Exposé - Automating administration panels with AI Agents

Felix Beer

fbeer.mmt-b2022@fh-salzburg.ac.at

FH Salzburg

26. Jänner 2025

Motivation

Web administration interfaces of modern software systems are crucial for managing and maintaining the underlying data. These interfaces provide administrators with the necessary tools to create content, configure settings, troubleshoot issues and get insight into key performance metrics. The design and usability of these interfaces can significantly impact the efficiency and effectiveness of system management tasks.

AI agents, which in short are LLMs (Large Language Models) that can execute a predefined set of functions, have the potential to simplify or even automate many of the routine tasks performed by administrators.

The concept of agents has been described in many different ways. The ReAct framework, introduced by Yao et al. 2023, describes a method to include reasoning traces based on task results. This allows the model to dynamically plan, execute, and adjust actions based on real-time feedback.

Li et al. 2023 introduced API-Bank, which benchmarked performance of different commonly used models across 73 APIs. The result showed that GPT-4 (OpenAI et al. 2024) managed to execute 70% of Plan/Retrieve/Call tasks correctly.

Research Question

Can AI agents replace traditionally complex administration interfaces in content management systems by simplifying user interaction?

Concept

The practical part of this thesis will focus on the development of an AI agent that can be used via a chat interface. This AI agent will be built to integrate with an existing content management

system, which allows administrators to manage resources such as blog posts, shop products, or online courses. Additionally, the agent should be implemented using the newly released Model Context Protocol [1], which would allow usage in a variety of different clients. To find out if the agent can actually simplify user interaction compared to the traditional interface, a user study will be conducted. The study would then take a set of features, such as:

- Updating a product's price
- Changing a blog post's hero image
- Creating a new online-course structure (name, chapters, lessons)

The users will be asked to perform these tasks using both the traditional web interface and the AI agent. To keep the bias at a minimum, the tests will be performed in a counterbalanced order, so that half of the participants will start with the traditional interface and the other half with the AI agent. By measuring the time it takes to complete every task, but also if the user was able to successfully complete the task at all, it is possible to quantitatively compare if the AI agent is more efficient than the traditional interface. Furthermore, it would make sense to include a survey like the well-established System Usability Scale (SUS) (Brooke 1995) to get an understanding of the user's satisfaction.

Outline

This is a rough outline that will be refined during the research process.

1. Introduction
    (a) Background and Motivation
    (b) Problem Statement
    (c) Methodology Overview
2. Theoretical Background
    (a) AI Agents
        i. Large Language Models (LLMs)
        ii. Capabilities and Limitations
            A. Function Calling
            B. Hallucination
            C. Reasoning
        iii. Model Context Protocol (MCP)
    (b) API Specification

1. https://www.anthropic.com/news/model-context-protocol

i. OpenAPI
ii. GraphQL
iii. Internal Function Calling

3. Concept and design
   (a) Integration methods with existing content management system
   (b) Proposed features of the AI agents
4. Implementation
5. User Study
   (a) Evaluation Metrics
      i. Time Savings
      ii. Task Completion
      iii. User Satisfaction
6. Discussion
7. Conclusions

References

Brooke, John. 1995. „SUS - A quick and dirty usability scale“ (November).

Li, Minghao, Yingxiu Zhao, Bowen Yu, Feifan Song, Hangyu Li, Haiyang Yu, Zhoujun Li, Fei Huang, and Yongbin Li. 2023. „API-Bank: A Comprehensive Benchmark for Tool-Augmented LLMs.“ In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, edited by Houda Bouamor, Juan Pino, and Kalika Bali, 3102–3116. EMNLP 2023. Singapore: Association for Computational Linguistics, December. Accessed January 24, 2025. https://doi.org/10.18653/v1/2023.emnlp-main.187. https://aclanthology.org/2023.emnlp-main.187/.

Lu, Junting, Zhiyang Zhang, Fangkai Yang, Jue Zhang, Lu Wang, Chao Du, Qingwei Lin, Saravan Rajmohan, Dongmei Zhang, and Qi Zhang. 2024. *Turn Every Application into an Agent: Towards Efficient Human-Agent-Computer Interaction with API-First LLM-Based Agents*, arXiv:2409.17140, September 25, 2024. Accessed January 24, 2025. https://doi.org/10.48550/arXiv.2409.17140. arXiv: 2409.17140[cs]. http://arxiv.org/abs/2409.17140.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, et al. 2024. *GPT-4 Technical Report*, arXiv:2303.08774, March 4, 2024. Accessed January 24, 2025. https://doi.org/10.48550/arXiv.2303.08774. arXiv: 2303.08774[cs]. http://arxiv.org/abs/2303.08774.

Ruan, Jingqing, YiHong Chen, Bin Zhang, Zhiwei Xu, Tianpeng Bao, Du Guo Qing, Shi Shiwei, et al. 2023. „TPTU: Task Planning and Tool Usage of Large Language Model-based AI Agents.“ NeurIPS 2023 Foundation Models for Decision Making Workshop. November 8, 2023. Accessed January 24, 2025. https://openreview.net/forum?id=GrkgKtOjaH.

Schick, Timo, Jane Dwivedi-Yu, Roberto Dessi, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. 2023. „Toolformer: Language Models Can Teach Themselves to Use Tools.“ *Advances in Neural Information Processing Systems* 36 (December 15, 2023): 68539–68551. Accessed January 24, 2025. https://proceedings.neurips.cc/paper_files/paper/2023/hash/d842425e4bf79ba039352da0f658a906-Abstract-Conference.html.

Wang, Lei, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, et al. 2024. „A survey on large language model based autonomous agents.“ *Frontiers of Computer Science* 18, no. 6 (March 22, 2024): 186345. ISSN: 2095-2236, accessed January 24, 2025. https://doi.org/10.1007/s11704-024-40231-1. https://doi.org/10.1007/s11704-024-40231-1.

Yao, Shunyu, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. *ReAct: Synergizing Reasoning and Acting in Language Models,* arXiv:2210.03629, March 10, 2023. Accessed January 18, 2025. https://doi.org/10.48550/arXiv.2210.03629. arXiv: 2210.03629[cs]. http://arxiv.org/abs/2210.03629.

Schedule

- January 26th - Submission of exposé
- February - Researching and testing suitable models and technologies
- February 17th - Application for supervision
- Early March - First meeting with supervisor
- March - Writing theoretical part and agent implementation
- Early April - User study and analysis
- Mid-April - Second meeting with supervisor
- Late April - Finalizing practical part
- May 5th - Submission of final thesis

Supervisor

Radomir Dinic, MSc