Filing of Summary and Discussion Reports of Reviewed Articles.

Question: **How can the collapse of generative AI models affect the accuracy and quality of responses provided by autonomous agents in interactions with human users in the corporate environment?**	
Phrases chosen for queries:	Collapse of generative AI models; Quality of AI responses; Autonomous AI agents; Autonomous AI agents in corporate applications; Impact of AI model degradation; Synthetic distributed data; Multi-agents in corporate applications; Bias in Multi-Agents; AI ethics and hallucinations; Agents AI in recommendation; Hallucinations in LLMs; Poisoning in data distribution for LLMs; Collapse models improve bias;
Sources (magazines, conferences, etc.) of papers used:	Nature; International Journal of Science and Research Archive; Frontiers of Computer Science; arXiv; IEEE; Electronic Commerce Research; Scientific Reports;
Name of selected papers:	AI models collapse when trained on recursively generated data
	Review of autonomous systems and collaborative AI agent frameworks
	A Survey on Large Language Model based Autonomous Agents
	Can We Trust AI Agents? An Experimental Study Towards Trustworthy LLM-Based Multi-Agent Systems for AI Ethics
	Transforming Competition into Collaboration: The Revolutionary Role of Multi-Agent Systems and

	Language Models in Modern Organizations
	Transforming Competition into Collaboration: The Revolutionary Role of Multi-Agent Systems and Language Models in Modern Organizations
	Multi-Agent Large Language Models for Conversational Task-Solving
	Strong Model Collapse
	How to Synthesize Text Data without Model Collapse?
	The Impact of Large Language Models in Academia: from Writing to Speaking
	Mitigating Social Bias in Large Language Models: A Multi-Objective Approach Within a Multi-Agent Framework
	Towards Implicit Bias Detection and Mitigation in Multi-Agent LLM Interactions
	Fairness in Multi-Agent AI: A Unified Framework for Ethical and Equitable Autonomous Systems
	Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey
	Personalized Recommendation Systems using Multimodal, Autonomous, Multi Agent Systems
	Consumer reactions to technology in retail: choice

	uncertainty and reduced perceived control in decisions assisted by recommendation agents
	Investigating Bias in LLM-Based Bias Detection: Disparities between LLMs and Human Perception
	Bias of AI-generated content: an examination of news produced by large language models
	A Multi-Agent Conversational Recommender System
	Poisoning and Backdooring Contrastive Learning
General Summary	
These studies show how LLM models and multi-agent systems are increasingly influencing technology, especially in areas such as recommendation, ethics and performance. One of the main points discussed is "model collapse", which happens when AI is trained with uncurated and synthetic data, which causes it to lose data diversity and impair its performance. This can also increase biases in models, which is concerning, especially in areas like decision-making and product recommendations. In the case of multi-agent systems, many studies explore how these agents can improve interaction with users, helping to control the flow of conversations and collect feedback to adjust model responses. These systems have shown potential in various industries, such as e-commerce, finance and healthcare, by personalizing experiences and improving adaptation to different contexts. Another big topic is bias in AI models. This has been a challenge, and several approaches are being tested to reduce these biases, such as using multiple agents and adjusting the prompts that are given to models. There are also discussions about how to deal with ethics in AI, creating systems that are more transparent and accountable.	

Report 1

Paper name:	AI models collapse when trained on recursively generated data	Reference:	SHUMAILOV, Ilia; SHUMAYLOV, Zakhar; ZHAO, Yiren; PAPERNOT, Nicolas; Anderson, Ross; GAL, Yarin. AI models collapse when trained on recursively generated data. *Nature*, v. 631, no. 8022, p. 755-759, 2024. Available at: https://www.nature.com/articles/s41586-024-07566-y. Accessed on: 10 Feb. 2025.

Understanding the Abstract:

Based on the summary read, it was observed that model collapse is a phenomenon that can occur in any generative model. This collapse is characterized as a degenerative process in which, in its initial phase, the original distribution of data begins to lose its tails, that is, the states of lower probability.

This phenomenon tends to become increasingly common, as the content available on the internet has been progressively influenced by texts generated by models, reducing the presence of content produced exclusively by humans. As a consequence, during data scraping, the distribution begins to reflect more and more artificially generated information, further accelerating the collapse of the models.

Understanding the Discussion/Conclusion:

In the “Discussion” section of the paper, it was proven that collapse really happens with generative AI models and, given this fact, it is necessary to ensure that low

probability events are considered during the regrouping of data distribution for model training, ensuring the generation of fair responses to the user. Furthermore, some actions were recommended in response to the collapse problem, such as ensuring that the original distribution of data, arising from interaction with human content, is part of the training of successor models. Furthermore, it was suggested the creation of an international coordination that ensures the sharing of information between different groups of developers, with the aim of guaranteeing the development of quality LLMs, based on reliable data distribution.

Report 2			
Paper name:	Review of autonomous systems and collaborative AI agent frameworks	Reference:	JOSHI, Satyadhar. (2025). Review of autonomous systems and collaborative AI agent frameworks. International Journal of Science and Research Archive. 14. 961-972. 10.30574/ijsra.2025.14.2.0439. Disponível em: https://www.researchgate.net/publication/389068903_Review_of_autonomous_systems_and_collaborative_AI_agent_frameworks. Accessed on: 22 Feb. 2025.
Understanding the Abstract:			
The paper in question presents an overview of the current use of AI agents, focusing on frameworks, highlighting the main tools, their advantages and disadvantages. Furthermore, it explores autonomous agent technology, emphasizing its concept, applications and technical aspects, including limitations and opportunities. It also			

discusses future trends, offering a comprehensive view on the topic.

The study continues with an analysis of the application of this technology in different sectors, such as finance, risk management and the corporate environment.

Finally, the paper serves as a guide, consolidating the main recent observations on the evolution and impact of autonomous agents.

Understanding the Discussion/Conclusion:

The conclusion of this paper provides an overview of the rapid advancement of AI agents and the main tools for building this technology, providing an overview of its use in different scenarios. Furthermore, the application of AI agents in highly complex tasks is emphasized. However, the text also highlights its weaknesses and recommends a more rigorous approach in terms of ethics and governance for this emerging technology.

Report 3			
Paper name:	A Survey on Large Language Model based Autonomous Agents	Reference:	WANG, Lei; MA, Chen; FENG, Xueyang; ZHANG, Zeyu; YANG, Hao; ZHANG, Jingsen; CHEN, Zhiyuan; et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, v. 18, no. 6, 2024, p. 186345. Available at: https://arxiv.org/abs/2308.11432. Accessed on: February 22nd. 2025.

Understanding the Abstract:
The study in question demonstrates the evolution of autonomous agents with the help of LLMs in carrying out complex tasks. It explores the advancement of these agents in increasingly diverse activities in different sectors, offering a holistic view of the topic. Furthermore, it presents the most common strategies for integrating these technologies and the challenges involved.
Understanding the Discussion/Conclusion:
The "Conclusion" section of the paper states that the study provided a detailed overview of the main advances in agents assisted by LLMs, addressing their construction, application and evolution, in addition to mentioning technical aspects. Finally, the conclusion emphasizes that the paper highlights the main challenges and shortcomings of these tools.

Report 4			
Paper name:	Can We Trust AI Agents? An Experimental Study Towards Trustworthy LLM-Based Multi-Agent Systems for AI Ethics	Reference:	**CERQUEIRA, José Antonio Siqueira de; et al.** Can we trust AI agents? An experimental study towards trustworthy LLM-based multi-agent systems for AI ethics. *arXiv preprint*, arXiv:2411.08881, 2024. Available at: https://arxiv.org/abs/2411.08881. Accessed on: 22 Feb. 2025.
Understanding the Abstract:			
In this study, we analyzed how LLMs can help in the development of ethical AI. A			

prototype called LLM-BMAS was created, which uses multiple agents to discuss real ethical issues, generating detailed ethical code. The system addressed topics such as bias, transparency, responsibility, consent and compliance.
Understanding the Discussion/Conclusion:
The study shows techniques to make AI models more reliable in the area of software engineering. To achieve this, a multi-agent system was created, where each one had a specific role within the process, helping to organize information and improve the quality of responses. Additionally, the system used structured debates and conversations to improve decision-making.

Report 5			
Paper name:	Transforming Competition into Collaboration: The Revolutionary Role of Multi-Agent Systems and Language Models in Modern Organizations	Reference:	CRUZ, Carlos Jose Xavier. Transforming competition into collaboration: The revolutionary role of multi-agent systems and language models in modern organizations. *arXiv preprint*, arXiv:2403.07769, 2024. Available at: https://arxiv.org/abs/2403.07769. Accessed on: 27 Feb. 2025.
Understanding the Abstract:			
This paper talks about how combining multi-agent systems (SMA) with large language models (LLM) can change the way humans interact with artificial agents. The idea is			

<table>
<tr><td>to use these agents to help with both day-to-day operational tasks and strategic decisions within companies.
The study approach proposes creating LLM-based agents with different profiles, which simulate specific behaviors and interact with each other in a guided conversation format.</td></tr>
<tr><td>Understanding the Discussion/Conclusion:</td></tr>
<tr><td>This text ends by emphasizing the interaction between multi-agents and LLMs, highlighting their high positive impact on tasks that require collaboration in organizations. It then presents a summary of the tasks most commonly performed using AI and concludes with a reflection on how this technological interaction will provide new forms of application, reducing complexity and encouraging the creative use of these technologies.</td></tr>
</table>

<table>
<tr><td colspan="4">Report 6</td></tr>
<tr><td>Paper name:</td><td>Multi-Agent Large Language Models for Conversational Task-Solving</td><td>Reference:</td><td>BECKER, Jonas. Multi-agent large language models for conversational task-solving. <i>arXiv preprint</i>, arXiv:2410.22932, 2024. Available at: https://arxiv.org/abs/2410.22932. Accessed on: 22 Feb. 2025.</td></tr>
<tr><td colspan="4">Understanding the Abstract:</td></tr>
<tr><td colspan="4">This work evaluates multi-agent systems in conversational tasks, analyzing their performance in different paradigms. I propose a taxonomy of 20 studies (2022-2024) and a framework for multi-agent LLMs.</td></tr>
</table>

Understanding the Discussion/Conclusion:
This conclusion addresses the main theme of the paper, the interaction between multi-agents in the context of communication. However, the main focus was the relationship between agents when solving tasks, highlighting their reactions in different scenarios and the impact of the duration of conversations on their performance. Furthermore, it is noteworthy that agents are able to guarantee ethics in their interactions, avoiding inappropriate topics. Finally, the text emphasizes that the introduction of LLMs to support multi-agents contributes significantly to the resolution of complex tasks and high performance.

Report 7			
Paper name:	Strong Model Collapse	Reference:	DOHMATOB, Elvis; FENG, Yunzhen; SUBRAMONIAN, Arjun; KEMPE, Julia. Strong model collapse. *arXiv preprint*, arXiv:2410.04840, 2024. Available at: https://arxiv.org/abs/2410.04840. Accessed on: 27 Feb. 2025.
Understanding the Abstract:			
This study analyzes model collapse in large neural networks caused by synthetic data in training. Even 1% synthetic data can lead to performance degradation, making augmenting the training set useless. The impact of increasing model size is also investigated, showing that larger models can amplify the collapse.			
Understanding the Discussion/Conclusion:			

Report 8

Paper name:	How to Synthesize Text Data without Model Collapse?	Reference:	ZHU, Xuekai; CHENG, Daixuan; LI, Hengli; ZHANG, Kaiyan; HUA, Ermo; LV, Xingtai; DING, Ning; LIN, Zhouhan; ZHENG, Zilong; ZHOU, Bowen. How to Synthesize Text Data without Model Collapse? *arXiv preprint*, arXiv:2412.14689, 2024. Available at: https://arxiv.org/abs/2412.14689. Accessed on: 22 Feb. 2025.

Understanding the Abstract:

The study analyzes the impact of synthetic data on training language models, showing that a greater proportion of synthetic data reduces model performance. Statistical analyzes indicate changes in data distribution and excess n-grams. To avoid model collapse, we propose editing tokens on human data to generate semi-synthetic data. Experiments confirm that this technique improves data quality and model performance.

Understanding the Discussion/Conclusion:

The paper in question concludes that the use of synthetic data can compromise the

effectiveness of pre-training when combined with human data, resulting in non-iterative model collapse. Furthermore, to mitigate this problem, the authors propose editing at the token level, adopting a resampling method guided by a pre-trained model.

Report 9			
Paper name:	The Impact of Large Language Models in Academia: from Writing to Speaking	Reference:	GENG, Mingmeng; CHEN, Caixi; WU, Yanru; CHEN, Dongping; WAN, Yao; ZHOU, Pan. The impact of large language models in academia: from writing to speaking. *arXiv preprint*, arXiv:2409.13686, 2024. Available at: https://arxiv.org/abs/2409.13686. Accessed on: 22 Feb. 2025.
Understanding the Abstract:			
The study shows that large language models (LLMs) are increasingly impacting human society, especially in textual information. The impact on speech is beginning to emerge and is likely to grow in the future, drawing attention to the implicit influence and ripple effect of LLMs on human society.			
Understanding the Discussion/Conclusion:			
This paper points out that, in the academic context, an increasing number of people use the response patterns generated by LLMs, influencing both writing and speaking, especially writing. Consequently, the text emphasizes the possible risk of the model collapsing, considering that, as more people use this tool, including in the academic area, the chance of obtaining answers from a collapsed model, that is, with biases,			

increases.

<table>
<tr><td colspan="4">Report 10</td></tr>
<tr><td>Paper name:</td><td>Mitigating Social Bias in Large Language Models: A Multi-Objective Approach Within a Multi-Agent Framework</td><td>Reference:</td><td>XU, Zhenjie; CHEN, Wenqing; TANG, Yi; LI, Xuanying; HU, Cheng; CHU, Zhixuan; REN, Kui; ZHENG, Zibin; LU, Zhichao. Mitigating social bias in large language models: A multi-objective approach within a multi-agent framework. <i>arXiv preprint</i>, arXiv:2412.15504, 2024. Available at: https://arxiv.org/abs/2412.15504. Accessed on: 22 Feb. 2025.</td></tr>
<tr><td colspan="4">Understanding the Abstract:</td></tr>
<tr><td colspan="4">In this study, a multi-objective approach within a multi-agent framework (MOMA) was proposed to reduce social bias in LLMs without significantly impairing performance. MOMA uses multiple agents to carry out causal interventions on the content related to bias in the questions, breaking the direct connection between this content and the answers.</td></tr>
<tr><td colspan="4">Understanding the Discussion/Conclusion:</td></tr>
<tr><td colspan="4">The conclusion of the paper highlights the techniques used to mitigate the bias of LLM models, one of the most effective being the use of multi-agents to address this problem.</td></tr>
</table>

Report 11			
Paper name:	Towards Implicit Bias Detection and Mitigation in Multi-Agent LLM Interactions	Reference:	BORAH, Angana; MIHALCEA, Rada. Towards implicit bias detection and mitigation in multi-agent LLM interactions. *arXiv preprint*, arXiv:2410.02584, 2024. Available at: https://arxiv.org/abs/2410.02584. Accessed on: 22 Feb. 2025.
Understanding the Abstract:			
In this study, LLM models are being used to gain insights into social aspects, so it is essential to mitigate biases. In this paper, the presence of implicit gender biases in multi-agent interactions with LLMs was investigated and two strategies were proposed to reduce them.			
Understanding the Discussion/Conclusion:			
The present study demonstrates the presence of bias in LLM models in the context of gender. Researchers developed analysis techniques that highlighted the occurrence of bias on a recurring basis. Furthermore, several conclusions were drawn throughout the study, the main ones being: LLMs generate biases even when trained with data produced by humans; LLM models with a greater number of parameters are more prone to bias; the interaction between multiple agents and LLMs can exacerbate bias; and fine-tuning can be an effective technique for mitigating bias in the context of interaction between generative AI models and multi-agent systems.			

Report 12

Paper name:	Fairness in Multi-Agent AI: A Unified Framework for Ethical and Equitable Autonomous Systems	Reference:	RANJAN, Rajesh; GUPTA, Shailja; SINGH, Surya Narayan. Fairness in multi-agent AI: A unified framework for ethical and equitable autonomous systems. *arXiv preprint*, arXiv:2502.07254, 2025. Available at: https://arxiv.org/abs/2502.07254. Accessed on: 22 Feb. 2025.
Understanding the Abstract:			
This paper provides a comprehensive overview of fairness in multi-agent AI, introducing a new framework that integrates fairness constraints, bias mitigation strategies, and incentive mechanisms to align agents' autonomous behaviors with social values, balancing efficiency and robustness.			
Understanding the Discussion/Conclusion:			
This paper emphasizes the objective of creating a collaborative environment between researchers to mitigate biases in the actions of multi-agent systems, promoting responsibility and transparency. Furthermore, the study highlights the need for techniques that minimize bias and ensure that agents act more fairly. Finally, the research was conducted using modification of the reward system as a strategy to mitigate unwanted actions by multi-agents.			

Report 13			
Paper name:	Agentic AI: Autonomous Intelligence for Complex	Reference:	ACHARYA, Deepak Bhaskar; KUPPAN, Carthigeyan; DIVYA, B. Agentic AI: Autonomous intelligence for complex goals –

	Goals—A Comprehensive Survey		A comprehensive survey. *IEEE Access*, 2025. Available at: https://ieeexplore.ieee.org/document/10849561. Accessed on: 10 Feb. 2025.
Understanding the Abstract:			
The study explores the fundamental concepts, unique characteristics and core methodologies that drive agent development. Furthermore, it discusses its applications in areas such as healthcare, finance and adaptive software, highlighting the advantages of implementing agent systems in real-world scenarios. The study also addresses the ethical challenges related to this technology, proposing solutions to issues such as objective alignment, resource constraints and adaptability to the environment.			
Understanding the Discussion/Conclusion:			
In the “Conclusion” section, this paper highlights the different facets of AI agents, addressing their concepts, applicability and challenges. Furthermore, it emphasizes the broad usability of these systems in different scenarios, but also highlights their limitations. Finally, it warns of the need for more robust governance in order to strengthen ethics in the application of this technology.			

Report 14			
Paper name:	Personalized Recommendation Systems using Multimodal, Autonomous,	Reference:	THAKKAR, Param; YADAV, Anushka. *Personalized Recommendation Systems using Multimodal, Autonomous, Multi Agent Systems*. arXiv preprint

	Multi Agent Systems		arXiv:2410.19855, 2024. Available at: https://arxiv.org/abs/2410.19855. Accessed on: 22 Feb. 2025.
Understanding the Abstract:			
The paper describes a personalized recommendation system using multimodal and multi-agent systems to improve the customer experience in e-commerce. The system is made up of three agents: the first recommends products, the second asks follow-up questions based on images and the third performs an autonomous search.			
Understanding the Discussion/Conclusion:			
The study demonstrates that there was collaboration between agents in a multi-agent system to assist users with product recommendations. It is important to highlight that the data distribution used went beyond the customer's history, also incorporating the use of images.			

Report 15			
Paper name:	Consumer reactions to technology in retail: choice uncertainty and reduced perceived control in decisions assisted by recommendatio	Reference:	ROHDEN, Simoni F.; ESPEARTEL, Lélis Balestrin. Consumer reactions to technology in retail: choice uncertainty and reduced perceived control in decisions assisted by recommendation agents. *Electronic Commerce Research*, v. 24, no. 2, p. 901-923, 2024. Available at: https://link.springer.com/article/10

	n agents		.1007/s10660-024-09808-7. Accessed on: 22 Feb. 2025.
Understanding the Abstract:			
The research highlights that recommendation agents can reduce choice overload and facilitate purchasing decisions, but they also generate greater uncertainty in decision making. Purchases assisted by these agents are perceived as more uncertain, with less perceived control over choices, resulting in lower satisfaction and purchase intentions.			
Understanding the Discussion/Conclusion:			
The present study highlights the positive effects of using agents in reducing the user's cognitive load when choosing and browsing products. However, experiments indicate that this technology can increase the user's perception of uncertainty regarding recommendations.			

Report 16			
Paper name:	Investigating Bias in LLM-Based Bias Detection: Disparities between LLMs and Human Perception	Reference:	LIN, Lean; WANG, Lingzhi; GUO, Jinsong; WONG, Kam-Fai. Investigating bias in LLM-based bias detection: disparities between LLMs and human perception. *arXiv preprint*, arXiv:2403.14896, 2024. Available at: https://arxiv.org/abs/2403.14896. Accessed on: 22 Feb. 2025.

Understanding the Abstract:
In this research, although robust large language models (LLMs) have emerged as fundamental tools for bias prediction, concerns persist about the biases inherent in these models. Furthermore, the presence and nature of bias in LLMs and its consequent impact on the detection of bias in the media were investigated.
Understanding the Discussion/Conclusion:
The text emphasizes the presence of bias in LLM models and insists on the urgency of policies, guidelines and governance to mitigate this problem.

Report 17			
Paper name:	Bias of AI-generated content: an examination of news produced by large language models	Reference:	FANG, X.; CHE, S.; MAO, M. et al. Bias of AI-generated content: an examination of news produced by large language models. *Sci Rep*, v. 14, p. 5224, 2024. Available at: https://doi.org/10.1038/s41598-024-55686-2. Accessed on: 22 Feb. 2025.
Understanding the Abstract:			
The study investigates gender and racial bias in AIGC produced by seven LLMs, including ChatGPT and LLaMA, using news articles from The New York Times and Reuters. Research reveals that LLMs demonstrate substantial biases, especially against women and individuals of color. ChatGPT has the lowest level of bias and is the only model capable of opting out of generating content with biased prompts.			

Understanding the Discussion/Conclusion:
The text emphasizes the presence of bias in LLM models, showing that the AIGC (AI-Generated Content) produced by these models presents gender and racial biases at different levels. The effectiveness of RLHF (Reinforcement from Human Feedback) in mitigating these biases stands out.

<table>
<tr><th colspan="4">Report 18</th></tr>
<tr><td>Paper name:</td><td>A Multi-Agent Conversational Recommender System</td><td>Reference:</td><td>FANG, Jiabao; GAO, Shen; REN, Pengjie; CHEN, Xiuying; VERBERNE, Suzan; REN, Zhaochun. A multi-agent conversational recommender system.*arXiv preprint*, arXiv:2402.01135, 2024. Available at: https://arxiv.org/abs/2402.01135. Accessed on: 22 Feb. 2025.</td></tr>
<tr><td colspan="4">Understanding the Abstract:</td></tr>
<tr><td colspan="4">The paper proposes the Multi-Agent Conversational Recommendation System (MACRS), which improves the dialogue flow and collection of user preferences. MACRS uses a multi-agent cooperative framework to generate and choose appropriate responses and a reflection mechanism to adjust dialogue planning based on user feedback.</td></tr>
<tr><td colspan="4">Understanding the Discussion/Conclusion:</td></tr>
<tr><td colspan="4">The study demonstrates the techniques used to improve the user recommendation</td></tr>
</table>

approach, adopting a multi-agent system, in which each agent is responsible for a part of the dialogue strategy, with the support of LLM models. Furthermore, a mechanism for continuous user feedback and integration of user information was employed to increase agent accuracy.

Report 19			
Paper name:	Poisoning and Backdooring Contrastive Learning	Reference:	CARLINI, Nicholas; TERZIS, Andreas. Poisoning and backdooring contrastive learning. *arXiv preprint*, arXiv:2106.09667, 2021. Available at: https://arxiv.org/abs/2106.09667. Accessed on: 10 Feb. 2025.
Understanding the Abstract:			
In this study, even poisoning just 0.01% of a dataset, it was shown that it is possible to induce the model to make errors, raising questions about the feasibility of training with uncurated data from the internet.			
Understanding the Discussion/Conclusion:			
The study shows how using unfiltered datasets can increase the risks of poisoning attacks on machine learning models. He explains that modern models train with large volumes of data taken from the Internet, without rigorous review, which makes it easier for adversaries to insert malicious information. Researchers have demonstrated that these attacks can be done with less effort than traditional methods and that increasing the amount of data does not prevent attacks. To solve this problem, the study suggests that new forms of defense be developed, as manually reviewing all the data is not viable.			