Pre-meeting for Development of LLM-driven GUI Agents

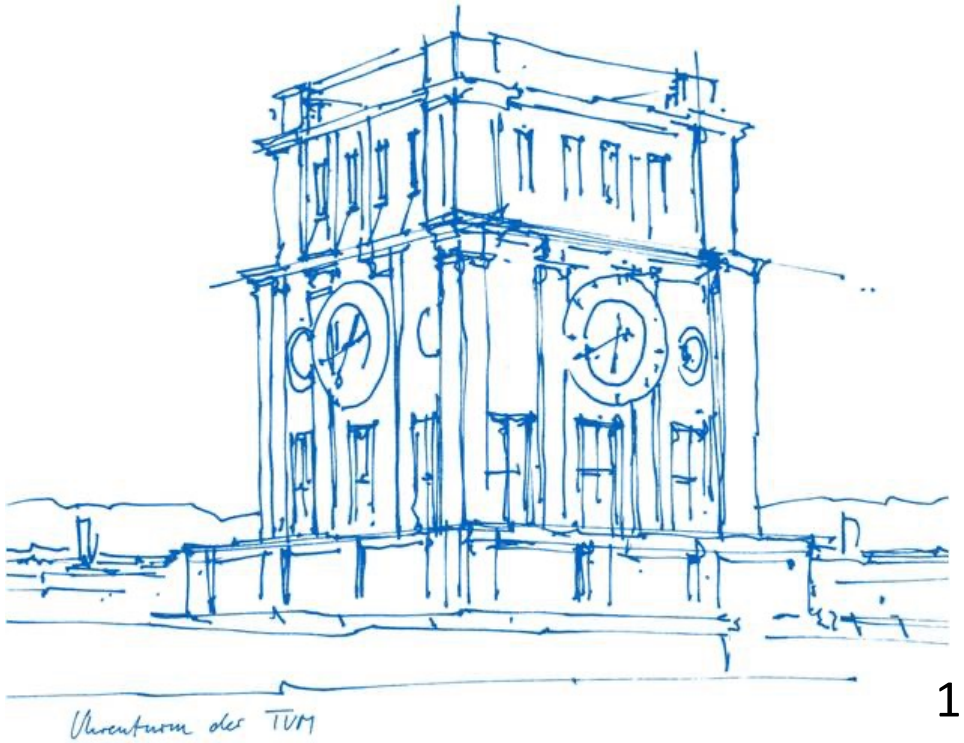
Dr. Shengcheng Yu

Chair of Software Engineering & AI

School of CIT, TUM

shengcheng.yu@tum.de

https://www.cs.cit.tum.de/en/seai/homepage/



Course Info

- **Pre-meeting time**: 10.07.2025
- **Deadline for preference list**: 20.07.2025
  - Basic information + Motivation statement
  - https://forms.gle/4C2xisLD4UehtuGn8
- **Dropping out deadline**: 19.10.2025 (one week after semester starting)
- **Upper student limit**: 12 (3*4groups)
- **Course time**: *expected* Mon 16:00 – 17:30

Context

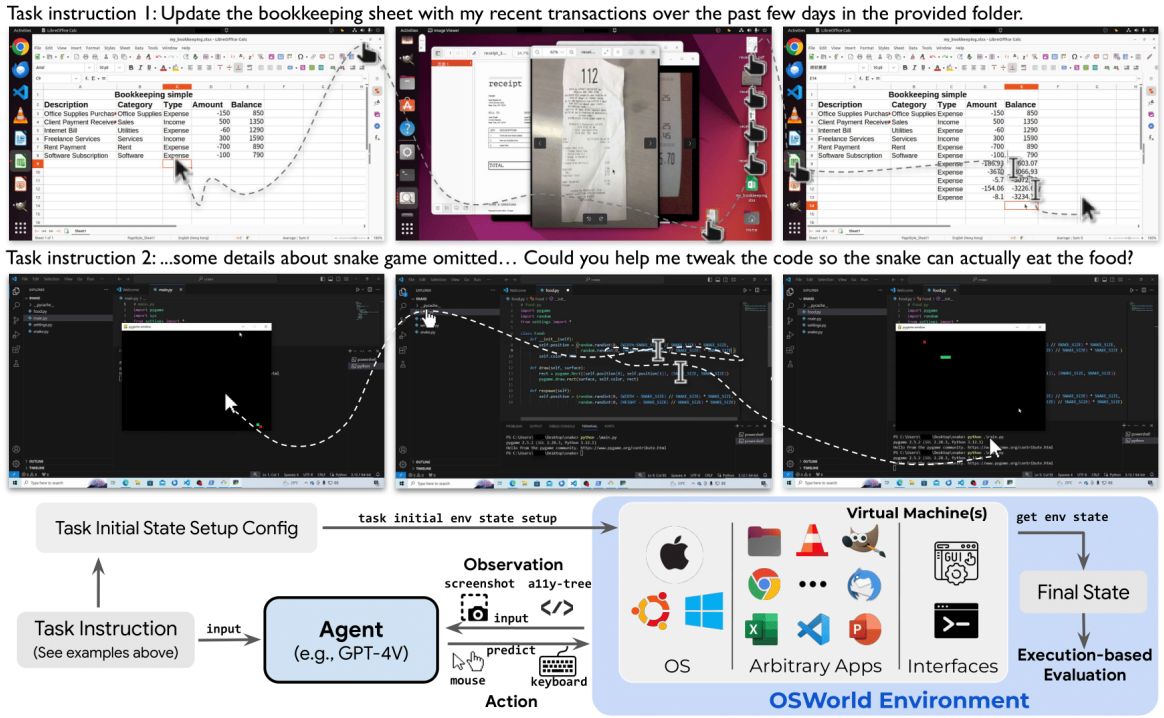
- Development of AI agents that autonomously interact with graphical user interfaces (GUI)
- Combination of:
  - Large Language Models (LLMs)
  - GUI Automation
  - Computer Vision
- Evaluation on standardized benchmarks

Content

- Learn about state-of-the-art LLM-driven GUI agents
- Implement an agent for a selected benchmark
- Evaluate and document the results
- Project Focus:
  - Implementation: Building agents that can autonomously interact with GUIs
  - Benchmark Performance: Meeting specific task criteria
  - Evaluation: Comparing against baseline metrics

Example Benchmark: OSWorld

- Real-world GUI tasks across platforms
- 369 standardized tasks (web, desktop, file operations)
- Current SOTA: 38.1% success rate
- Provides reproducible evaluation metrics
- See: https://os-world.github.io



Structure

- Foundation Phase
  - Introduction to LLM-driven GUI agents
  - Overview of relevant technologies and frameworks
  - Formation of groups (2~3 people) and selection of benchmark
- Research & Prototype Phase
  - Working on prototypes
  - Weekly meetings with assigned tutor
  - Midterm presentation (10%) - progress check
- Implementation Phase
  - Complete implementation (50%)
  - Benchmark evaluation and documentation (30%)
  - Final presentation (10%) - demonstrating achievements

Expectation

- Working Prototype
  - Demonstrable on real examples
  - Reproducible results
  - Well-structured implementation
- Documentation
  - Clear code structure
  - Key methods explained
  - Setup and usage instructions
- Presentation
  - Midterm: Show clear progress and planning
  - Final: Demonstration of achievements

Additional Information

- All implementations should use open-source LLMs
- Computing Resources
  - Three NVIDIA RTX 4090 GPUs available for student use
  - Dedicated for running open-source LLMs
  - Suitable for models like:
    - Llama variants
    - Mistral
    - DeepSeek R1

Q & A

Contact

Dr. Shengcheng Yu

shengcheng.yu@tum.de

https://www.cs.cit.tum.de/en/seai/homepage/

https://www.seysc.com