A Framework for Integrated Digital Forensic Investigation Employing AutoGen AI Agents

Akila Wickramasekara
School of Computer Science
University College Dublin
Dublin, Ireland
akila.wickramasekara@ucdconnect.ie

Mark Scanlon
School of Computer Science
University College Dublin
Dublin, Ireland
mark.scanlon@ucd.ie

***Abstract*—The increasing frequency and rapidity of criminal activities require faster digital forensic (DF) investigations. Currently, most DF phases involve manual procedures, requiring significant human effort and time, often facing evolving requirements. This paper proposes an integrated framework employing AutoGen Artificial Intelligence (AI) agents and Large Language Models (LLMs) such as LLAMA, and StarCoder. The suggested framework utilizes AI agents and LLMs to perform tasks articulated in natural language by a human agent. The proposed architecture presents a significant advantage by alleviating the investigative workload and shortening the learning curve for investigators. However, it is still combined with risks such as information accuracy, hallucination impact, and legal barriers. Although, this research contributes to the ongoing discourse on optimizing DF processes in response to the evolving landscape of criminal activities and the corresponding demands placed on investigative resources.**

***Keywords*—Digital Forensics, Large Language Models, AI Agents, AutoGen**

I. INTRODUCTION

The proliferation of criminal activities and their increasing complexity has required an increased allocation of human resources for investigative purposes. Insufficient expertise and experience among investigators can impede the Digital Forensic (DF) investigation process, leading to unnecessary delays and cost overruns. Simultaneously, the complex patterns of data and their voluminous nature, together with an insufficiency of standardization and limitations in existing tools, present an additional set of challenges [1]. These challenges contribute to delayed and costly investigations. The rapid growth of diverse LLMs has opened up new avenues for exploration in various research domains, with DF being no exception [2]. This surge in LLM capabilities presents opportunities for innovative research directions and advancements within the field of DF.

The DF process model conventionally comprises five pivotal phases: Incident Recognition, Collection & Seizure, Preservation, Examination, Analysis, and Reporting [3]. Recognizing the evolving landscape of technological advancements, recent research efforts highlight the potential integration of LLMs within these phases to improve the productivity and efficiency of DF investigations [4, 5].

Current research underscores the potential of LLMs to seamlessly integrate artificial intelligence (AI) agents within the DF process model [4]. By incorporating LLMs, it becomes feasible to implement AI agents capable of executing designated tasks through effective communication. This cohesiveness between LLMs and AI agents holds the promise of optimizing the investigative workflow of DF, offering novel insights, and fostering advancements in the field of DF.

A. Contribution of this Work

This paper proposes a novel architecture of using LLMs for DF investigations on top of AutoGen framework. Below are key contributions of this work.

- Formulate a framework for DF investigations that operates based on natural language inputs.
- Put forth an innovative architecture for the reusability of subtasks, specifically tailored for repetitive prompts.
- Introduce the concept of prompt engineering in the context of DF, aiming to generate subtasks from intricate and sequential tasks.

II. RELATED WORK

The literature on DF reveals a variety of implemented frameworks. One example is the Next Generation Digital Forensic Investigation Model (NGDFIM), which provides a comprehensive approach to DF investigations. This model emphasizes reducing investigation time, ensuring data integrity, and protecting privacy, incorporating advanced techniques like customized content imaging and on-site triage to enhance the efficiency and effectiveness of DF analysis, especially in addressing challenges related to data volume and privacy concerns [6].

The integration of AI into DF investigations and frameworks is an area of growing interest, but faces significant challenges, including the scarcity of comprehensive training datasets, explainability issues (both for the investigator and when used for expert testimony), and nuanced management of sensitive data [7]. Furthermore, automation of tasks within DF frameworks is gaining traction [8], facilitated by developments in natural language processing, AI, and LLMs.

979-8-3503-3036-6/24/$31.00 ©2024 IEEE

SimpleTOD offers a unified methodology for task-oriented dialogue based on GPT-2, streamlining the creation and optimization of tasks including dialogue state tracking, action decisions, and response generation [9]. Similarly, Yao et al. introduce ReAct, a novel approach that combines reasoning and action within LLMs to improve task creation, incorporating a Human-in-the-Loop Correction mechanism for fine-tuning model behavior [10]. Other notable contributions, such as ADAPT and SMART-LLM, demonstrate the potential of LLMs in dynamically decomposing complex tasks for improved success rates and adaptive task management, showing particular promise for applications in challenging environments [11, 12].

With the rapid advancements and surge in LLM research following ChatGPT's introduction in late 2022, there has been a significant increase in LLM-related studies, many of which are solely available as preprints on platforms like arXiv or OpenReview. In particular, foundational papers on GPT-4 [13] and LLaMA [14], despite being in preprint form, have attracted extensive citations. These documents are pivotal, providing crucial insights for current research and discussions in the field. The inclusion of preprint papers as part of the related work in this paper is vital for presenting the most current knowledge and viewpoints.

Recent releases like LLaMA and StarCoder are notable for their exceptional performance across various benchmarks, with LLaMA highlighted for its training across different parameter ranges and StarCoder for its specialization in coding tasks, achieving significant success rates [14, 15]. The literature also identifies a variety of projects that leverage LLMs for task automation, including Auto-GPT, BabyAGI, ChatDev, AutoGen, CAMEL, GPT-in-the-Loop, and TaskBench [16, 17, 18, 19, 20, 21, 22]. These initiatives illustrate a trend toward the use of LLMs to help users perform specific tasks, using bots with predefined roles to collaborate and achieve desired results [23]. Among these, AutoGen stands out for its facilitation of AI agent-based application development, offering an open source framework that simplifies the creation of intricate workflows and the orchestration of tasks, supported by a GUI for easy code-free solution development [19].

In the existing literature, while manual frameworks for DF are well-defined, there is a noticeable absence of frameworks that incorporate the automation enabled by LLMs. Additionally, a discernible gap exists in the literature on specifically trained LLMs tailored for DF applications. Despite advances in natural language processing, comprehensive and specialized LLMs designed explicitly for DF investigations are lacking in the current body of research. Such a DF focused LLM could have significant potential in enabling the querying and analysis by both DF personnel and non-DF experts, e.g., non-DF law enforcement officers, prosecutors, lawyers, etc., but has a variety of risks that require careful consideration and mitigation.

III. Design Considerations

The framework proposed as part of this paper aims to enhance the efficiency of DF investigations across the Examination, Analysis, and Reporting stages, with a primary focus on accelerating the investigative process while ensuring result accuracy and reliability. A critical feature of the framework is its ability to process and understand input from natural languages, distinguishing between specific tasks and irrelevant information. It emphasizes recognizing various language patterns and understanding technical DF terminology, improving the model's accuracy in interpreting user commands.

Task decomposition is identified as a crucial component, impacting outcomes and reporting. The framework introduces a systematic approach to decomposition using the "5W+H" or "5W1H" (Who, What, When, Where, Why, and How) method, alongside a validation mechanism that leverages a baseline data set for effective implementation. The strategy includes refining a specific LLM for task decomposition, in order to accurately identify pertinent subtasks. The Language Feedback Benchmark (LLF-Bench) is utilized to assess the LLM's performance in sub-task decomposition within the DF domain, offering a comprehensive evaluation through eight benchmarks [24].

Anticipated enhancements to the framework involve analyzing language input patterns and frequent queries to refine the AI agents' responsiveness to DF investigators' needs. The compilation of language inputs, task breakdowns, and outcomes will create a substantial validation dataset for ongoing refinement, with stored queries and responses facilitating quicker information retrieval.

The framework's integrity is intricately tied to the precision and clarity achieved through prompt engineering, establishing a pivotal dependency. Each AI agent's role is meticulously defined using engineered prompt language, dictating a precise workflow for task execution. The implementation phase demands a standardized and well-documented set of prompts, ensuring a seamless and accurate communication channel between human operators and AI agents. This emphasis on prompt engineering is foundational, guiding the framework's behavior and interactions, ultimately determining its efficacy in translating human natural language instructions into actionable tasks for optimal DF outcomes.

Furthermore, compliance with standards will be evaluated using the Computer Forensic Tool Testing Program (CFTT) established by the National Institute of Standards and Technology (NIST). NIST's provision of evaluation test cases serves as a benchmark for standardizing any tool or framework, ensuring consistent and comprehensive validation [25]. Testing against the NIST CFTT benchmark test cases not only assures the accuracy of the framework but also highlights any potential limitations or areas where the framework may not be suitable in the domain of DF.

Since AutoGen is an open-source framework, it provides the flexibility to customize its code to suit the specific needs of different agencies, stakeholders, and governments. The

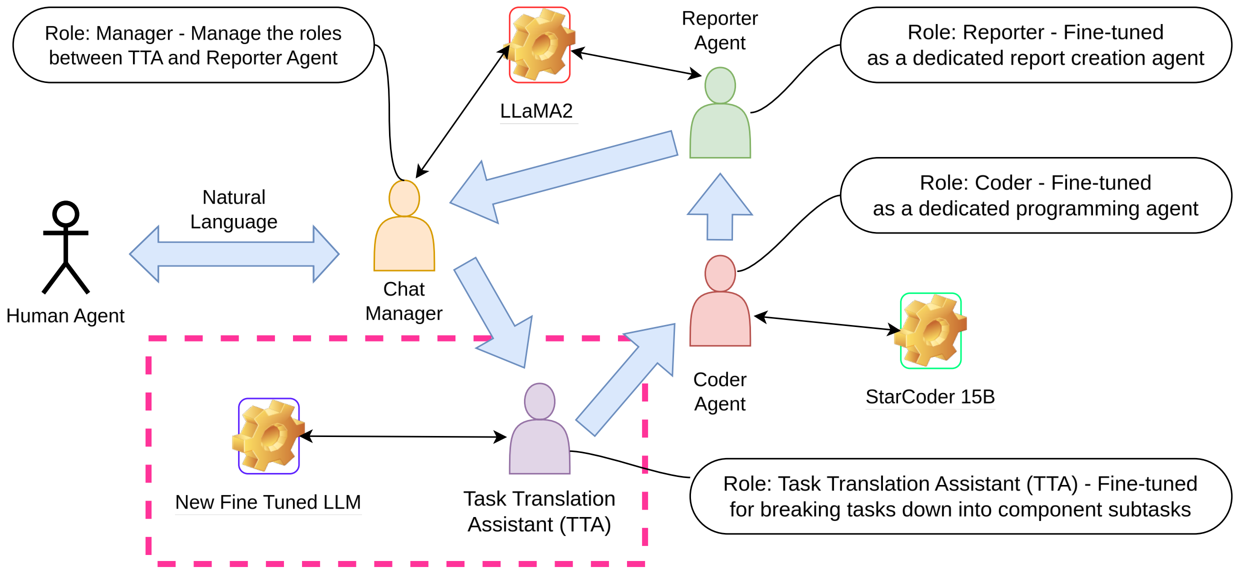

Fig. 1 Architecture of the Proposed Framework

ability to customize skills for AI agents enables seamless integration with other existing systems, frameworks, and tools. For instance, integration with the Hansken system, an open DF platform that uses the Hansken Query Language (HQL) to retrieve DF information stored in the cloud, becomes feasible [26]. With a predefined Application Programming Interface (API) layer and agent skills, this integration offers the flexibility of utilizing existing resources through natural language queries.

IV. Framework

The proposed framework will be built upon the AutoGen framework, integrating LLaMA and StarCoder LLMs alongside four AI agents. Figure 1 depicts the high-level architecture of the framework, and outlines the role definitions of the AI agents and specifying the respective LLMs they will utilize. The Coder Agent will be endowed with a predefined set of skills, each skill defined by a Python function with the specification of its functionality.

The interconnections among the AI agents of the framework and their respective roles within the system are described in Figure 1. After the entry of a natural language query by the human agent, the Chat Manager Agent assumes the responsibility of orchestrating activities between the Task Translation Assistant (TTA) and the Reporter Agent. The TTA, in turn, engages in direct communication with the Coder Agent, tasked with crafting scripts and test cases in accordance with the specified task received from the TTA. Once the Coder Agent generates results, they will be transmitted to the Reporter Agent. Subsequently, the Reporter Agent undertakes the creation of a comprehensive report, which is then conveyed back to the Chat Manager. The Chat Manager, in its final role, presents the generated report to the human agent, facilitating effective communication and comprehension of the investigative outcomes.

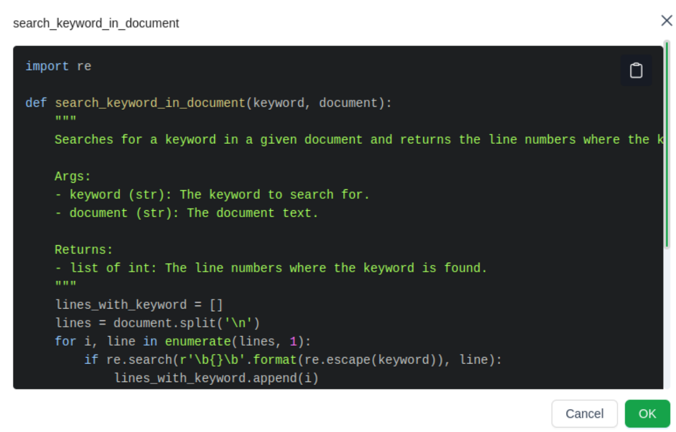

Fig. 2 Example Skill for Coder Agent

It is crucial to clearly define the purpose of the function, identify the necessary parameters, and specify the expected output in a skill. In Figure 2, a simple example skill is demonstrated to search for a specified keyword within a document. The required parameters are the keyword (a string) and the document text (also provided as a string). The function will return an integer value representing the line number in the document where the specified keyword is found. This concise articulation of the function's functionality, input requirements, and output ensures a clear and effective integration of a skill into the proposed framework.

Similarly, unique sets of skills will be defined for both the Report Agent and Chat Manager. Each of these agents will be equipped with the LLaMA LLM, while the Coder Agent will be powered by the StarCoder LLM.

The roles of each agent must be precisely defined to ensure that they comply with the given instructions. To achieve this, prompt engineering is essential. Jules et al. [27] has introduced a catalog of prompt patterns specifically designed for LLMs and this catalog can be utilized to formulate the roles and behaviors of each agent, ensuring accurate and consistent responses as per the instructions provided.

The TTA assumes a critical role in translating human natural language tasks into decomposed subtasks and instructions, easily interpretable by the Coder Agent. This role requires definition with a predefined skill set that includes identifying complex tasks, determining the necessary number of subtasks, and selecting the optimal set of subtasks to maintain investigation efficiency. As indicated with the dash line in Figure 2 the TTA will be powered by a specially fine-tuned LLM tailored for DF task decomposition. This ensures a high level of understandability for complex natural language tasks, provided by human agents. Simultaneously, TTA plays a crucial role in discerning whether the natural language query necessitates decomposition or warrants a direct response for the human agent. This analytical process, integrated within the framework, contributes to the system's adaptability by determining the appropriate course of action based on the intricacies inherent in the query.

As outlined in Section III, task decomposition is carried out systematically taking into account the key aspects of who needs to execute the task, the nature of the task itself, the designated time frame for execution, the specific location for execution of the task and the methodological approach to task execution. This comprehensive approach ensures that human-input tasks are properly comprehended and appropriately decomposed within the framework. By addressing these aspects of each task, the system aims to improve clarity and precision in the execution of tasks within the DF framework. As identified in the feasibility study, task understanding and decomposition require a substantial amount of time. When a user inputs the same query repeatedly, the TTA would otherwise need to decompose the subtask each time, incurring significant costs and time expenditure. To mitigate this time-consuming process for repetitive or similar user queries, pre-generated decomposition tasks will be employed. As depicted in Figure 3, a memory-based NoSQL database will be utilized to store instructions based on the predefined skills (programs) of the TTA. This database will observe input query patterns, determining if a new query is similar to previous ones. In case of similarity, the pre-stored decomposed sub-tasks will be retrieved and sent to the Coder Agent. For new queries not present in the database, the same TTA skill will store the instructions and the generated sub-tasks. The implementation is planned to leverage Redis database for its efficiency, being a fast NoSQL database capable of handling a large volume of data expeditiously [28].

Upon receiving instructions, the Coder Agent will generate the corresponding code along with unit tests. Following the predefined instructions provided to the Coder Agent, it will then pass the obtained information and data, after conducting unit testing to verify accuracy, to the Report Agent. The Report Agent, equipped with predefined skills and role instructions,

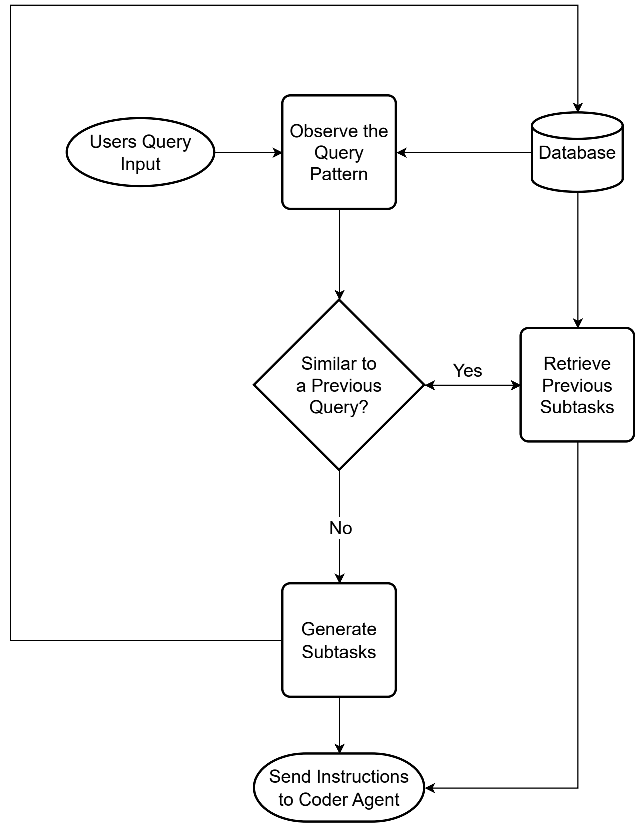

Fig. 3 Flow Diagram of the Task Translation Assistant

will formulate a report and transmit it to the Chat Manager. The Chat Manager, upon confirmation that the instructions are met, will prompt the human agent for verification and validation. If the human agent is dissatisfied, they may redefine the input query and seek clarification.

V. BENEFITS AND RISKS

LLMs offer significant productivity benefits, but are not without limitations, including biases, information hallucinations, and challenges in explaining [29]. The proposed framework integrates LLM advantages while addressing these limitations, aiming to transform DF investigations. It can reduce the learning curve for investigators and reduce the need for in-depth technical knowledge. By eliminating the requirement for coding expertise, the framework makes DF tools more accessible, broadening the investigative field to include professionals of diverse backgrounds without compromising on technical requirements.

In addition, the framework significantly improves the efficiency of DF investigations by automating information extraction and reporting processes. This not only speeds up case resolutions but also contributes to a more productive DF workflow overall. A key feature is its ability to track and refine responses to user prompts through dynamic interaction with stored prompts and decomposed tasks, improving its utility and adaptability for real-world applications.

Additionally, the framework's API ensures easy integration with existing DF tools, allowing it to act as a valuable plugin that bridges natural language input with robust DF solutions. This interoperability further cements its value within the DF ecosystem, offering simplicity and enhanced functionality without sacrificing the capabilities of existing systems.

However, the framework's reliance on natural language input introduces risks, such as inaccuracies due to the investigator's language proficiency. Therefore, the accuracy of information retrieval can vary, affecting the reliability of investigative outcomes. The framework also faces challenges from LLM hallucinations, which could affect the accuracy of generated reports by producing content that deviates from factual information. Furthermore, adversarial attacks targeting LLMs pose a risk to the integrity of the framework [30]. Ensuring robust security measures is crucial to safeguard against such threats and maintain the reliability of the investigative process.

Lastly, despite efforts to include human verification, acceptance issues regarding the generated data and reports may arise across different legal jurisdictions. The framework must navigate diverse legal standards and ensure compliance to facilitate its integration into the global DF landscape effectively.

VI. Conclusion

The introduction of a novel framework incorporating AI agents and LLMs is presented in this paper to revolutionize traditional DF investigations. This innovation is expected to significantly improve efficiency, saving investigators considerable time and contributing to the reduction of pending case backlogs. The integration of task decomposition and LLM role management opens new avenues for DF research, particularly in the realms of decision-making and timeline creation for cases.

Despite the promising impacts, it is acknowledged that certain risks are associated with the proposed framework, emphasizing the necessity of human verification in each information retrieval process. Future improvements are envisioned, leveraging data gathered from investigators' input prompts, decomposed tasks, and the codes and reports generated in the process. Enhancements may include the integration of multiple AI agents and the introduction of auto-defined agents, enabling the framework to generate its agents dynamically based on the complexity of human agents' input queries.

The openness of the framework facilitates extensible API integration, allowing seamless integration with various DF tools and platforms. Emphasizing the precision of the information recovered is crucial, warranting the development of a dedicated evaluation mechanism for this DF framework in the future. Evaluating the accuracy of information retrieval will be a key factor in assessing the efficacy of the framework, ensuring its reliability and suitability for diverse DF scenarios.

The implementation of the framework, reliant on multiple LLMs, necessitates high-end processing hardware, including servers with large Random Access Memory (RAM) and Graphics Processing Unit (GPU) capabilities. This infrastructure requirement, while contributing to initial implementation and maintenance costs, is a crucial consideration. The upfront investment in powerful hardware is offset by the framework's usability and significant time-saving advantages, positioning it as a strategic and impactful solution within the realm of DF investigations. Despite the associated costs, the long-term benefits in terms of efficiency and productivity may outweigh the initial financial considerations, making the investment in robust hardware infrastructure a prudent choice for optimal framework performance.

However, challenges such as jurisdictional limitations and infrastructure costs pose potential hindrances to the widespread adoption of this novel framework. These considerations contribute to the argument that the framework's status as the future of DF investigations may be subject to debate. Ongoing refinements and addressing these challenges will play a crucial role in determining the framework's viability and acceptance within the broader landscape of DF.

References

[1] H. Dubey, S. Bhatt, and L. Negi, "Digital Forensics Techniques and Trends: A Review," *The International Arab Journal of Information Technology (IAJIT)*, vol. 20, no. 4, pp. 644–654, 2023.

[2] M. Scanlon, B. Nikkel, and Z. Geradts, "Digital forensic investigation in the age of ChatGPT," *Forensic Science International: Digital Investigation*, vol. 44, p. 301543, 2023.

[3] E. Casey, *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet*, 3rd ed. USA: Academic Press, Inc., 2011.

[4] M. Scanlon, F. Breitinger, C. Hargreaves, J.-N. Hilgert, and J. Sheppard, "ChatGPT for digital forensic investigation: The good, the bad, and the unknown," *Forensic Science International: Digital Investigation*, vol. 46, p. 301609, 2023.

[5] G. Michelet and F. Breitinger, "ChatGPT, Llama, can you write my report? An experiment on assisted digital forensics reports written using (local) large language models," *Forensic Science International: Digital Investigation*, vol. 48, p. 301683, 2024.

[6] A. A. Thakar, K. Kumar, and B. V. Patel, "Next Generation Digital Forensic Investigation Model (NGDFIM) - Enhanced, Time Reducing and Comprehensive Framework," *Journal of Physics: Conference Series*, vol. 1767, no. 1, p. 012054, Feb 2021.

[7] X. Du, C. Hargreaves, J. Sheppard, F. Anda, A. Sayakkara, N.-A. Le-Khac, and M. Scanlon, "SoK: Exploring the State of the Art and the Future Potential of Artificial Intelligence in Digital Forensic Investigation," in *Proceedings of the 15th International Conference on Availability, Reliability and Security*, ser. ARES '20. New York, NY, USA: Association for Computing Machinery, 2020.

[8] F. Breitinger, J.-N. Hilgert, C. Hargreaves, J. Sheppard, R. Overdorf, and M. Scanlon, "DFRWS EU 10-year review and future directions in Digital Forensic Research," *Forensic Science International: Digital Investigation*, vol. 48, p. 301685, 2024.

[9] E. Hosseini-Asl, B. McCann, C.-S. Wu, S. Yavuz, and R. Socher, "A Simple Language Model for Task-Oriented Dialogue," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 20 179–20 191. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/e946209592563be0f01c844ab2170f0c-Paper.pdf

[10] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, "ReAct: Synergizing Reasoning and Acting in Language Models," vol. abs/2210.03629, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2210.03629

[11] A. Prasad, A. Koller, M. Hartmann, P. Clark, A. Sabharwal, M. Bansal, and T. Khot, "ADaPT: As-Needed Decomposition and Planning with Language Models," *CoRR*, vol. abs/2311.05772, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2311.05772

[12] S. S. Kannan, V. L. N. Venkatesh, and B. Min, "SMART-LLM: Smart Multi-Agent Robot Task Planning Using Large Language Models," *CoRR*, vol. abs/2309.10062, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2309.10062

[13] OpenAI, "GPT-4 technical report," *CoRR*, vol. abs/2303.08774, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2303.08774

[14] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample, "LLaMA: Open and Efficient Foundation Language Models," *CoRR*, vol. abs/2302.13971, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2302.13971

[15] R. Li, L. B. Allal, Y. Zi, N. Muennighoff, D. Kocetkov *et al.*, "StarCoder: May the Source Be with You!" *CoRR*, vol. abs/2305.06161, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2305.06161

[16] M. Firat and S. Kuleli, "What If GPT-4 Became Autonomous: The Auto-GPT Project and Use Cases," *Journal of Emerging Computer Technologies*, vol. 3, no. 1, p. 1–6, 2023.

[17] J. Bieger, K. R. Thórisson, and P. Wang, "Safe Baby AGI," in *Artificial General Intelligence*, J. Bieger, B. Goertzel, and A. Potapov, Eds. Cham: Springer International Publishing, 2015, pp. 46–49.

[18] C. Qian, X. Cong, C. Yang, W. Chen, Y. Su, J. Xu, Z. Liu, and M. Sun, "Communicative Agents for Software Development," *CoRR*, vol. abs/2307.07924, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2307.07924

[19] Q. Wu, G. Bansal, J. Zhang, Y. Wu, S. Zhang, E. Zhu, B. Li, L. Jiang, X. Zhang, and C. Wang, "AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation Framework," 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2308.08155

[20] G. Li, H. A. A. K. Hammoud, H. Itani, D. Khizbullin, and B. Ghanem, "CAMEL: Communicative Agents for "Mind" Exploration of Large Scale Language Model Society," 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2303.17760

[21] N. M. do Nascimento, P. S. C. Alencar, and D. D. Cowan, "GPT-in-the-Loop: Adaptive Decision-Making for Multiagent Systems," 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2308.10435

[22] Y. Shen, K. Song, X. Tan, W. Zhang, K. Ren, S. Yuan, W. Lu, D. Li, and Y. Zhuang, "TaskBench: Benchmarking Large Language Models for Task Automation," *CoRR*, vol. abs/2311.18760, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2311.18760

[23] S. Hong, X. Zheng, J. Chen, Y. Cheng, J. Wang, C. Zhang, Z. Wang, S. K. S. Yau, Z. Lin, L. Zhou, C. Ran, L. Xiao, and C. Wu, "MetaGPT: Meta Programming for Multi-Agent Collaborative Framework," 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2308.00352

[24] C. Cheng, A. Kolobov, D. Misra, A. Nie, and A. Swaminathan, "LLF-Bench: Benchmark for Interactive Learning from Language Feedback," *CoRR*, vol. abs/2312.06853, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2312.06853

[25] W. A. Bhat, A. AlZahrani, and M. A. Wani, "Can Computer Forensic Tools Be Trusted in Digital Investigations?" *Science & Justice*, vol. 61, no. 2, pp. 198–203, 2021.

[26] H. van Beek, J. van den Bos, A. Boztas, E. van Eijk, R. Schramp, and M. Ugen, "Digital Forensics as a Service: Stepping Up the Game," *Forensic Science International: Digital Investigation*, vol. 35, p. 301021, 2020.

[27] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt, "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT," *CoRR*, vol. abs/2302.11382, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2302.11382

[28] J. L. Carlson, *Redis in Action*. Shelter Island, NY: Manning, 2013.

[29] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J. Nie, and J. Wen, "A Survey of Large Language Models," *CoRR*, vol. abs/2303.18223, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2303.18223

[30] A. Zou, Z. Wang, J. Z. Kolter, and M. Fredrikson, "Universal and Transferable Adversarial Attacks on Aligned Language Models," *CoRR*, vol. abs/2307.15043, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2307.15043