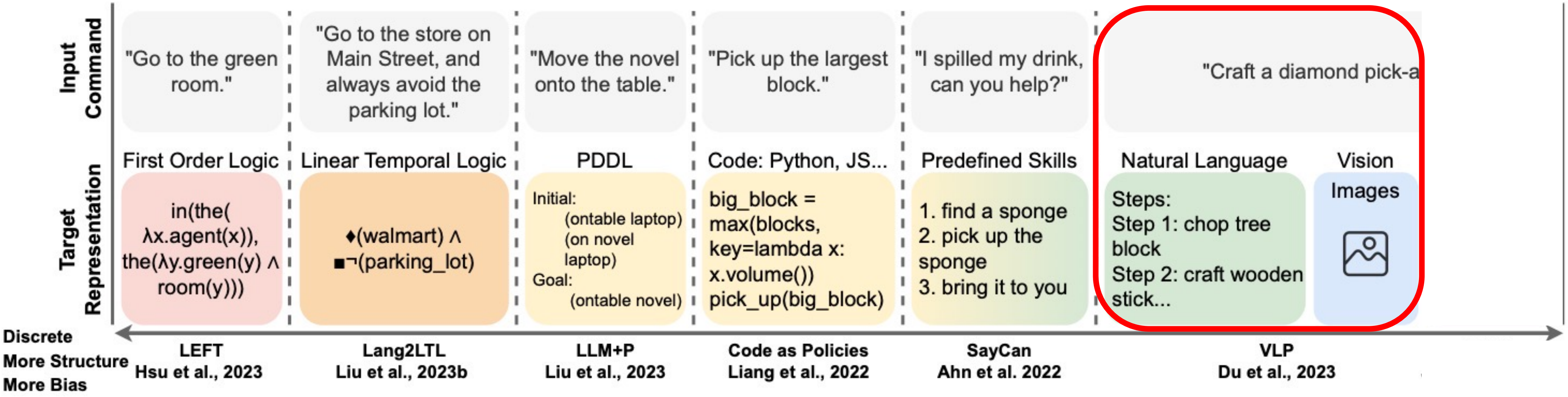
Grounding Language to Subgoals: VLP



Video Language Planning (VLP)

- Tree search
- VLM proposes language subgoals
- Video model conditioned on text generates image subgoals
- Policy conditioned on image executes the plan

Grounding Language to Subgoals

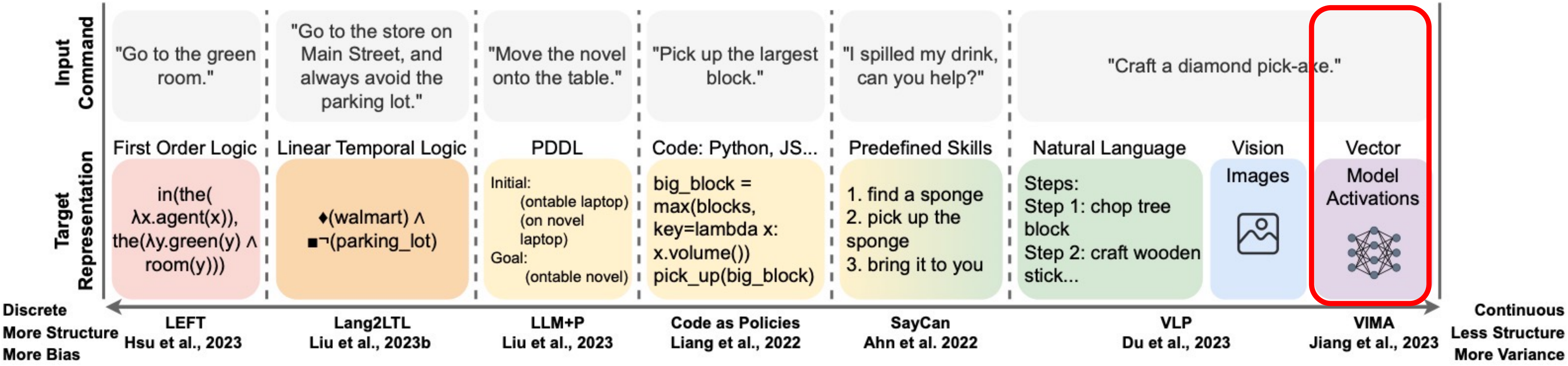


More Papers

- Zero-Shot Robotic Manipulation with Pretrained Image-Editing Diffusion Models [Black et al. 2023]
- UniSim: A Neural Closed-Loop Sensor Simulator [Yang et al. 2023]
- GAIA-1: A Generative World Model for Autonomous Driving [Hu et al. 2023]

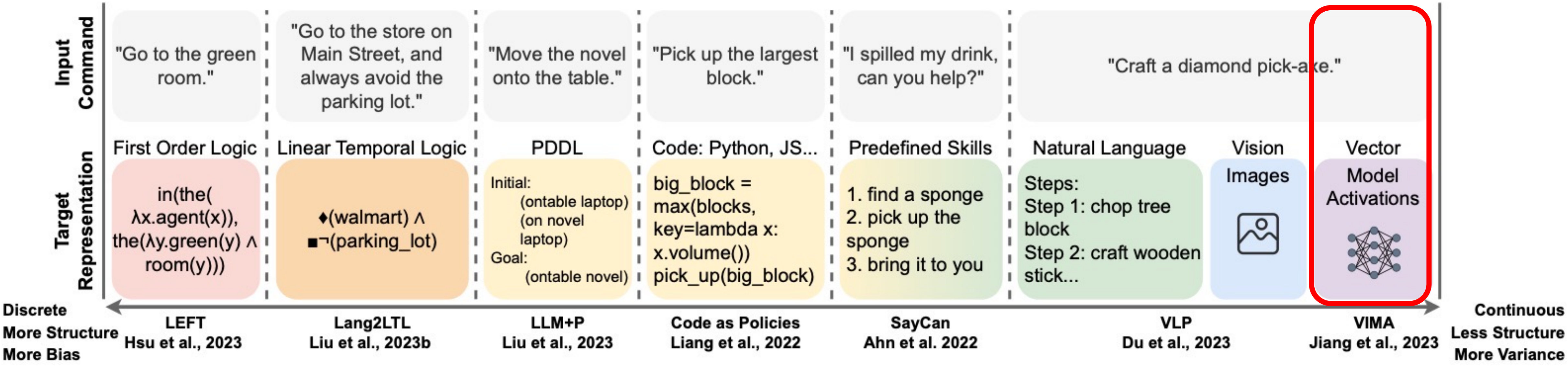
Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

Grounding Language to Embeddings



Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

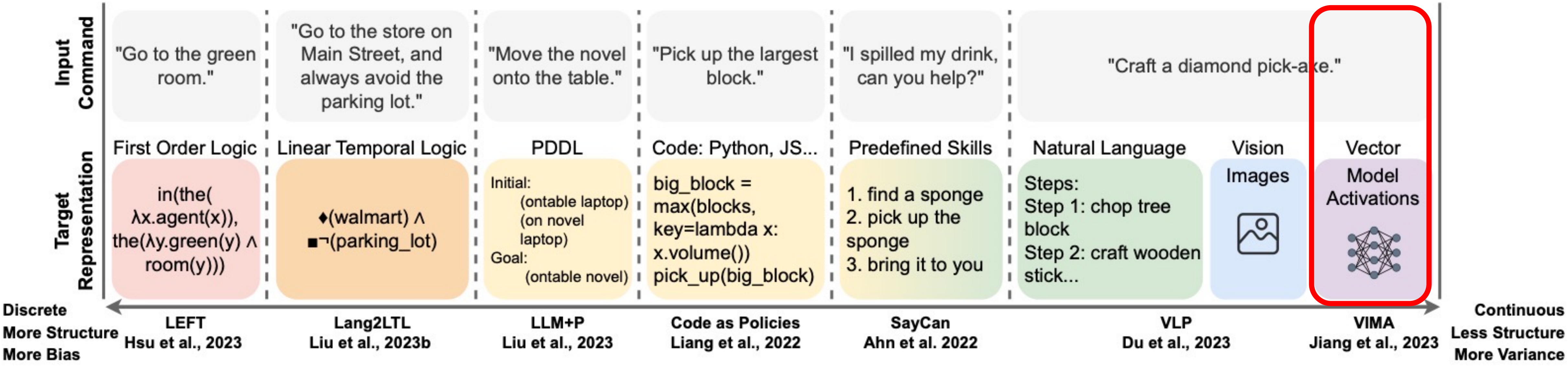
Grounding Language to Embeddings



Pros

- Adaptive

Grounding Language to Embeddings



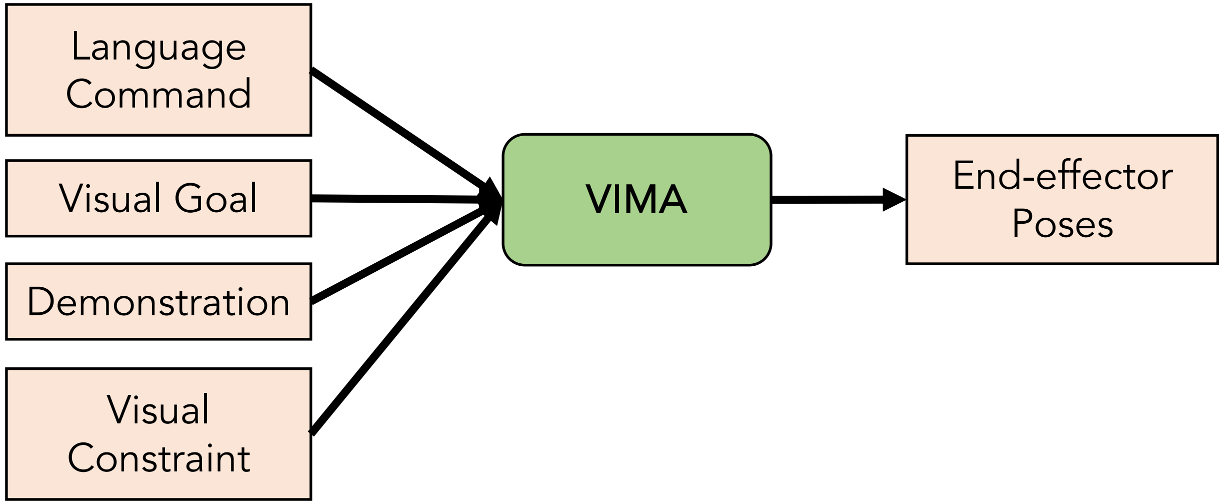
Pros

- Adaptive

Cons

- Large training set and compute
- Possibly incorrect actions

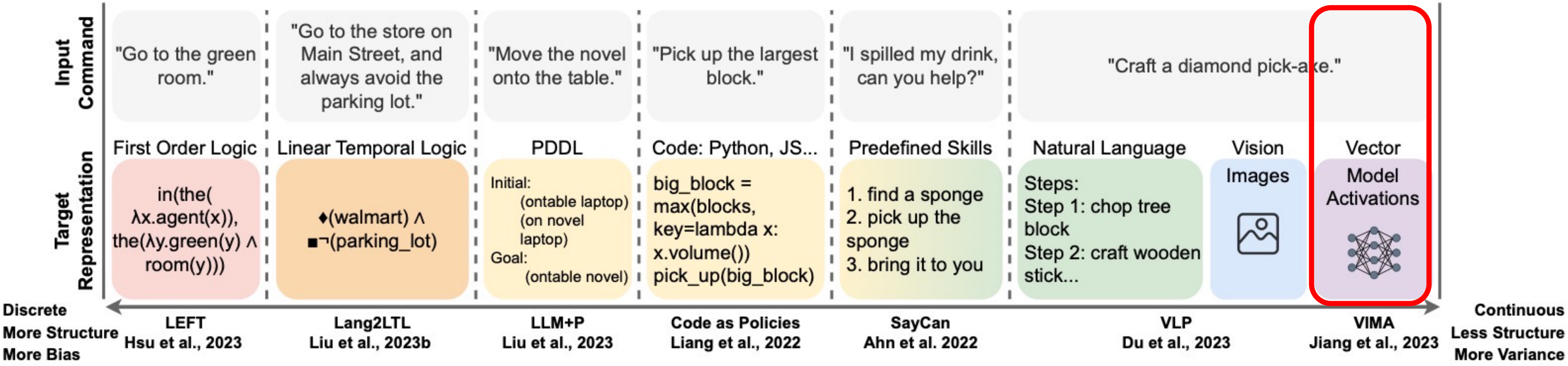
Grounding Language to Embeddings: VIMA



VIMA

- Tokenize multimodal input
- Transformer architecture
- Output end-effector poses

Grounding Language to Embeddings

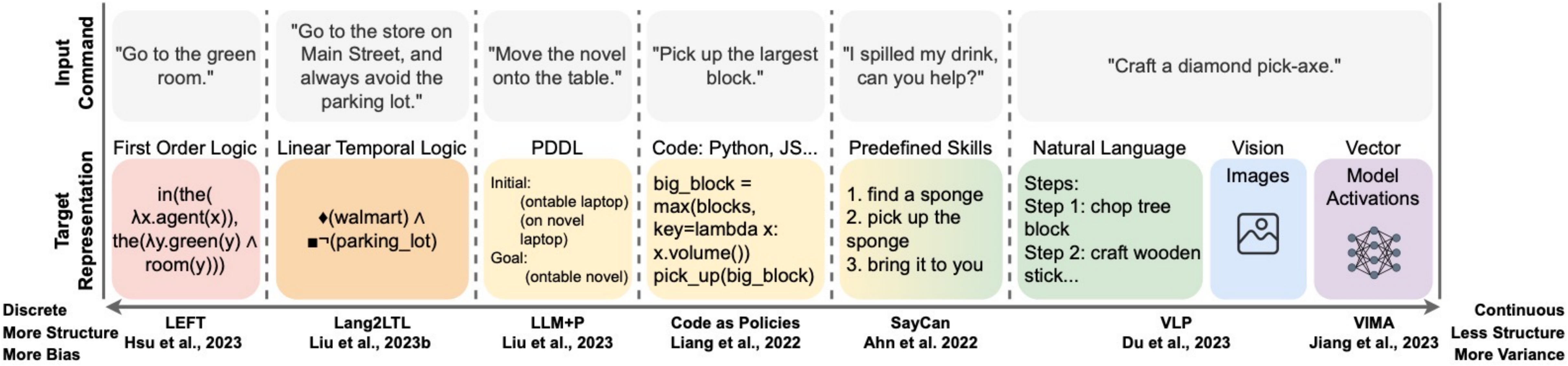


More Papers

- Octo: An Open-Source Generalist Robot Policy [Octo Model Team 2024]
  Open X-Embodiment: Robotic Learning Datasets and RT-X Models [Open X-Embodiment Collaboration 2024]
- RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control [Brohan et al. 2023]
- RT-1: Robotics Transformer for Real-World Control at Scale [Brohan et al. 2023]
- PaLM-E: an Embodied Multimodal Language Model [Driess et al. 2023]
- Vision-Language Foundation Models as Effective Robot Imitators [Li et al. 2023]
- GATO: A Generalist Agent [Reed et al. 2022]
- Perceiver-Actor: A Multi-Task Transformer for Robotic Manipulation [Shridhar et al. 2022]
- Video PreTraining (VPT): Learning to Act by Watching Unlabeled Online Videos [Baker et al. 2022]

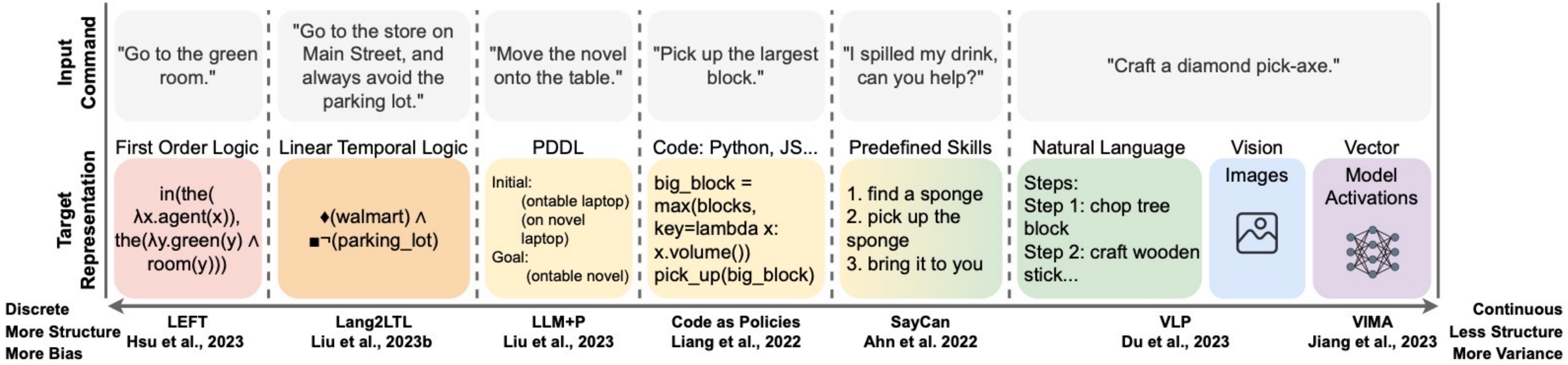
Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

Language Grounding for Robots



Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

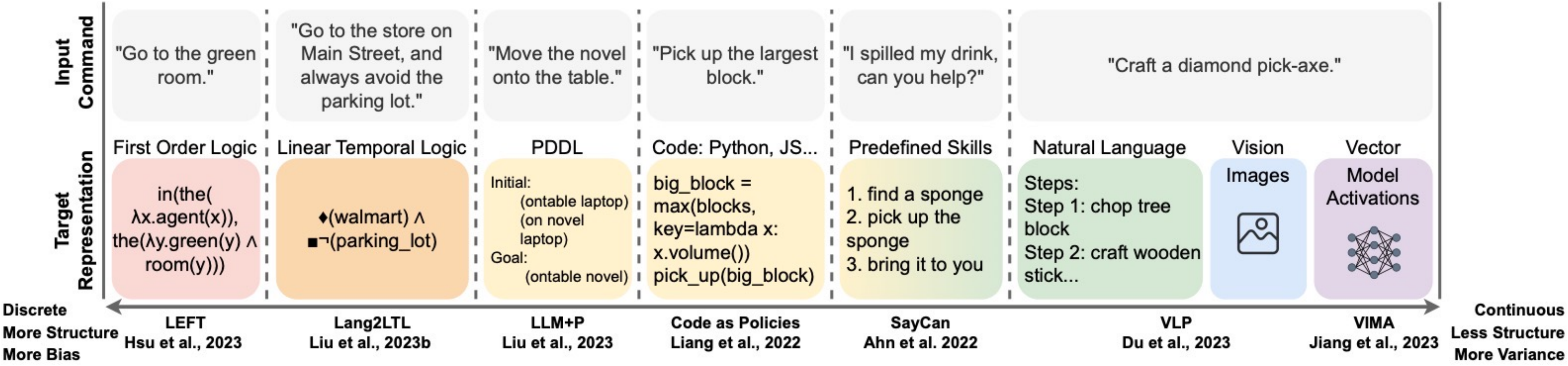
Language Grounding for Robots



Discrete Symbols

- Logic
- Planning domain definition language (PDDL)
- Code
- Descriptions of predefined skills

Language Grounding for Robots



Discrete Symbols

- Logic
- Planning domain definition language (PDDL)
- Code
- Descriptions of predefined skills

High-dimensional Embeddings

- Language and image subgoals
- Neural embeddings

Liu et al. IJCAI 2024 Survey Track. A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

Open Problems and Future Directions

Open Problems and Future Directions

- Neuro-symbolic Approach
    - POMDP and PDDL planners
    - Deep learning models with generalizable representations
    - E.g., Jointly learn symbols in the embedding space and skills

Open Problems and Future Directions

- **Neuro-symbolic Approach**
    - POMDP and PDDL planners
    - Deep learning models with generalizable representations
    - E.g., Jointly learn symbols in the embedding space and skills
- **Multimodal Dataset**
    - E.g., text, audio, RGB images, point clouds, voxels, videos, demonstrations
    - Semantically diverse

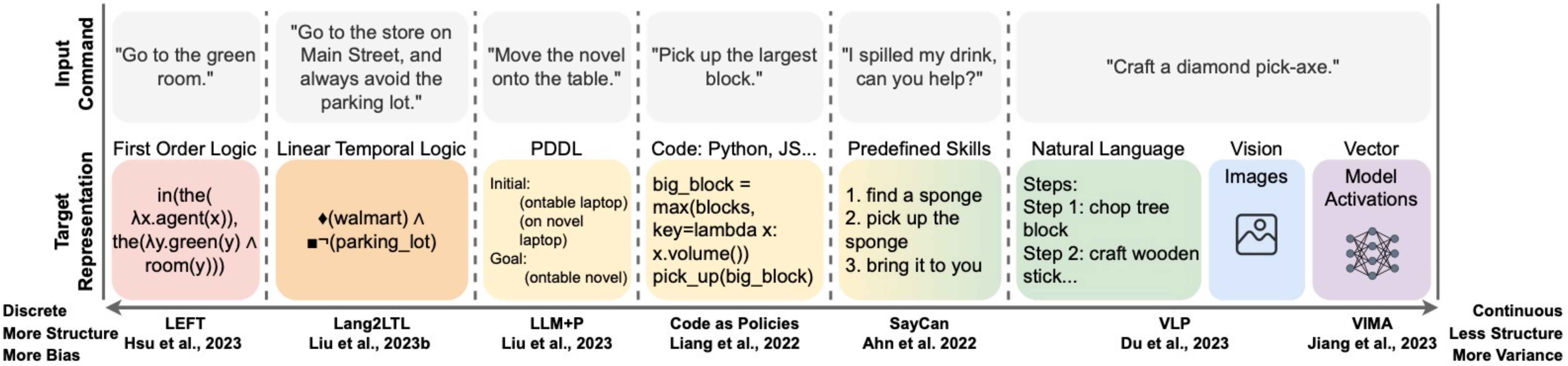
Open Problems and Future Directions

- **Neuro-symbolic Approach**
  - POMDP and PDDL planners
  - Deep learning models with generalizable representations
  - E.g., Jointly learn symbols in the embedding space and skills
- **Multimodal Dataset**
  - E.g., text, audio, RGB images, point clouds, voxels, videos, demonstrations
  - Semantically diverse
- **Modular Approach**
  - Existing robot modules
  - E.g., SLAM, motion planning and object detection

Open Problems and Future Directions

- Neuro-symbolic Approach
  - POMDP and PDDL planners
  - Deep learning models with generalizable representations
  - E.g., Jointly learn symbols in the embedding space and skills
- Multimodal Dataset
  - E.g., text, audio, RGB images, point clouds, voxels, videos, demonstrations
  - Semantically diverse
- Modular Approach
  - Existing robot modules
  - E.g., SLAM, motion planning and object detection
- Verification and Safety
  - Formal methods

Conclusion



A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings

https://arxiv.org/abs/2405.13245

Jason Xinyu Liu

https://jasonxyliu.github.io/

xinyu_liu@brown.edu

Poster Location:
E15