Robot Learning based on Long-Horizon Visual-Language-Action Models

Mentor: Zlatan Ajanović

E-mail: ajanovic.zlatan@gmail.com | Naučna oblast: AI and Robotics

Opis teme

The project addresses a central challenge in embodied intelligence: enabling robots to interpret language instructions, ground them in visual observations, and execute coherent multi-step behaviors over extended temporal horizons.
Recent advances in multimodal learning have led to strong progress in language-conditioned robotic control. However, long-horizon tasks remain challenging due to compounding errors, partial observability, delayed dependencies, and the need for robust subgoal reasoning, temporal abstraction, and recovery from failure. This internship focuses on learning-based approaches for building VLA systems that can support reliable sequential decision-making in complex robotic environments.

Zadaci i ciljevi

The internship may involve one or more of the following directions:

- Visual-language-action models for long-horizon manipulation or mobile manipulation tasks
- Language-conditioned policy learning from multimodal robotic data
- Sequence modeling of actions, observations, and task progress over extended horizons
- Hierarchical policy learning, skill composition, and subgoal prediction
- Memory-based architectures for temporally extended decision-making
- Comparative analysis of learning-based and planning-based methods for long-horizon robotics

Responsibilities

- Participation in a robotics team with research responsibilities.
- Design and implement visual-language-action models for robotic task execution
- Train models using multimodal datasets containing visual observations, actions, trajectories, and language annotations
- Investigate architectures for multimodal grounding, sequence prediction, memory, and hierarchical control
- Evaluate model performance in simulation and, where applicable, on robotic platforms
- Analyze generalization across tasks, environments, and instructions

- Study failure modes and improve robustness in long-horizon execution
- Document experiments and contribute to research reports or publications

Lista referenci

[1] J. Luijkx, R. Ma, Z. Ajanović, and J. Kober, "LLM-Guided Task- and Affordance-Level Exploration in Reinforcement Learning," in 2026 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2026. doi: 10.48550/arXiv.2509.16615.
[2] J. Luijkx, Z. Ajanovic, L. Ferranti, and J. Kober, "ASkDAgger: Active Skill-level Data Aggregation for Interactive Imitation Learning," Transactions on Machine Learning Research, Apr. 2025, Accessed: Aug. 07, 2025. [Online]. Available: https://openreview.net/forum?id=987Az9f8fT

Tražene vještine od studenata

- Strong programming skills in Python
- Experience with PyTorch or similar deep learning frameworks
- Background in machine learning, robot learning, or robotics
- Familiarity with one or more of the following (imitation learning or reinforcement learning; transformer-based sequence models; robotic manipulation, VLAs)
- Interest in long-horizon decision-making for robotic systems

Link za grafički element

https://llm-tale.github.io/