DOI:10.17951/h.2026.60.2.27-42

ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN – POLONIA

VOL. LX, 2 SECTIO H 2026

MARIUSZ HOFMAN

mariusz.hofman@mail.umcs.pl
Maria Curie-Skłodowska University. Faculty of Economics
pl. Marii Curie-Skłodowskiej 5, 20-031 Lublin, Poland
ORCID ID: https://orcid.org/0000-0002-6866-3906

ANDRZEJ SOBCZAK

sobczak@sgh.waw.pl
Warsaw School of Economics. Institute of Information Systems and Digital Economy
al. Niepodległości 162, 02-554 Warsaw, Poland
ORCID ID: https://orcid.org/0000-0002-1271-1251

LLM-Based Agents: Can They Transform Organizational Realities? An Experimental Verification Attempt

Keywords: AI agents; LLM-based AI agents; AI agents in management

JEL: M15; O33

How to quote this paper: Hofman, M., & Sobczak, A. (2026). LLM-Based Agents: Can They Transform Organizational Realities? An Experimental Verification Attempt. *Annales Universitatis Mariae Curie-Skłodowska, sectio H – Oeconomia*, *60*(2), 27–42.

Abstract

Theoretical background: Large language models (LLMs) have transformed artificial intelligence (AI) and are increasingly considered for use in organizational management. However, they have limitations, especially in generalization and performance on topics they were not trained on. Despite these constraints, LLMs show potential to support multiple management-related activities, particularly when deployed as autonomous or collaborative agents.

Purpose of the article: The article aims to explain how LLMs can be applied to organizational management despite their current limitations. It focuses on the emerging concept of LLM-based autonomous agents

that can reason, plan, and act in an environment. It also highlights the value of multi-agent collaboration coordinated by an orchestrator.

Research methods: The paper primarily takes a conceptual and exploratory approach by synthesizing existing ideas and emerging research on LLMs in management contexts. It discusses potential architectures, such as individual agents and orchestrated multi-agent systems, rather than reporting a controlled experiment. The article identifies key organizational areas where these approaches could be applied.

Main findings: LLM-based agents can potentially perform complex organizational tasks by combining reasoning, planning, and action capabilities. Collaborative multi-agent solutions managed by an orchestrator may improve effectiveness compared to single-agent setups. The article concludes that more empirical research is needed to validate these solutions in real organizations and to assess their effects on efficiency and workplace adaptation.

Introduction

In recent years, the emergence of large language models (LLMs) has revolutionized. LLMs show great potential but also have their weaknesses (Händler, 2023). Their answers might seem correct but may be incorrect or hallucinated, which can negatively affect trust in the information provided by these models (Ji et al., 2023). In addition, LLMs lack knowledge of the specifics of a particular organization, the environment in which company operates as well as the nature of the problem the model helps to solve. The presence of these limitations has caused the use of these models in enterprises to be inadequate compared to their potential (Korzynski et al., 2023). It has also led to the development of more sophisticated and flexible AI architectures, including AI agent solutions.

LLMs can be used as a mechanism that allows AI agents to operate effectively in a variety of settings (Park et al., 2023). LLMs gives such agents the ability to reason, plan and act within the environment in which they operate (Kaddour et al., 2023; Padgham & Winikoff, 2005). This means that agents, whether embodied or not, can be used to perform even complex and demanding tasks or processes in the organizational domain. LLM-powered agents can achieve user-prompted goals by employing a proper strategy and breaking it down into smaller manageable tasks (Xi et al., 2023). More advanced solutions anticipate the collaboration of multiple agents to solve complex problems by leveraging the cognitive synergy of multiple autonomous agents (Hong et al., 2023; Park et al., 2023). LLM-based agents can also acquire interactive capabilities by learning from feedback, and self-evolving (Schick et al., 2023). This results in LLM-based agents having the potential to perform even complex tasks or processes that are currently within the domain of humans.

The purpose of this article is to show and verify the applicability of LLM-based agents in companies. Different tasks may be delegated to these agents, allowing them to act on behalf of employees. In this scenario, humans do not have full control over tasks and process outputs. Rather than being adversaries, workers and LLM agents combine their complementary strengths, enabling mutual learning and multiplying

their capabilities (Raisch & Krakowski, 2021). Thus, verifying the ability of these agents to automate or augment tasks and processes is crucial for determining their place and role in future enterprises.

Literature review

The concept of an agent has a long history that has been explored and interpreted in many fields, including organizational and management theory (Eisenhardt, 1989; Lan & Heracleous, 2010). AI agents are understood as artificial entities that are capable of perceiving their environment, making decisions, and then taking actions (Wooldridge & Jennings, 1995; Yao et al., 2023). Thus, user-prompted tasks can be performed by AI agents, which are equipped with dedicated roles and LLM reasoning capabilities. Taking the right action requires those agents to have the ability to plan. By planning, agents generate a sequence of steps to complete the task and achieve goals (often exploring multiple paths) (Wei et al., 2023). Moreover, agents can employ introspection to modify their plans, ensuring they align better with real-world circumstances, leading to adaptive and successful task execution (Xi et al., 2023). Action, in turn, requires LLM agents to be able to use tools and contextual resources, such as data sets or other models (Händler, 2023; Qin & Hu, 2023). These agents can use available tools or build new ones when needed (Hussein et al., 2017). In the course of their operation, they collect and store in short-term memory necessary information on an ongoing basis. They may also have long-term memory, allowing them to accumulate knowledge and experience over a longer period (Zhu et al., 2023).

Single-agent solutions use specialised agents to perform specific task. These solutions focus on the internal mechanisms of the agent, including the effectiveness of tool usage and their ability to interact with the environment. The first framework that provided the capability to build solutions using LLM-based agents was LangChain (Chase, 2023). Multi-agent solutions are based on the cooperation of many agents to deal with complex goals (Park et al., 2023). In such solutions, agents assume distinct roles encompassing agent characteristics, behaviours and capabilities (Guo et al., 2024). Multiple LLM-powered agents jointly perform tasks, each equipped with unique strategies and engaged in communication with one another (Xi et al., 2023). These agents may cooperate or compete with each other, while information exchange between them can occur in a centralized or decentralized manner. An important advantage of multi-agent solutions is that agents can obtain feedback not only from the user or the environment, but also from other agents (Wang et al., 2023). This enables agents to adjust their profiles or goals, rather than just learning from historical interactions (Guo et al., 2024). It also allows agents to acquire new capabilities and utilize new tools (Zhang et al., 2023). The following multi-agent frameworks are currently available: CAMEL (Li et al., 2023) and AutoGen (Wu et al., 2023). The latter is currently the most widely used due to its versatility and flexibility (Guo et al., 2024).

Research methods

The article is conceptual in nature and is an attempt to determine the potential of using LLM-based agents in organizational settings. The approach used involves a combination of research methods, including systematic literature review (SLR), analysis of framework documentation, and quasi-experiments. The research work began with an analysis of the literature on the subject. Due to the novelty of the issue and the rapidly changing state of knowledge in this area, the "arXiv" database was used. This repository provides access to the latest scientific articles and preprints, making it possible to access the most up-to-date knowledge at any given time. The search criteria included the following two phrases: LLM-based agents and LLM-based multi-agents. The research procedure adopted was to find 25 recent articles addressing the issue of LLM-based agents and their subsequent analysis using the Latent Dirichlet Allocation (LDA) (Mo et al., 2015). This made it possible to determine the existing state of knowledge in this field and identify the most frequently addressed research topics in the literature (Snyder, 2019).

The next stage was an in-depth analysis of the LangChain (https://www.langchain.com/agents), CrewAI (https://www.crewai.io), and AutoGen (https://microsoft.github.io/autogen) frameworks. The focus was on these frameworks because they enable the design and commissioning of a wide range of solutions using LLM agents. They are complex, open-source, and currently the most widely used libraries. Analyzing their documentation allowed us to define the main LLM-based agent types, understand how they work, and indicate the potential for agent customization (e.g., prompt engineering, equipping them with memory, and predefined or custom tools). The analysis also involved assessing the usefulness of these frameworks for building multi-agent solutions, in which multiple agents work together to perform user-prompted tasks. Analyzing the documentation of the libraries mentioned above also allowed us to identify the limitations of LLM-based agents and indicate their potential applications in organizations.

The final stage of the research was a series of quasi-experiments. Conducting the experiments allowed for comparing the methods and accuracy of selected types of agents in conditions similar to real-world scenarios (Ross & Morrison, 2004). The experiments involved designing and implementing single-agent and multi-agent solutions within the previously described frameworks and evaluating them. In the case of single-agent solutions, the experiments allowed us to determine how specific types of agents operate, whether they correctly use various data sources and entrusted tools, and thus verify their usefulness for performing specialized tasks currently carried out by humans in enterprises. For multi-agent solutions, the conducted experiments allowed us to verify the usefulness of LLM-based agents for team problem solving (Ma et al., 2024). A key aspect of the experiments was the evaluation of the outputs generated by the agents. For single-agent solutions, the focus was on utility, as the effectiveness of an agent in properly executing assigned tasks is the key aspect of

evaluation in this case (Wang et al., 2023). For multi-agent solutions, the assessment focused on agents' sociability, as the ability for team problem solving is the most important aspect in evaluating the agents' work results in this scenario (Xi et al., 2023). The evaluation of agents' outputs had both subjective and objective components to consider both task completion and the quality of the underlying processes (Wang et al., 2023). The experiments were conducted in April 2004. The results of the experiments allowed us to determine the usefulness of LLM agent solutions and identify potential weaknesses of these solutions in the enterprise domain. The quasi-experiments were carried out using the Python programming language, specialized frameworks (described above), the "Google Colab" software, and the "OpenAI" gpt-3.5-turbo model.

Results

SLR results

The results of the SLR are included in the Appendix. Between September and December 2023, 13 articles on LLM-based agents were posted on arXiv, while between January and February 2024, a further 12 were posted. This clearly indicates that the topic of AI agents is rapidly developing and becoming an important part of artificial intelligence research. Content analysis of the retrieved articles was carried out using the LDA method. Its use allows for the discovery of key themes based on the words that appear in these documents (Blei et al., 2003). In LDA, soft clustering is performed, meaning a sample word can belong to several topics simultaneously. This makes it possible to detect semantic structures that are overlooked when solely reading the content of the articles (Mo et al., 2015). To conduct the LDA analysis, ten key research themes were identified. Each was described by five keywords and assigned weights. The results are presented in Table 1.

Table 1. Results of LDA analysis

Topic	Keywords and their weights				
1	0.021 : agents	0.014 : agent	0.007 : llm	0.007 : language	0.005 : dylan
2	0.001 : agents	0.001 : agent	0.001 : language	0.001 : models	0.000 : llm
3	0.011 : personality	0.008 : models	0.006 : traits	0.006 : agents	0.006 : others
4	0.011 : model	0.009 : date	0.008 : task	0.008 : agents	0.007 : language
5	0.021 : agents	0.015 : agent	0.012 : language	0.008 : models	0.007 : learning
6	0.012 : agent	0.010 : agents	0.008 : code	0.007 : autogen	0.006 : llm
7	0.018 : agents	0.010 : social	0.010 : agent	0.009 : llm	0.007 : llms
8	0.022 : agents	0.009 : agent	0.009 : language	0.008 : models	0.005 : memory
9	0.012 : agents	0.011 : opinion	0.009 : agent	0.006 :xyz	0.005 : bias
10	0.009 : llm	0.009 : language	0.008 : agent	0.008 : agents	0.007 : text

Source: Authors' own study.

The keywords "agent" and "agents" are present in almost every extracted topic, indicating a strong focus of the literature on the concept of AI agents itself. It is also apparent that the analyzed articles treat the functioning of single agents and multiple agents as highly interrelated phenomena.

The first topic focuses on the interaction of agents with LLMs. The appearance of the word "Dylan" (Dynamic LLM-Agent Network) suggests considering these issues in the context of different forms of agent interaction (Liu et al., 2023). The second topic is of low relevance due to the low weights assigned to its keywords. It captures general or less clearly defined issues related to LLM-based agents, natural language, and large language models. The third topic focuses on the attribution of specific personality and traits to agents. An agent's specific profile determines its behaviour and influences the actions it takes. The inclusion of the word "others" suggests an interest in the nature of relationships that can occur between agents with different profiles, as well as between agents with different personalities and users. The fourth and fifth topic point to the importance of data in the process of creating language models (training or fine-tuning them), which are then used by agents to perform specific tasks. The sixth topic explores the ability of LLM agents to automatically generate, run, and test code. The word "autogen," which appears within this topic, refers to a specific solution that enables such applications of AI agents (Wu et al., 2023). The seventh topic explores the social aspects of the functioning of LLM agents, including their ability to mimic human behaviour and imitate social relationships. This topic also addresses the impact of LLM agents on society. The eighth identified topic highlights the role of memory in the functioning of LLM agents, enabling them to accumulate information, knowledge, and experience over time. The ninth theme indicates how certain opinions, viewpoints, and biases embedded in LLMs can influence the way AI agents operate. This theme highlights the potential risks associated with this (the word "xyz"). The tenth topic addresses the ways in which agents process and generate text using large language models, applicable in a wide variety of contexts, from automatic content generation to text comprehension and interpretation.

An analysis of the literature supports the authors' belief that LLM-based agents can be useful for automating routine tasks, replacing humans in these tasks, and in co-creating new and innovative solutions with humans by enhancing their capabilities (Guo et al., 2024; Wang et al., 2023; Wu et al., 2023; Xi et al., 2023).

Quasi-experiments results

The quasi-experiments were designed to test the effectiveness of different types of agents within single and multi-agent solutions, thereby verifying the potential usefulness of these solutions in the business domain. Single-agent solutions were prepared using the libraries LangChain (Chase, 2023) and CrewAI (https://www.

crewai.com/). The experiments analysed the performance of a retriever agent, a data agent and an SQL agent. Each agent had access to data, had four tools at its disposal, and was expected to perform five tasks. Those tasks allowed to assess the agents' reasoning, tool utilisation, interaction with the environment, memory use and generalization (Xi et al., 2023). Each experiment involved five consecutive repetitions to determine a task success metric and objectively evaluate performance. Subjective evaluation was done by humans and LLM experts on a scale of 1 (lowest rating) to 10 (highest rating) (Wang et al., 2023). The results of the quasi-experiments on single-agent solutions are provided in Tables 2–4.

Table 2. Evaluation of retriever agents

Framework	Agent Class	Category	Task success	LLM_Expert	Human_Expert
LangChain	Retriever_agent	Reasoning	4/5	7	8
LangChain	Retriever_agent	Tools utilization	5/5	8	9
LangChain	Retriever_agent	Environment interaction	4/5	9	9
LangChain	Retriever_agent	Memory utilization	3/5	8	7
LangChain	Retriever_agent	Generalization	5/5	9	8
Mean:				**8.2**	**8.2**
CrewAI	Retriever_agent	Reasoning	5/5	9	8
CrewAI	Retriever_agent	Tools utilization	5/5	7	8
CrewAI	Retriever_agent	Environment interaction	2/5	7	6
CrewAI	Retriever_agent	Memory utilization	0/5	5	2
CrewAI	Retriever_agent	Generalization	5/5	9	9
Mean:				**7.4**	**6.6**

Source: Authors' own study.

Table 2 shows the results of the evaluation of retriever agents running in two different frameworks (one in LangChain, the second in CrewAI). Each agent had access to information stored in a PDF file, was equipped with memory, and four different tools: one leading tool (retriever), two predefined tools, and one customized tool allowing interaction with the environment (saving information to disk in defined format). The results show that retriever agents are good at analysing the content of documents, extracting relevant information and interpreting it in different contexts. They make good use of tools by selecting them according to the type of task at hand. In the case of agents of this type deployed in CrewAI, there is a problem with memorizing and saving the acquired information in an appropriate format and location. Nevertheless, it can be concluded that agents of this type running in both LangChain and CrewAI generate satisfactory and reproducible results, and can be successfully used for tasks requiring document analysis, structured information retrieval, and reasoning or inference.

Table 3. Evaluation of data agents

Framework	Agent Class	Category	Task success	LLM_Expert	Human_Expert
LangChain	Data_agent	Reasoning	4/5	8	8
LangChain	Data_agent	Tools utilization	4/5	8	8
LangChain	Data_agent	Environment interaction	4/5	8	7
LangChain	Data_agent	Memory utilization	0/5	2	1
LangChain	Data_agent	Generalization	5/5	8	9
Mean:				**6.8**	**6.6**
CrewAI	Data_agent	Reasoning	4/5	3	4
CrewAI	Data_agent	Tools utilization	4/5	7	8
CrewAI	Data_agent	Environment interaction	2/5	4	5
CrewAI	Data_agent	Memory utilization	0/5	2	1
CrewAI	Data_agent	Generalization	5/5	3	3
Mean:				**3.8**	**4.2**

Source: Authors' own study.

Table 3 shows the results of evaluations based on LLM data agents operating in the LangChain and CrewAI frameworks. The results of the quasi-experiments clearly indicate that the agents perform worse with data, than with text analysis. The PandasDataFrame agent (LangChain), which leverages the analytical potential of the Pandas library, performs better. The lower values obtained by the data agent running in the CrewAI framework are due to the limitations of the spectator tool (CSVSearchTool). This tool allows semantically searching for queries in the content of a specified CSV file, and does not always work well when searching for data of a quantitative nature. The agent also had problems with the proper use of the customized tool (as did the retriever agent) and inference (due to errors in searching for the correct data). A drawback of both agents is their use of available memory. This is a major problem that severely limits the capabilities of these agents and their potential use in enterprises.

Table 4. Evaluation of SQL agents

Framework	Agent Class	Category	Task success	LLM_Expert	Human_Expert
LangChain	SQL_agent	Reasoning	4/5	7	7
LangChain	SQL_agent	Tools utilization	2/5	3	3
LangChain	SQL_agent	Environment interaction	2/5	6	5
LangChain	SQL_agent	Memory utilization	0/5	3	2
LangChain	SQL_agent	Generalization	5/5	9	7
Mean:				**5.6**	**4.8**
CrewAI	SQL_agent	Reasoning	4/5	8	9
CrewAI	SQL_agent	Tools utilization	4/5	9	9
CrewAI	SQL_agent	Environment interaction	2/5	6	6
CrewAI	SQL_agent	Memory utilization	0/5	2	1
CrewAI	SQL_agent	Generalization	4/5	8	9
Mean:				**6.6**	**6.8**

Source: Authors' own study.

Table 4 presents the results of the experiments for the SQL agents running using the LangChain and CrewAI frameworks. Due to problems with the correct operation of the leading tool (i.e., PGSearchTool), the CrewAI agent was equipped with the same set of tools available to the LangChain agent. Surprisingly, the SQL agent running in CrewAI made good use of this toolkit and performed better (in terms of reasoning, tool utilization, and generalization) than the SQL agent running in LangChain, for which these tools were designed. As in the previous cases, both agents had problems using the standard implemented memory module. They also did not perform well in using the customized tool for storing acquired and generated information on disk. For now, SQL agents can be useful for retrieving information from databases and interpreting it. Nevertheless, their operation currently requires human supervision.

The final stage of the quasi-experiments was the analysis of multi-agent solutions. The experiment was carried out with three agents running with the LangChain, CrewAI and AutoGen frameworks. Their task was to conduct a discussion and answer the question posed at the outset. Each agent had three tools at their disposal: a tool to retrieve information from an attached file (retriever), a tool to search the Internet and a tool to search a database of scientific articles. The evaluation of the results of the agents included the following aspects: role-playing ability, discussion effectiveness and problem-solving quality. The evaluation was conducted in the same manner as for single-agent solutions objectively evaluating performance using task success metric, subjective evaluation was done by humans and LLM experts (see Table 5).

Table 5. Evaluation of multi-agents problem solving

Framework	Agent Class	Category	Task success	LLM_Expert	Human_Expert
LangChain	Multi_agents	Role playing	5/5	9	8
LangChain	Multi_agents	Discussion effectiveness	5/5	8	9
LangChain	Multi_agents	Tool utilization	4/5	9	9
LangChain	Multi_agents	Problem solving	4/5	8	8
Mean:				**8.5**	**8.5**
AutoGen	Multi_agents	Role playing	4/5	9	9
AutoGen	Multi_agents	Discussion effectiveness	5/5	8	8
AutoGen	Multi_agents	Tool utilization	5/5	8	9
AutoGen	Multi_agents	Problem solving	4/5	7	7
Mean:				**8.0**	**8.3**
CrewAI	Multi_agents	Role playing	5/5	9	9
CrewAI	Multi_agents	Discussion effectiveness	5/5	8	8
CrewAI	Multi_agents	Tool utilization	4/5	9	9
CrewAI	Multi_agents	Problem solving	4/5	8	9
Mean:				**8.5**	**8.8**

Source: Authors' own study.

The analysis of the results shows that agents can be profiled to present specific viewpoints and positions in the discussion. They efficiently use the tools entrusted

to them, acquiring the information needed to solve problems. This indicates the considerable usefulness of this type of AI agent configuration in identifying problems, searching for their sources, and generating potential solutions.

Findings

The systematic literature review indicates that LLM-based agents are becoming a significant part of the scientific discourse in the field of AI. The analysis of the content of the articles carried out using the LDA method has made it possible to identify research topics related to LLM-based agents. These topics address issues related to their nature, operation, interaction, learning and self-evolution, as well as the risks involved.

In turn, analysis of the documentation of the LangChain, CrewAI and AutoGen libraries, has made it possible to identify agents (and their configurations) that may be useful in enterprises. The research shows that these frameworks offer the possibility to design, run, and utilize LLM-based agents. Experimental results demonstrate that agents are able to plan effectively and are good at using the tools at their disposal to carry out assigned tasks .They are also capable of generalizing and drawing conclusion. This enables them to automate routine tasks or collaborate with employees, thereby enhancing their capabilities. However, the problem of agent memory utilization, especially in single-agent solutions, persists. While conceptually simple, in practice agent memory provisioning requires complex prompt engineering techniques (especially for data agents and SQL agents). This aspect of agent LLM configuration needs to be refined and more seamlessly integrated into the way the agent operates. Enhancing agents with effective long-term memory will increase their potential to learn and grow. Currently, these are external vector databases where agents collect relevant information. The results of the experiments also indicate that, for now, the operation of LLM-based agents and the results of their work still require human supervision. Therefore, we cannot speak of full automation of the tasks or processes entrusted to LLM-based agents.

The results of the experiments also allow conclusions to be drawn regarding the purpose of the individual frameworks. CrewAI is a solution that will be useful for automating routine tasks. The commissioning agents within this environment produce reproducible, structured and format-compliant results. They make very good use of the tools available in the CrewAI library, as well as those offered by the LangChain library. In turn, agents running in LangChain framework are more flexible, making them more versatile. They interact well with predefined and customized tools. As already mentioned, they are more flexible, enabling them to perform more complex and comprehensive tasks using different tools. Additionally, this framework offers the possibility of running agents tailored to perform specific tasks (e.g., PandasDataFrame agent, SQL agent) along with toolkits dedicated to such agents. In terms of

solutions allowing multiple agents to solve a problem by using cognitive multi-agents, the agents running within each framework achieved similar results. This includes aspects such as role playing capacity, discussion effectiveness, tool utilization and problem-solving quality. However, AutoGen library offers the widest spectrum of possible configurations and provides the greatest flexibility for solutions involving the collaboration of multiple LLM agents (including state or graph agents).

Conclusions

The issue of LLM-based agents is an important research strand in the area of AI. These agents demonstrate significant potential for automating routine tasks and supporting employees in executing more complex and comprehensive activities. Literature surveys, analyses, and quasi-experiments collectively confirm that LLM agents are capable of effective planning, utilizing tools and executing assigned tasks. The capabilities of LLM-based agents also include the ability to learn and develop, indicating they should be regarded as more than tools. However, despite their great potential for the application of AI agents in organisations, their current operations requires human supervision. One of the key problems is the limited memory of LLM agents and the variability in reproducibility of results across iterations.

Emerging formulations of agents, such as graphs (LangGraph), or StateFlow (AutoGen), address these limitations, leading to the prediction that in the future LLM agents will be full-fledged actors in organisations, capable of performing complex tasks autonomously. Thus, further research and development of LLM agent technology is essential to fully harness their potential in organisational context.

References

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Chase, H. (2023). *Langchain*. https://docs.langchain.com/docs/

Eisenhardt, K.M. (1989). Agency theory: An assessment and review. *The Academy of Management Review*, *14*(1), 57–74.

Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N., Wiest, O., & Zhang, X. (2024). Large Language Model-based Multi-Agents: A Survey of Progress and Challenges. *exarXiv:2402.01680v1 [cs.CL]*.

Händler, T. (2023). Balancing autonomy and alignment: A multi-dimensional taxonomy for autonomous LLM powered multi-agent architectures. *arXiv preprint arxiv:2310.03659*. **https://doi.org/10.48550/arXiv.2310.03659**

Hong, S., Zheng, X., Chen, J., Cheng, Y., Zhang, C., Wang, Z., Yau, S.K.S., Lin, Z., Zhou, L., Ran, C., Schmidhuber, J., Wu, C., Xiao, L., Zhuge, M., & Wang, J. (2023). MetaGPT: Meta programming for multi-agent collaborative framework. *arXiv preprint arXiv:2308.00352*.

Hussein, A., Gaber, M.M., Elyan, E., & Jayne, C.(2017). Imitation learning: A survey of learning methods. *ACM Computing Surveys*, *50*(2), 1–35.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1–38. **https://doi.org/10.1145/3571730**

Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. (2023). Challenges and applications of large language models. *arXiv preprint arXiv:2307.10169*.

Korzynski, P., Mazurek, G., Altmann, A., Ejdys, J., Kazlauskaite, R., Paliszkiewicz, J., Wach, K., & Ziemba, E. (2023). Generative artificial intelligence as a new context for management theories: Analysis of ChatGPT. *Central European Management Journal*, *31*(1), 3–13. **https://doi.org/10.1108/CEMJ-02-2023-0091**

Lan, L.L., & Heracleous, L. (2010). Rethinking agency theory: The view from law. *Academy of Management Review*, *35*(2), 294–314.

Li, G., Hammoud, H.A.A.K., Itani, H., Khizbullin, D., & Ghanem, B. (2023). CAMEL: Communicative agents for "mind" exploration of large scale language model society. *arXiv preprint arXiv:2303.17760*.

Liu, Z., Zhang, Y., Li, P., Liu, Y., & Yang, D.(2023). Dynamic LLM-agent network: An LLM-agent collaboration framework with agent team optimization. *arXiv preprint arXiv:2310.02170*.

Ma, Q., Xue, X., Zhou, D., Yu, X., Liu, D., Zhang, X., Zhao, Z., Shen, Y., Ji, P., Li, J., Wang, G., & Ma, W. (2024). Computational experiments meet large language model based agents: A survey and perspective. *arXiv:2402.00262v1*.

Mo, Y., Kontonatsios, G., & Ananiadou, S. (2015). Supporting systematic reviews using LDA-based document representations. *Systematic Reviews*, *4*(172), 1–12. **https://doi.org/10.1186/s13643-015-0117-0**

Padgham, L., & Winikoff, M. (2005). *Developing Intelligent Agent Systems: A Practical Guide*. John Wiley & Sons.

Park, J.S., O'Brien, J.O., Cai, C.J., Morris, M.R., Liang, P., & Bernstein, M.S. (2023). Generative agents: Interactive simulacra of human behaviour. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, NY.

Qin, Y.S., & Hu, Y. (2023). Tool learning with foundation models. *CoRR*, *abs/2304.08354*.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, *46*, 192–210.

Ross, S., & Morrison, G. (2004). Experimental research methods. In D.J. Jonassen (Ed.), *Handbook of Research on Educational Communications and Technology* (pp. 1021–1043). Lawrence Erlbaum Associates.

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., Scialom, & Toolformer, T. (2023). Language models can teach themselves to use tools. *arXiv preprint*. *arXiv:2302.04761*.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W.X., Wei, Z., & Wen, J.-R. (2023). A survey on large language model-based autonomous agents. *CoRR*, *abs/2308.11432*.

Wei, J., Shuster, K., Szlam, A., Weston, J., Urbanek, J., & Komeili, M. (2023). Multi-party chat: Conversational agents in group settings with humans and models. *CoRR*, *abs/2304.13835*. **https://doi.org/10.48550/arXiv.2304.13835**

Wooldridge, M.J., & Jennings, N.R. (1995). Intelligent agents: Theory and practice. *Knowledge Engineering Review*, *10*(2), 115–152.

Wu, Q., Bansal, G., Zhang, J., Wu, J., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A.H., White, R.W., Burger, D., & Wang, C. (2023). AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. *arXiv:2308.08155v2 [cs.AI]*.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng,

R., Cheng, W., Zhang, Q., Qin, W., Zheng, Y., Qiu, X., Huang, X., & Gui, T. (2023). The rise and potential of large language model based agents: A survey. *arXiv:2309.07864*.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). React: Synergizing reasoning and acting in language models. *arXiv:2210.03629v3 [cs.CL]*.

Zhang, C., K. Yang, S. Hu, Wang, Z., Li, G., Sun, Y., Zhang, C., Zhang, Z., Liu, A., Zhu, S.-C., Chang, X., Zhang, J., Yin, F., Liang, Y., & Yang, Y. (2023). ProAgent: Building proactive cooperative AI with large language models. *CoRR*, *abs/2308.11339*.

Zhu, X., Chen, Y., Tian, H., Tao, C., Su, W., Yang, C., Huang, G., Li, B., Lu, L., Wang, X., Qiao, Y., Zhang, Z., & Dai, J. (2023). Ghost in the Minecraft: Generally capable agents for open-world environments via large language models with text-based knowledge and memory. *arXiv preprint arXiv:2305.17144*.

Appendix. A summary of the retrieved articles on the basis of which the LDA analysis was carried out

No.	Title	Authors	Main thesis of the article	Article URL
1	"DS-Agent: Automated Data Science by Empowering Large Language Models with Case-Based Reasoning"	(Guo et al., 2024) 2024-02-27	The main thesis of the article is development and deployment of a system called DS-Agent that utilizes large language models	http://arxiv.org/pdf/2402.17453v1.pdf
2	"Language Agents as Optimizable Graphs"	(Zhuge et al., 2024) 2024-02-26	The article introduces the new approach by describing LLM-based agents as computational graphs	http://arxiv.org/pdf/2402.16823v2.pdf
3	"Large Language Models as Urban Residents: An LLM Agent Framework for Personal Mobility Generation"	(Wang et al., 2024) 2024-02-22	The article propose a framework that leverages Large Language Models (LLMs) to simulate urban mobility and improve urban mobility management	http://arxiv.org/pdf/2402.14744v1.pdf
4	"KG-Agent: An Efficient Autonomous Agent Framework for Complex Reasoning over Knowledge Graph"	(Jiang et al., 2024) 2024-02-17	The article aim is to improve the reasoning ability of large language models (LLMs) over knowledge graphs (KGs) to answer complex questions	http://arxiv.org/pdf/2402.11163v1.pdf
5	"Watch Out for Your Agents! Investigating Backdoor Threats to LLM-Based Agents"	(Yang et al., 2024) 2024-02-17	The main thesis of the article is to study and highlight the security threat posed by backdoor attacks on LLM-based agents	http://arxiv.org/pdf/2402.11208v1.pdf
6	"Affordable Generative Agents"	(Yu et al., 2024) 2024-02-03	The article discusses the performance of these agents and aims to create more believable agents in the future	http://arxiv.org/pdf/2402.02053v1.pdf
7	"Formal-LLM: Integrating Formal Language and Natural Language for Controllable LLM-based Agents"	(Li et al., 2024) 2024-02-01	The main thesis of the article is to introduce the Formal-LLM framework for LLM-based agents, which combines the power of natural language comprehension with the precision of formal language	http://arxiv.org/pdf/2402.00798v2.pdf
8	"Computational Experiments Meet Large Language Model Based Agents: A Survey and Perspective"	(Ma et al., 2024) 2024-02-01	Article discusses the advantages that LLM-based Agents bring to computational experiments and how computational experiments can improve the explainability of LLM-based Agents	http://arxiv.org/pdf/2402.00262v1.pdf
9	"Large Language Model based Multi-Agents: A Survey of Progress and Challenges"	(Guo et al., 2024) 2024-01-21	The article aims to provide a comprehensive overview of LLM-based Multi-Agent (LLM-MA) systems, explain the fundamental concepts involved in establishing multi-agent systems	http://arxiv.org/pdf/2402.01680v1.pdf
10	"Open Models, Closed Minds? On Agents Capabilities in Mimicking Human Personalities through Open Large Language Models"	(La Cava et al., 2024) 2024-01-13	The article presents findings that suggest some open models exhibit a remarkable ability to mimic human personality traits, potentially enhancing interactions between humans and agents, especially in educational contexts	http://arxiv.org/pdf/2401.07115v1.pdf

No.	Title	Authors	Main thesis of the article	Article URL
11	"Speech Agents: Human-Communication Simulation with Multi-Modal Multi-Agent Systems"	(Zhang et al., 2024) 2024-01-08	The article presents the idea that leveraging LLM agents for simulating human communication can enhance our understanding of language and interaction, exploring cognitive processes and social mechanisms in human society	http://arxiv.org/pdf/2401.03945v1.pdf
12	"Exploring Large Language Model based Intelligent Agents: Definitions, Methods, and Prospects"	(Cheng et al., 2024) 2024-01-07	The article presents the effectiveness of LLM-based agents in advancing scientific inquiry, addressing mathematical challenges, and improving problem-solving capabilities in intricate mathematical problems	http://arxiv.org/pdf/2401.03428v1.pdf
13	"How Far Are We from Believable AI Agents? A Framework for Evaluating the Believability of Human Behavior Simulation"	(Xiao et al., 2023) 2023-12-28	The main thesis of the article is to propose two novel metrics, long consistency, and robustness, to measure the believability level of agents in storytelling	http://arxiv.org/pdf/2312.17115v1.pdf
14	"Do LLM Agents Exhibit Social Behavior?"	(Leng & Yuan, 2023) 2023-12-23	The main thesis of the article is to contribute to the fundamental understanding of the social behaviors of LLM based agents, particularly focusing on how they interact with humans and other LLM agents	http://arxiv.org/pdf/2312.15198v2.pdf
15	"Simulating Opinion Dynamics with Networks of LLM-Based Agents"	(Chuang et al., 2023) 2023-11-16	The article explores how LLM based agents form and potentially change their opinions based on true and false framings presented to them	http://arxiv.org/pdf/2311.09618v2.pdf
16	"LLM-Based Agent Society Investigation: Collaboration and Confrontation in Avalon Gameplay"	(Lan et al., 2023) 2023-10-23	The article contribute to a better understanding of the role of LLM-based agents in social and strategic contexts, shedding light on the implications of these behaviors in such environments	http://arxiv.org/pdf/2310.14985v1.pdf
17	"Large Language Model-Empowered Agents for Simulating Macroeconomic Activities"	(Li et al., 2023a) 2023-10-16	The article presents a framework where LLM-empowered agents emulate real-world economic actors and exhibit human-like decision-making patterns	http://arxiv.org/pdf/2310.10436v1.pdf
18	"Theory of Mind for Multi-Agent Collaboration via Large Language Models"	(Li et al., 2023b) 2023-10-16	The study aims to evaluate LLMs' high-order Theory of Mind (ToM) capabilities, encompassing dynamic belief state evolution and rich intent communication between multiple agents	http://arxiv.org/pdf/2310.10701v2.pdf
19	"Dynamic Task Sampling Using LLM-Generated Sub-Goals for Reinforcement Learning Agents"	(Shukla et al., 2023) 2023-10-14	The article presents LgTS, using LLMs to improve planning in reinforcement learning agents with reduced environmental interactions, without specific data or pre-training	http://arxiv.org/pdf/2310.09454v1.pdf

No.	Title	Authors	Main thesis of the article	Article URL
20	"Formally Specifying the High-Level Behavior of LLM-Based Agents"	(Crouse et al., 2023) 2023-10-12	The main thesis of the article is the introduction of a new agent architecture called PASS, which leverages a framework that allows for flexibility in generating text or executing actions using large language models (LLMs)	http://arxiv.org/pdf/2310.08535v3.pdf
21	"Dynamic LLM-Agent Network: An LLM-agent Collaboration Framework with Agent Team Optimization"	(Liu et al., 2023) 2023-10-03	The article presents DyLAN, a dynamic framework improving LLM agent collaboration for tasks like reasoning and code generation, achieving significant performance gains with optimized agent selection	http://arxiv.org/pdf/2310.02170v1.pdf
22	"AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation"	(Wu et al., 2023) 2023-10-03	The articles introduces AutoGen, an open-source framework for creating LLM applications with customizable agents that communicate to perform tasks, supporting a mix of LLMs, human inputs, and tools	https://arxiv.org/pdf/2308.08155.pdf
23	"The Rise and Potential of Large Language Model Based Agents: A Survey"	(Xi et al., 2023) 2023-09-19	The abstract reviews a survey on LLM-based AI agents, highlighting the lack of a general model for designing adaptable AI agents and positioning LLMs as potential AGI foundations	https://arxiv.org/pdf/2309.07864.pdf
24	"A Survey on Large Language Model Based Autonomous Agents"	(Wang et al., 2023) 2023-09-19	The article reviews LLM-based autonomous agents, offering a unified framework, exploring applications, evaluation methods, challenges, and future directions to understand their progress toward human-level intelligence	https://arxiv.org/pdf/2308.11432.pdf
25	"Enhancing Pipeline-Based Conversational Agents with Large Language Models"	(Foosherian et al., 2023) 2023-09-07	The main thesis of the article is that the latest advancements in AI and deep learning, particularly large language models (LLMs) like GPT-4, can enhance pipeline-based conversational agents	http://arxiv.org/pdf/2309.03748v1.pdf

Source: Authors' own study.