

# Chat with the Environment: Interactive Multimodal Perception Using Large Language Models



Matcha



**Xufeng Zhao**,



Mengdi Li,



Cornelius Weber,



Muhammad Burhan Hafez,



Stefan Wermter  
Knowledge Technology, University of Hamburg



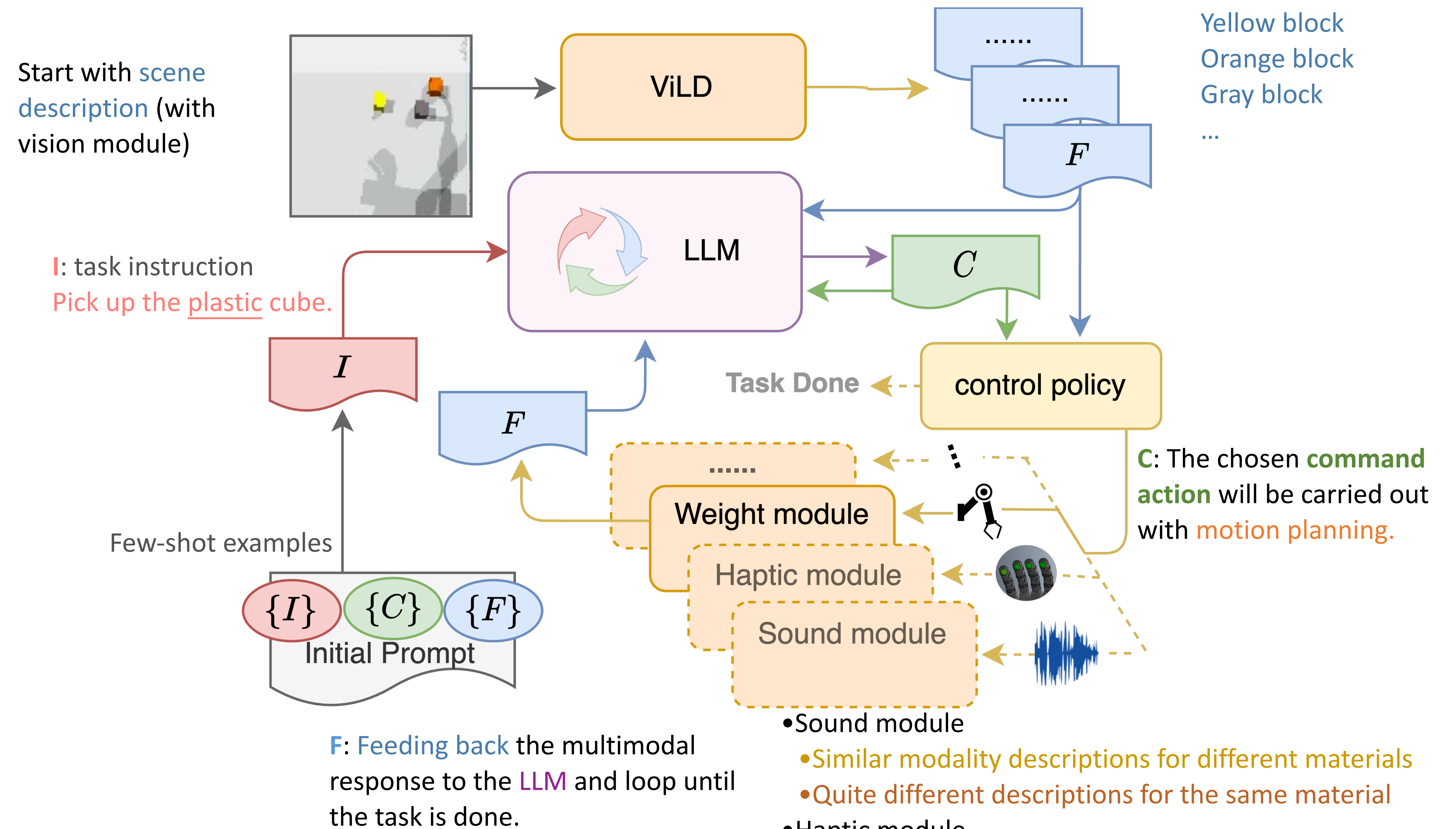
We propose the **Matcha framework**, comprising an LLM and multiple **multimodal modules**, enabling the robot to engage with its surroundings through high-level **LLM planning**.

(**M**ultimodal environment **c**hattering **a**gent)

## Robotic Perceptions

- **Passive perceptions**
- **Epistemic uncertainty**
- **Active perceptions**
- **Increased complexity**
- **Generalizability**
- **Robots with LLMs**
- **Causal reasoning ability with distilled human knowledge inside**
- **In-context learning ability with few-shot prompts**

## Matcha Architecture



LLM	Type of Description	Success Rate
text-ada-001	Indistinct	19.05%
	Distinct	28.57%
text-davinci-003	Indistinct	56.67%
	Distinct	90.57%

\*Random guess in principle: 33.33%

- NICOL robot
- Coppeliasim simulator
- LLM: OpenAI API text-davinci-003
- Works without any fine-tuning

