A Dialog-Based Multi-Agent System for Context-Aware Financial Analysis in SMEs

Miriam Velasco[1], Simone Braun[2]

[1] bimetrics GmbH
miriam.velasco@bimetrics.de [2] Offenburg University of Applied Sciences **simone.braun@hs-offenburg.de**

Abstract. This paper presents a dialog-based AI financial assistant designed to support small and medium-sized enterprises (SMEs) in interpreting their financial indicators. The system combines a modular multi-agent architecture with context engineering techniques (explicit agent roles, controlled function calls, and structured hand-offs) to analyze and explain financial data in natural language. A prototypical cash-flow analysis, embedded within a production platform, demonstrates how the assistant enables interactive communication of financial information, thereby improving access to managerial and financial knowledge for non-experts.

Keywords: Dialog-based AI; Multi-Agent Systems; Context Engineering; LLMs; Agentic RAG; Financial Analysis; Digital Financial Literacy; SMEs

1 Introduction

Small and medium-sized enterprises (SMEs) often lack the in-house expertise to interpret their financial metrics independently. This gap in financial literacy hinders informed decision-making and risk management [1]. Existing analytics dashboards visualize data but seldom provide accessible explanations tailored to non-experts, thereby limiting essential digital financial literacy [2].

This work introduces a dialog-based AI financial assistant that integrates large language models (LLMs) into a modular multi-agent system for grounded, natural-language analysis and explanation. Unlike generic chatbots, the system operates directly on structured financial data and provides transparent, contextualized insights. The main contribution is the design and evaluation of a scalable multi-agent architecture that combines retrieval-augmented generation, controlled function calling, and deterministic context transfer – enabling reliable, explainable financial analysis for SMEs [1-3].

2 Approach

The financial assistant is implemented as a multi-agent system integrating LLMs with targeted context engineering [3]. Specialized agents perform distinct tasks – such as intent recognition, function execution, and natural-language response generation – while maintaining a strict separation of concerns between data access, semantic processing, and output generation [4,5]. This modular design enhances scalability and reduces hallucinations by ensuring each agent operates within a well-defined context [6,7].

Figure 1 illustrates the architecture's core components, which transform structured financial information into dialog-based explanations. The system is embedded in an operational financial platform, connected via API to real booking and document data. A cash-flow analysis was implemented as proof of concept to assess the assistant's ability to analyze and explain financial metrics interactively.

To ensure transparency and reliability in sensitive financial contexts, the system employs controlled function calling restricting LLM access to validated data from a preceding analytics layer [3]. Deterministic temperature settings (e.g., temperature = 0.3) ensure consistent tool selection, while structured handovers of contextual data between agents preserve semantic coherence across the dialogue [8]. These mechanisms form the foundation of the assistant's context-management strategy, key to reproducible, trustworthy analyses.

A user-centered evaluation (n = 10) confirmed the approach's practical applicability: all participants stated they would use the cash-flow analysis regularly, 88% rated the explanations as easy to understand, and 83% considered the results trustworthy. These findings underline that consistent context handling and transparent system logic are as important to user trust as analytical performance itself.

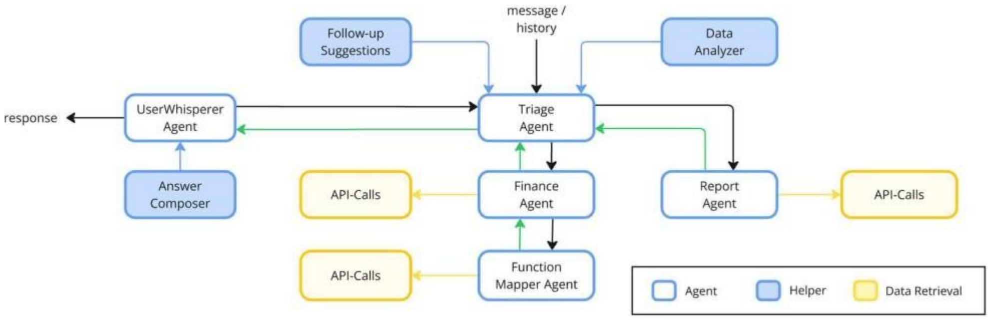

Fig. 1. Architecture of the multi-agent system underlying the financial assistant.

3 Conclusion

This paper demonstrates how combining multi-agent architectures with context-controlled LLM integration enables reliable, dialog-based financial analysis for SMEs [3-5]. The system translates structured financial data into natural-language explanations, allowing non-experts to explore business metrics interactively. The development process revealed that the architecture's effectiveness relies on clearly defined interfaces, controlled function calling, and structured context transfer between agents, which together reduce hallucinations, maintain semantic coherence, and foster user trust [3].

A user-centered evaluation confirmed that participants found the explanations highly comprehensible and the results reliable, indicating that transparent reasoning and consistent context handling are as important to user acceptance as analytical performance. The system securely integrates with real financial data and complies with privacy requirements [3-5].

While currently focused on cash-flow analysis with a small user sample, future work will extend the approach to other financial domains and a larger user base. Overall, the approach provides a reproducible blueprint for deploying LLM-based assistants in domain-specific decision-support workflows for SMEs, with significant potential to improve digital financial literacy and decision quality [1,2].

References

1. Hammer, T. (2024). Implications of Financial Literacy on Entrepreneurship in Germany. [Doctoral thesis] Universidad Católica de Murcia. URL https://repositorio.ucam.edu/bitstream/handle/10952/7984/Tesis.pdf
2. Imjai, N., Meesook, K., Somwethee, P., Usman, B., & Aujirapongpan, S. (2025). Exploring the impact of digital financial literacy to effective financial planning and control: Perspectives on competitiveness of Thai micropreneurs. *Social Sciences & Humanities Open*, *11*, 101307. doi:10.1016/j.ssaho.2025.101307
3. Mei, L., Yao, J., Ge, Y., Wang, Y., Bi, B., Cai, Y., … & Liu, S. (2025). A Survey of Context Engineering for Large Language Models. *arXiv preprint arXiv:2507.13334*.
4. Han, S., Zhang, Q., Yao, Y., Jin, W., Xu, Z., & He, C. (2024). LLM Multi-Agent Systems: Challenges and Open Problems. *arXiv preprint arXiv:2402.03578*.
5. Li, X., Wang, S., Zeng, S., Wu, Y., & Yang, Y. (2024). A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges. *Vicinagearth, 1*(9), 1-43. doi:10.1007/s44336-024-00009-2
6. Lu, Y., Aleta, A., Du, C., Shi, L., & Moreno, Y. (2024). LLMs and generative agent-based models for complex systems research. *Physics of Life Reviews, 51*, 283-293.
7. Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., ... & Li, Y. (2024). Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanities and Social Sciences Communications, 11*(1), 1-24.
8. Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2019). The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751.*