Development of LLM-driven GUI Agents – Pre-Meeting

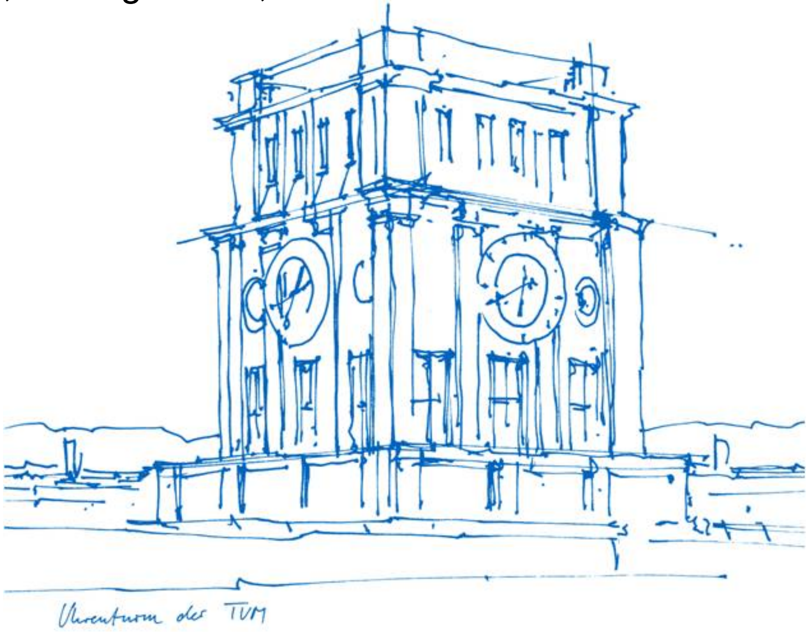
Prof. Dr. Chunyang Chen, Dr. Shencheng Yu, Sidong Feng, Ludwig Felder, Shen Hu

Technical University of Munich

School of Computation, Information and Technology (CIT)

Chair of Software Engineering & AI

11.02.2025



Context

Development of AI agents that autonomously interact with graphical user interfaces

Combination of:

- Large Language Models (LLMs)
- GUI Automation
- (Computer Vision)

Evaluation on standardized benchmarks

Content

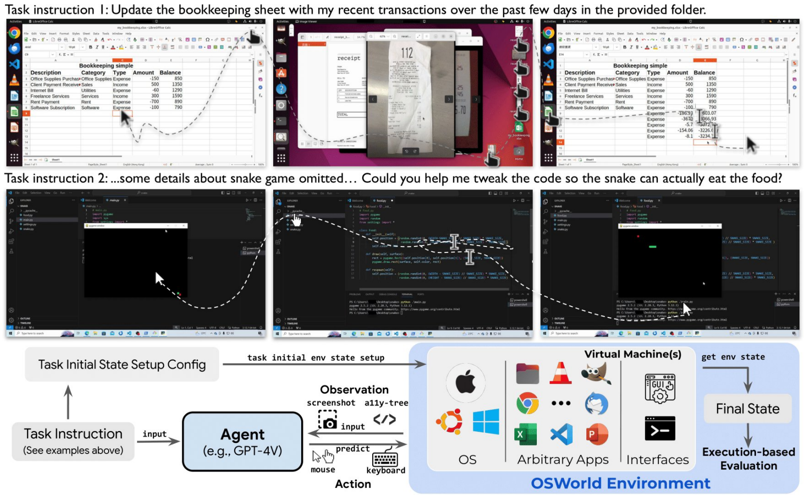
1. Learn about state-of-the-art LLM-driven GUI agents
2. Implement an agent for a selected benchmark
3. Evaluate and document the results

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



Task instruction 1: Update the bookkeeping sheet with my recent transactions over the past few days in the provided folder.

Task instruction 2: ...some details about snake game omitted… Could you help me tweak the code so the snake can actually eat the food?

Structure

- Week 1-3: Foundation Phase
  - Introduction to LLM-driven GUI agents
  - Overview of relevant technologies and frameworks
  - Formation of groups (2 people) and selection of benchmark
- Week 4-7: Research & Prototype Phase
  - Working on prototypes
  - Weekly meetings with assigned tutor
  - Midterm presentation (10%) - progress check
- Week 8-13: 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
- Presentations
  - 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

Question & Answer

- Main Tutors
  - Shen Hu, shen.hu@tum.de
  - Ludwig Felder, ludwig.felder@tum.de

- To be preferred in the matching, please fill our application form (deadline 18.02, 23:59): https://forms.gle/iabNFWLeM2ecGYTY6