A Review on Large Language Models and Generative AI in Banking

Daniel Staegemann[1][a], Christian Haertel[1][b], Christian Daase[1][c], Matthias Pohl[2][d], Mohammad Abdallah[3][e] and Klaus Turowski[1][f]

[1]*Magdeburg Research and Competence Cluster VLBA, Otto-von-Guericke University Magdeburg, Magdeburg, Germany*
[2]*Institute of Data Science, German Aerospace Center (DLR), Jena, Germany*
[3]*Department of Software Engineering, Al-Zaytoonah University of Jordan, Amman, Jordan*

Keywords: Generative AI, GenAI, Large Language Model, LLM, Literature Review, Banking, Finance.

Abstract: Since ChatGPT was presented to the public in 2022, generative artificial intelligence and especially large language models (LLM) have attracted a lot of interest in academia and industry alike. One of the arguably most interesting domains in that regard is banking. This is because it could, theoretically, heavily benefit from their application but also brings very strict regulations and demands. To provide an overview of the current state of research in this field of tension, a literature review across four major scientific databases was conducted and the identified papers were analysed to, inter alia, determine, which types of studies are common, for which tasks the use of LLMs is explored, and which challenges and concerns became apparent. Further, the findings are discussed and some general observations are made.

1 INTRODUCTION

Since the release of ChatGPT to the public in 2022, large language models (LLM) and generative artificial intelligence (GenAI) have attracted increasing interest inside and outside of academia (Chang et al. 2024; Raiaan et al. 2024). With their ability to produce complex outputs based on a provided prompt, many see them as a promising avenue to significantly increase productivity across numerous domains (Brynjolfsson et al. 2023; Filippucci et al. 2024; Simons et al. 2024).

However, despite their great potential, they also suffer from significant drawbacks. Besides the challenge of providing suitable prompts to obtain the best possible output, one of the arguably biggest issues of GenAI and LLMs is the correctness of the created output. Their trustworthiness can especially suffer due to the so-called *hallucinations* (Huang et al. 2024; Perković et al. 2024). These occur when the models make up information or references yet present them as based on existing facts. While for some cases (e.g., suggesting suitable formulations for writing an email or giving the synopsis of a movie) this issue is rather negligible, in other scenarios (e.g., in medical settings or the legal domain) this can be highly problematic. A domain where the use of GenAI could potentially yield tremendous benefits, since huge numbers of transactions and activities have to be processed quickly, yet the significance of errors and inaccuracies is high, is banking.

To explore the potential of implementing such solutions as well as the accompanying challenges is the goal of this work, for which a structured literature review (SLR) will be conducted. Therefore, within this paper, the following research question (RQ) shall be answered:

RQ: What is the current state of incorporating GenAI, respectively LLMs, in banking, according to the scientific literature?

[a] https://orcid.org/0000-0001-9957-1003
[b] https://orcid.org/0009-0001-4904-5643
[c] https://orcid.org/0000-0003-4662-7055
[d] https://orcid.org/0000-0002-6241-7675
[e] https://orcid.org/0000-0002-3643-0104
[f] https://orcid.org/0000-0002-4388-8914

Staegemann, D., Haertel, C., Daase, C., Pohl, M., Abdallah, M. and Turowski, K.
A Review on Large Language Models and Generative AI in Banking.
DOI: 10.5220/0013472600003956

In *Proceedings of the 7th International Conference on Finance, Economics, Management and IT Business (FEMIB 2025)*, pages 267-278
ISBN: 978-989-758-748-1; ISSN: 2184-5891

To answer the RQ, the remainder of this publication will be structured as follows. After the introduction, the review protocol that was followed for the SLR is presented, which is ensued by a section that is dedicated to the description of the papers that were found in the search. Based on these, a discussion follows. Finally, a conclusion is given, and avenues for future work are outlined.

2 REVIEW PROTOCOL

To answer the RQ, a SLR was conducted. Since the value of a SLR largely depends on its rigour and reproducibility (Kraus et al. 2022; vom Brocke et al. 2009), before starting the search process, following common practices (Okoli 2015; vom Brocke et al. 2015), a protocol was developed to guide the process. The prescribed steps, as well as the corresponding considerations and results, are described in the following.

To identify potentially suitable literature, *Scopus*[1] and *IEEE Xplore*[2] (IEEE) were harnessed. While the former was chosen due to its comprehensive coverage across many scientific databases and publishers, the latter was added because of IEEE's significance in the computer science domain. These were complemented by the ACM Digital Library[3] (ACM), belonging to the world's largest computing society (ACM History Committee 2025), and the AIS electronic Library[4] (AISeL), which, inter alia, contains the proceedings of some of the premiere conferences in the information systems domain.

All of them were queried with the same search term that consisted of two components.

The first part aims at making sure that the concept of GenAI is covered with a broad range of common terms and spellings to assure comprehensiveness:

llm OR "large language model" OR "generative ai" OR "gen ai" OR genai OR gpt

While there are several other LLMs besides ChatGPT in use, it currently is the most popular one, which is why it was explicitly included in the term, whereas others were not.

In the second part, the banking domain is addressed. However, for the purpose of this paper, only "core" activities are considered, therefore, related activities such as stock market trading are not included. Thus, the corresponding term was as follows:

bank OR credit OR lend* OR financ* OR fintech*

These two parts were connected with an AND, to make sure that both aspects are significant in the found papers. Moreover, to further strengthen the focus, they had to appear in the *document title*.

Hence, the final search term, as used in Scopus, with the others using the same parameters, was as follows:

(TITLE (llm OR "large language model" OR "generative ai" OR "gen ai" OR genai OR gpt) AND TITLE (bank OR credit OR lend OR financ* OR fintech))*

To ensure the necessary quality, only *conference papers* and *journal articles* were included, whereas book chapters were not, since the latter are usually not peer-reviewed. This is also the reason why, despite their timeliness, which is especially relevant in a quickly emerging domain like GenAI, preprint services like arXiv[5] were not utilised as additional sources of papers, since there are "concerns about the research accuracy, quality, and credibility of preprints" (Adarkwah et al. 2024).

Moreover, only papers that were written in English were considered, which led to the exclusion of one paper that was written in Chinese. This decision was made since none of the authors possesses the necessary language skills to adequately analyse it and the use of AI-based tools for translation could potentially lead to misrepresentations of the content that could not be detected.

Based on the aforementioned stipulations, the search in Scopus resulted in the identification of a total of 87 papers, of which 29 were journal articles and 58 from conferences. IEEE, in turn, yielded 24 papers, 2 from journals and 22 from conferences. Through ACM, 1 journal article and 21 conference papers were found, and AISeL contributed 1 additional journal article. Thus, overall, the keyword search brought 134 items, with 33 from journals and 101 from conferences.

However, since multiple databases were used for the search, several duplicates occurred that were removed in the next step. After doing so, 109 items remained, with 31 being from journals and 78 from conferences. Naturally, not each of these papers fit the intended scope, which made additional filtering necessary. Aligned with common practices (vom Brocke et al. 2015), this was performed in multiple steps to assure a high degree of diligence while still maintaining efficiency.

[1] https://www.scopus.com
[2] https://ieeexplore.ieee.org
[3] https://dl.acm.org/
[4] https://aisel.aisnet.org/
[5] https://arxiv.org/

For all of these phases, a joint set of inclusion and exclusion criteria, as depicted in Table 1, was defined in advance to serve as the foundation of the filter process. Hereby, for a paper to be deemed suitable, each of the inclusion criteria had to be met, whereas when at least one of the exclusion criteria applied, it was removed from the set.
As already highlighted in the description of the keyword search, to be included, a paper had to be published either in the proceedings of a scientific conference or in a scientific journal. Further, it has to be written in English. Due to this work's RQ, it also has to focus on the banking sector. This is further sharpened by the exclusion of adjacent or auxiliary activities such as stock market prediction, respectively trading, or the automated analysis of financial documents. Moreover, primarily technical considerations (e.g., benchmarking and performance comparisons of different tools) or papers describing (the creation of) datasets for training or benchmarking purposes were also excluded. The same applied when the focus was not on banking itself, but it was merely used as a domain to research something else (e.g., the use of LLMs in the development/testing for banking software). It also led to exclusion if a paper rather generically addressed aspects such as LLMs' impact on organisations, matters of acceptance/trust, or privacy considerations. Instead, only papers that provide insights into (potential) application scenarios of LLMs in banking were sought after.

In the first step, the papers were filtered based on their title. Whenever it clearly indicated that a publication does not fit the intended scope, it was excluded from the list. Following this, 26 journal articles and 66 conference papers were left for consideration.

Yet, since titles have a rather limited capability of conveying a paper's content, the former measure could not be handled too strictly, which is why, afterward, the abstracts and keywords were consulted to further narrow down the considered literature. For instance, since the automated analysis of financial documents and reports to populate database tables with their content is an auxiliary activity, corresponding papers were not further regarded. This step reduced the number of remaining papers significantly, to 27, of which 8 were journal articles and 19 conference papers.

Finally, to further assess the suitability of the papers to the RQ's scope, they were read in total, and those that did not fit were excluded. In doing so, 13 more papers across were dropped. Thus, the final literature set comprises 14 papers, of which 9 are from conferences and 5 appeared in journals.

Table 1: The search's inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
The paper is published in the proceedings of a scientific conference or in a scientific journal	The paper is not actually focused on banking but only uses it as an example or as a mean to research something else
The paper is written in English	The paper focusses on stock market predictions or trading
The paper focusses on the application of LLMs in banking	The paper primarily deals with the automated analysis of financial reports
The paper discusses application scenarios or (potential) use cases	The paper primarily focusses on technical considerations (e.g., benchmarking and comparisons of tools)
	The paper just presents a new banking focussed data set for training or benchmarking purposes
	The paper rather generically addresses aspects such as LLMs' impact on organisations, matters of acceptance/trust, or privacy considerations without discussing actual application scenarios

3 FINDINGS

Resulting from the search, 14 publications were identified that, despite no time frame being specified to limit the extent of the search, all emerged in 2023 or 2024. This is, however, not surprising since the tremendous interest in LLMs only started rather recently, and the search was carried out too early for many papers from 2025 already being available for consideration.

An overview of the identified papers is given in Table 2. There, besides the title, publication year, reference, and publication type, it is also depicted where the paper was found, and an ID for further use within the publication at hand is assigned. The latter allows it to refer to specific papers in a more convenient fashion, which will be used within Table 3 but adds no value beyond that. The former, however, reveals an interesting insight that is also important outside of the scope of this particular literature review. Namely, it emphasises the importance of not only querying one single database but instead multiple when attempting to create a comprehensive picture of the domain, since there is no single all-encompassing source.

Table 2: The identified papers.

ID	Title	Year	Type	Found in	Reference
1	Applications of Generative AI in Fintech	2023	Conference	Scopus	(Barde and Kulkarni 2023)
2	A Study On Generative Ai And Its Impact On Banking And Financial Services Sector: Data Privacy & Sustainable Perspective	2023	Conference	IEEE Xplore, Scopus	(Ramaswamy and Bagrecha 2023)
3	Enhancing Credit Risk Reports Generation using LLMs: An Integration of Bayesian Networks and Labeled Guide Prompting	2023	Conference	ACM Digital Library	(Teixeira et al. 2023)
4	From fiction to fact: the growing role of generative AI in business and finance	2023	Journal	Scopus	(Chen et al. 2023)
5	LLMs for Financial Advisement: A Fairness and Efficacy Study in Personal Decision Making	2023	Conference	ACM Digital Library	(Lakkaraju et al. 2023)
6	AI versus AI in Financial Crimes & Detection: GenAI Crime Waves to Co-Evolutionary AI	2024	Conference	ACM Digital Library	(Kurshan et al. 2024)
7	An Intelligent LLM-Powered Personalized Assistant for Digital Banking Using LangGraph and Chain of Thoughts	2024	Conference	IEEE Xplore	(Easin et al. 2024)
8	Bankruptcy Prediction: Data Augmentation, LLMs and the Need for Auditor's Opinion	2024	Conference	ACM Digital Library	(Sideras et al. 2024)
9	Credit scoring model for fintech lending: An integration of large language models and FocalPoly loss	2024	Journal	Scopus	(Xia et al. 2024)
10	Empowering financial futures: Large language models in the modern financial landscape	2024	Journal	Scopus	(Cao et al. 2024)
11	Enhancing Graph Database Interaction through Generative AI-Driven Natural Language Interface for Financial Fraud Detection	2024	Conference	IEEE Xplore, Scopus	(Simran and Geetha 2024)
12	Generative AI in Shariah Advisory in Islamic Finance: An Experimental Study	2024	Journal	AIS electronic Library	(Jokhio and Jaffer 2024)
13	LLMs in Banking: Applications, Challenges, and Approaches	2024	Conference	ACM Digital Library	(Fan 2024)
14	New Paradigm for Economic and Financial Research With Generative AI: Impact and Perspective	2024	Journal	IEEE Xplore, Scopus	(Zheng et al. 2024)

The first paper, *Applications of Generative AI in* The order of the papers within the table is purely based on the publication year and the alphabetical order of the titles and holds no further meaning. However, in the following introduction of these papers, it will still be adhered to, to increase clarity.

The first paper, *Applications of Generative AI in Fintech* (Barde and Kulkarni 2023) aims to provide an overview of the different ways that GenAI can be incorporated by companies that are active in the financial technology sector. What is really noteworthy, however, is that it particularly focuses on its use by specific institutions. Thus, it compiles valuable insights into how global leaders such as, inter alia, Bloomberg, Goldman Sachs, and Wells Fargo harness the new opportunities to advance their operations.

While not the only content of the presented work, the arguably most relevant part with regard to this paper of *A Study On Generative Ai And Its Impact On Banking And Financial Services Sector: Data Privacy & Sustainable Perspective* (Ramaswamy and Bagrecha 2023) is the conducted survey amongst a somewhat heterogeneous group of participants (even though with a strong emphasis on employees) from India to explore their opinions on and sentiment towards GenAI in banking.

The development of a prompt-engineering method (referred to as “Labeled Guide Prompting”) is described in *Enhancing Credit Risk Reports Generation using LLMs: An Integration of Bayesian Networks and Labeled Guide Prompting* (Teixeira et al. 2023). Here, it is demonstrated how ChatGPT can be used to create high-quality credit risk reports when provided with suitable examples and appropriate structure and guidance through the prompt. For the evaluation, data from credit applications was used, and human credit analysts were tasked to assess the quality of LLM and human-generated reports in a blinded setting. Hereby, the LLM reports were usually preferred, highlighting the approach’s potential.

A non-exhaustive overview of different types of tasks for GenAI related to banking is given in *From fiction to fact: the growing role of generative AI in business and finance* (Chen et al. 2023). For each of them, identical requests are sent to *ChatGPT 3.5*, *ChatGPT 4*, and Google’s *Bard*, and the corresponding responses are shown and compared. Further, the paper also comprises a case study for sentiment analysis and contains considerations regarding ethical concerns, technical limitations, and legal aspects.

Another comparison of the suitability of ChatGPT and Bard for certain financial tasks was presented in *LLMs for Financial Advisement: A Fairness and Efficacy Study in Personal Decision Making* (Lakkaraju et al. 2023). However, this time, the focus was on the advisement of customers on credit card-related questions. Hereby, not only general requests to deliver information were considered, but also how specific scenarios should be handled, which required the assistants to perform mathematical calculations and compare the parameters of different products. Furthermore, it was also examined if the (likely; based on the name) gender or ethnicity of the user impacted the provided answer. This, in turn, adds an important aspect to the overall discourse, since avoiding such biases is an important duty when developing automated systems.

How GenAI can be harnessed to combat financial crime but also which challenges are encountered in this endeavour, and in which ways it can be abused by criminals is addressed in *AI versus AI in Financial Crimes & Detection: GenAI Crime Waves to Co-Evolutionary AI* (Kurshan et al. 2024). Even though the latter is not within the scope of this study, it is still highly important for all actors in the financial system to be aware of the potential exploits and the associated risks, to reduce the likelihood of falling victim to them. Hereby, the paper provides a high-level overview that can constitute a valuable starting point for further research into the respective areas that are most relevant for one’s situation.

Another study that deals with the use of LLMs as personal banking assistants is *An Intelligent LLM-Powered Personalized Assistant for Digital Banking Using LangGraph and Chain of Thoughts* (Easin et al. 2024). However, in contrast to the study of (Lakkaraju et al. 2023), here, instead of credit card consulting, support with general banking activities (e.g., adding money or paying bills) is targeted. For this, at first, a single-agent system was proposed, which was later amended by the development of a multi-agent architecture. In both cases, the customer interacts with a single virtual assistant, which assures the convenience of using it. However, whereas in the first approach, the assistant accesses all the relevant tools, in the second one, instead, it communicates with another set of agents, of which each is specifically developed to handle one distinct task. It then gets passed the results and presents them to the user. This way, a high degree of modularity and specialisation can be achieved, similar to, for instance, a microservice architecture (Shakir et al. 2021), while not negatively impacting usability.

An approach to improve bankruptcy predictions through the use of LLMs is presented in *Bankruptcy Prediction: Data Augmentation, LLMs and the Need for Auditor's Opinion* (Sideras et al. 2024). Here, it is suggested to incorporate the opinions of auditors that is included in financial reports as an additional input to the prediction algorithm that shall determine if a company will go bankrupt in the foreseeable future. Hereby, one challenge was the small percentage of companies that actually go bankrupt, leading to a heavily imbalanced distribution of the data. To deal with this issue, LLMs were harnessed to generate realistic synthetic data. Besides this data augmentation, also the idea of directly tasking LLMs with making corresponding predictions is explored. However, while LLMs have been found to perform well in many different tasks (Chang et al. 2024), here the performance was not deemed sufficient. Yet, this does not necessarily mean that the general idea is unsuitable. Potentially, more sophisticated prompts or future improvements in the LLMs might yield better results.

The idea of incorporating narrative data into the decision-making process is also explored in *Credit scoring model for fintech lending: An integration of large language models and FocalPoly loss* (Xia et al. 2024). This time, however, the focus is on credit scoring. Within the paper, several LLMs are compared regarding their ability to extract valuable information that can improve the accuracy of the risk prediction. Hereby, the authors found that incorporating the LLMs indeed increased the performance. Furthermore, they also showed that using a LLM tailored to the language of the use case (in this case Chinese) can lead to better performance compared to, for instance, ChatGPT, which is primarily trained on English sources.

In *Empowering financial futures: Large language models in the modern financial landscape* (Cao et al. 2024), an overview of numerous potential application areas of LLMs in the financial sector in general is given, of which many are also relevant when it specifically comes to banking. Additionally, several challenges are discussed. While these are not necessarily just applicable to the financial sector, due to its critical and impactful nature, they are especially significant and, thus, need to be addressed appropriately.

Another attempt at dealing with financial fraud is shown in *Enhancing Graph Database Interaction through Generative AI-Driven Natural Language Interface for Financial Fraud Detection* (Simran and Geetha 2024). Here, a pipeline is built that simplifies the analysis by allowing the user to control the application with natural language requests via a web interface, significantly increasing user-friendliness. These are then transformed into a query and forwarded to a database to retrieve the relevant data. Subsequently, a LLM is provided with the data, analyses them, and predicts if a transaction is fraudulent. The results are then shown to the user. Furthermore, for the LLM, different alternatives are compared regarding their performance.

A rather unique, yet very interesting, case is presented in *Generative AI in Shariah Advisory in Islamic Finance: An Experimental Study* (Jokhio and Jaffer 2024). To guide the decisions of banks that have to or aspire to comply with shariah regulations, experts are needed that are well-versed in both domains, shariah regulations and banking. Yet, this particular combination is relatively rare, potentially creating a corresponding bottleneck. Aiming to alleviate this issue, the authors explored how feasible the use of (different) LLMs is to identify shariah compliance issues, provide corresponding references from the shariah, and give guidance on how to proceed.

Another overview that highlights how LLMs can support banking operations is given in *LLMs in Banking: Applications, Challenges, and Approaches* (Fan 2024). Here, various application avenues are outlined and, using real world examples, it is highlighted how these can bring tangible business value. Moreover, similar to several of the priorly introduced papers, potential challenges are discussed, and potential mitigation strategies are mentioned. Additionally, brief strategic recommendations are given for banks that intend to utilize LLMs in their operations.

Finally, an outlook on research in the field of GenAI application in finance is given in *New Paradigm for Economic and Financial Research With Generative AI: Impact and Perspective* (Zheng et al. 2024). While the focus is somewhat different from the other papers and not directly aimed at the incorporation of LLMs in banking but instead on the scientific side, it also prominently discusses potential application areas as well as challenges that have to be considered. Thus, it contributes to the corresponding discourse.

4 DISCUSSION

Even though the focus of this review was intentionally kept rather narrow, the versatility of GenAI in the banking sector still shows in the plethora of different tasks and approaches that are

Table 3: Overview of the presented papers.

ID	Type of Research Results	Addressed Area	What Was Done?	Tasks Mentioned	Challenges/ Concerns	Used Model(s)	Prompting
1	Overview	The financial sector in general	An overview of diverse applications scenarios of GenAI was given	Credit risk evaluations; Customer service operations; Banking operations; Data analysis	Biases	Not applicable	Not applicable
2	Survey	Banking in general	Survey on opinions/ sentiment regarding GenAI	Customer service operations; Financial planning	Data privacy; User acceptance	Not applicable	Not applicable
3	Specific development: Prompting strategy	Credit risk analysis	Development and evaluation of a prompting strategy	Generation of credit risk reports	The prioritization of ChatGPT in presenting information	GPT-4	Prompts shown in parts; "Labeled Guide Prompting" proposed; Few-shot prompting applied
4	Overview; Case study	The financial sector in general	Comparison of GPT 3.5, GPT 4, and Bard for different tasks; Sentiment analysis case study	Customer service operations; Risk management; Decision support	Data privacy; Lack of legislation; Quality of responses; Overreliance; Sensitivity to prompting template; Energy consumption; Impact on labour market	GPT-3.5; GPT-4; Bard	Not shown
5	Specific development: Application scenario	LLM as banking assistant	Feasibility of LLM-chatbots as assistant/ advisor for rather challenging tasks tested	Customer service operations; Financial advisor	Biases	ChatGPT (version not stated); Bard	Not shown
6	Overview	Fraud detection	An overview on GenAI-based crimes and opportunities for crime detection through LLMs was given	Fraud detection; money laundering detection	Potential of LLMs for use in criminal activities	Not applicable	Not applicable

Table 3: Overview of the presented papers (cont.).

ID	Type of Research Results	Addressed Area	What Was Done?	Tasks Mentioned	Challenges/ Concerns	Used Model(s)	Prompting
7	Specific development: Application	LLM as banking assistant	Development of a (multi-agent) personalized assistant for digital banking	Various banking tasks (e.g., add money or pay bills)	Not mentioned	GPT-3.5	Short prompts shown; Chain of Thoughts prompting mentioned
8	Specific development: Application scenario	Bankruptcy prediction	Use of LLM to predict if a company will go bankrupt based on auditor's opinion in a report	Narrative extraction to improve bankruptcy prediction; Use of LLM to predict bankruptcy	Low quality of LLM predictions	Llama-3; Finance-chat (fine-tuned Llama-2 model)	Prompts shown; Zero-shot prompting
9	Specific development: Application scenario	Credit risk analysis	Extraction of narrative data from credit report to enhance credit risk assessment model	Extraction of narrative data	Data security/ privacy; Information extraction capability may be language-dependent	GPT-4; GPT-3.5; Bert; ERNIE 4.0; Turbo; Doubao	Not mentioned
10	Overview	The financial sector in general	An overview of diverse applications potentials of LLMs in finance as well as challenges was given	Customer service operations; Fraud detection/ prevention; Market analysis; Financial advisor; Regulatory compliance; Legal document analysis; Data analysis	Biases; Ethical considerations; Data security/ privacy; Quality of responses; User acceptance	Not applicable	Not applicable
11	Specific development: Application scenario	Fraud detection	Automated conversion of natural language into graph database queries to make fraud detection tasks more accessible; Fraud prediction by LLM	Conversion of natural language into database queries; Fraud detection	Scalability challenges with increasing transaction volumes impacting real-time processing	T5 model; Llama-2; FinBERT; RoBERTa; DistilBERT	Not mentioned

Table 3: Overview of the presented papers (cont.).

ID	Type of Research Results	Addressed Area	What Was Done?	Tasks Mentioned	Challenges/ Concerns	Used Model(s)	Prompting
12	Specific development: Application scenario	Policy adherence support	Evaluation of the capacity of generic LLMs to provide shariah advisory in Islamic finance based on ten hypothetical financing scenarios	Identify shariah compliance issues; Provide the corresponding sharia references; Offer shariah guidance on handling the issues	Limitations in providing shariah guidance	GPT-4; Gemini; Meta AI	Not mentioned
13	Overview	Banking in general	An overview of diverse applications potentials of LLMs in banking as well as challenges was given	Customer acquisition and relationship management; Account management; Customer service operations; Loans and credit management; Investment and wealth management; Regulatory compliance; Risk management	Data privacy/ security; Biases; Interpretability and transparency; Technical challenges; Maintenance	Not applicable	Not applicable
14	Overview	The financial sector in general	An overview of diverse applications potentials of LLMs in finance was given	Fraud detection; Policy analysis; Extreme scenario analysis; Economic and financial predictions; Portfolio management	Data privacy/ security; Biases; Ethical considerations; Quality of the results; Transparency; Dependence on major technology corporations; Impact on labour market	Not applicable	Not applicable

presented in the identified papers. To provide a comprehensive overview of their contents, in Table 3, a matrix is shown that summarizes the most important aspects (Webster and Watson 2002).

This comprises firstly the general type of research results that were obtained, which area was addressed, and a brief summary of what was actually done with regards to this study's scope. Moreover, it is depicted, which tasks for LLMs were mentioned and which challenges and concerns related to the use of GenAI and LLMs in banking were highlighted.

Finally, for those cases where it was applicable and stated, it is noted which LLMs were used in the described research endeavour and which prompting strategies were applied.

- Auffällig, dass wenig/kein ChatGPT. Erklärung: Daten sind sensibel

When looking at the type of research results, it becomes apparent that many of the papers attempt to provide an overview of the application potentials. This emphasises that the novelty of the domain goes along with a great sense of uncertainty and exploration regarding the potential of this technology. Whereas more established topics are usually advanced by specific developments and theories that add incremental knowledge, here, just understanding its actual significance is already a challenge in its own.

Yet, none of the aforementioned papers is a structured literature review, highlighting the significance of the study at hand in providing a more systematised overview of the domain.

The current lack of maturity is also emphasized when scrutinizing the specific developments, be it tools, prompting strategies, or further attempts at exploring potential application scenarios.

Initially, it was intended to add another column to the table to indicate if the specific developments were evaluated in real-life scenarios or in an experimental way. Yet, after analysing the literature, it was found that all of them took place in experimental settings, and not a single one was already (at least at the time these papers were written) used productively. This is, however, not surprising, factoring in the lacking maturity of the technology in combination with the critical nature, strict regulations, and high demands of the banking industry as well as the competitive advantages that can be achieved through corresponding solutions that are superior compared to the competition's ones. Nevertheless, describing the use of LLMs in real-world settings, as already to some degree done in (Barde and Kulkarni 2023), could provide valuable additional insights and would most likely be appreciated by many.

The general description of (potential) tasks for LLMs is, however, done plentiful across the identified papers. One of the most frequently mentioned ones is the dealing with customer service operations, respectively, the role of personalized assistants. The use of LLMs as financial advisors or planners was also frequently mentioned. However, this would, naturally, require highly sophisticated and trustworthy solutions, yet, currently, the public's trust in AI for those tasks is rather limited (Ramaswamy and Bagrecha 2023).

Other popular tasks include the extraction of information to, for instance, amend actually existing processes and varying prediction tasks. Hereby, especially credit risk assessment and fraud detection or prevention seem to be popular research directions. Here, some of the results are surprisingly impressive (Simran and Geetha 2024), indicating that LLMs are already very competent in this field.

Increasing accessibility by acting as an easy-to-use interface, for instance for the use of databases (Simran and Geetha 2024), also appears as a promising approach. Moreover, the creation of realistic synthetic (text-based) data, which can be used for varying purposes such as testing or the training of algorithms (Staegemann et al. 2023), is also a strength of LLMs.

The final big group of tasks that stood out in the identified papers comprised the analysis of policies, the analysis of legal texts, and the provisioning of guidance on related matters. Even though the corresponding quality is not yet sufficient to replace the respective experts (Jokhio and Jaffer 2024), providing some support can already bring significant benefits.

Nevertheless, there are also considerable challenges associated with the use of LLMs in general and especially in the banking sector. The ones that are mentioned the most are the threat of biases influencing the results, and issues regarding data privacy and security as well as transparency. Ethical considerations and a potentially negative impact on the labour market are also stated. Another big concern is, as mentioned earlier, the quality of the results that is oftentimes insufficient for productive use in critical tasks. Consequently, trust, respectively a lack of it, as also highlighted before, is, therefore, another big barrier for LLMs in many finance-related roles. Additionally, as to be expected for a rather new type of tool, technical challenges are also a big factor that needs to be dealt with.

While many other obstacles are also pointed out, a major one is the legal situation around LLMs and their use. This is not restricted to the financial sector and also applies to many other areas (Barqawi and Abdallah 2024), but is, naturally, especially significant in such a strictly regulated domain.

Even though this might not be a challenge per se, it was also experienced that language-specific LLMs outperform general ones, when dealing with other languages than English (Xia et al. 2024). This is in line with other works (Noels et al. 2024; Zhang et al. 2024) and suggests that organizations should make their model-choice under consideration of the language that the LLM shall operate in, or potentially even run several specialized LLMs that are addressed based on the language relevant to the respective request. This way, one LLM could be used as the point of contact and forward the requests to the underlying LLM most suited for the task and/or language. This would be similar to the solution suggested in (Easin et al. 2024).

Currently, however, the use of language-specific LLMs is still rather rare, at least based on the literature, and general LLMs are the most common ones. This is also visible in the identified papers, where ChatGPT is the most commonly found LLM. While this is not surprising, due to its popularity, it is, in contrast to other options, not specialized on tasks in the financial sector. With growing maturity of the domain, a development towards the use of more specialized models for such tasks appears to be likely.

Further, unfortunately, the low maturity of the domain also shows in a lack of standards for the reporting of LLM projects. Therefore, in many cases, relevant information such as the applied prompting strategy/strategies, the prompts themselves, or even the specific version of the LLM that was used are missing. The same applies to a more detailed breakdown of the evaluations. This, in turn, makes it harder to contextualize the findings.

5 CONCLUSION

With the use of GenAI and LLMs being in its infancy, many domains are trying to find ways to harness their power. An example with especially high stakes is the banking sector since it could hugely benefit but also brings strict regulations. To obtain an overview of the research on the use of LLMs in core tasks in banking that can be used as a starting point for future research endeavours, a structured literature review was conducted. To this end, four scientific databases were searched and the found papers were subsequently analysed to identify application scenarios, challenges and concerns, and current themes. In the future, this could be expanded by also incorporating other facets of finance such as stock trading.

REFERENCES

ACM History Committee. (2025). "ACM History," available at https://www.acm.org/about-acm/acm-history, accessed on Jan 10 2025.

Adarkwah, M. A., Islam, A. Y. M. A., Schneider, K., Luckin, R., Thomas, M., and Spector, J. M. (2024). "Are Preprints a Threat to the Credibility and Quality of Artificial Intelligence Literature in the ChatGPT Era? A Scoping Review and Qualitative Study," *International Journal of Human–Computer Interaction*, pp. 1-14 (doi: 10.1080/10447318.2024.2364140).

Barde, K., and Kulkarni, P. A. (2023). "Applications of Generative AI in Fintech," in *The Third International Conference on Artificial Intelligence and Machine Learning Systems,* Bangalore India. 25.10.2023 - 28.10.2023, New York, NY, USA: ACM, pp. 1-5 (doi: 10.1145/3639856.3639893).

Barqawi, L., and Abdallah, M. (2024). "Copyright and generative AI," *Journal of Infrastructure, Policy and Development* (8:8), p. 6253 (doi: 10.24294/jipd.v8i8.6253).

Brynjolfsson, E., Li, D., and Raymond, L. (2023). "Generative AI at Work," *NBER Working Paper Series* 31161, Cambridge, MA: National Bureau of Economic.

Cao, X., Li, S., Katsikis, V., Khan, A. T., He, H., Liu, Z., Zhang, L., and Peng, C. (2024). "Empowering financial futures: Large language models in the modern financial landscape," *EAI Endorsed Transactions on AI and Robotics* (3) (doi: 10.4108/airo.6117).

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., and Xie, X. (2024). "A Survey on Evaluation of Large Language Models," *ACM Transactions on Intelligent Systems and Technology* (15:3), pp. 1-45 (doi: 10.1145/3641289).

Chen, B., Wu, Z., and Zhao, R. (2023). "From fiction to fact: the growing role of generative AI in business and finance," *Journal of Chinese Economic and Business Studies* (21:4), pp. 471-496 (doi: 10.1080/14765284.2023.2245279).

Easin, A. M., Sourav, S., and Tamás, O. (2024). "An Intelligent LLM-Powered Personalized Assistant for Digital Banking Using LangGraph and Chain of Thoughts," in *2024 IEEE 22nd Jubilee International Symposium on Intelligent Systems and Informatics (SISY),* Pula, Croatia. 19.09.2024 - 21.09.2024, IEEE, pp. 625-630 (doi: 10.1109/SISY62279.2024.10737601).

Fan, M. (2024). "LLMs in Banking: Applications, Challenges, and Approaches," in *Proceedings of the International Conference on Digital Economy, Blockchain and Artificial Intelligence,* Guangzhou China. 23.08.2024 - 25.08.2024, New York, NY, USA: ACM, pp. 314-321 (doi: 10.1145/3700058.3700107).

Filippucci, F., Gal, P., Jona-Lasinio, C., Leandro, A., and Nicoletti, G. (2024). "The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges," *OECD Artificial Intelligence Papers,* OECD.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., and Liu, T. (2024). "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," *ACM Transactions on Information Systems* (doi: 10.1145/3703155).

Jokhio, M. N., and Jaffer, M. A. (2024). "Generative AI in Shariah Advisory in Islamic Finance: An Experimental Study," *Business Review* (19:2), pp. 74-92 (doi: 10.54784/1990-6587.1665).

Kraus, S., Breier, M., Lim, W. M., Dabić, M., Kumar, S., Kanbach, D., Mukherjee, D., Corvello, V., Piñeiro-Chousa, J., Liguori, E., Palacios-Marqués, D., Schiavone, F., Ferraris, A., Fernandes, C., and Ferreira, J. J. (2022). "Literature reviews as independent studies: guidelines for academic practice," *Review of Managerial Science* (16:8), pp. 2577-2595 (doi: 10.1007/s11846-022-00588-8).

Kurshan, E., Mehta, D., and Balch, T. (2024). “AI versus AI in Financial Crimes & Detection: GenAI Crime Waves to Co-Evolutionary AI,” in *Proceedings of the 5th ACM International Conference on AI in Finance,* Brooklyn NY USA. 14.11.2024 - 17.11.2024, New York, NY, USA: ACM, pp. 745-751 (doi: 10.1145/3677052.3698655).

Lakkaraju, K., Jones, S. E., Vuruma, S. K. R., Pallagani, V., Muppasani, B. C., and Srivastava, B. (2023). “LLMs for Financial Advisement: A Fairness and Efficacy Study in Personal Decision Making,” in *4th ACM International Conference on AI in Finance,* Brooklyn NY USA. 27.11.2023 - 29.11.2023, New York, NY, USA: ACM, pp. 100-107 (doi: 10.1145/3604237.3626867).

Noels, S., Blaere, J. de, and Bie, T. de. (2024). “A Dutch Financial Large Language Model,” in *Proceedings of the 5th ACM International Conference on AI in Finance,* Brooklyn NY USA. 14.11.2024 - 17.11.2024, New York, NY, USA: ACM, pp. 283-291 (doi: 10.1145/3677052.3698628).

Okoli, C. (2015). “A Guide to Conducting a Standalone Systematic Literature Review,” *Communications of the Association for Information Systems* (37), pp. 879-910 (doi: 10.17705/1CAIS.03743).

Perković, G., Drobnjak, A., and Botički, I. (2024). “Hallucinations in LLMs: Understanding and Addressing Challenges,” in *2024 47th MIPRO ICT and Electronics Convention (MIPRO),* Opatija, Croatia. 20.05.2024 - 24.05.2024, IEEE, pp. 2084-2088 (doi: 10.1109/MIPRO60963.2024.10569238).

Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., Ahmad, J., Ali, M. E., and Azam, S. (2024). “A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges,” *IEEE Access* (12), pp. 26839-26874 (doi: 10.1109/ACCESS.2024.3365742).

Ramaswamy, S., and Bagrecha, C. (2023). “A Study On Generative Ai And Its Impact On Banking And Financial Services Sector: Data Privacy & Sustainable Perspective,” in *2023 IEEE Technology & Engineering Management Conference - Asia Pacific (TEMSCON-ASPAC),* Bengaluru, India. 14.12.2023 - 16.12.2023, IEEE, pp. 1-5 (doi: 10.1109/TEMSCON-ASPAC59527.2023.10531592).

Shakir, A., Staegemann, D., Volk, M., Jamous, N., and Turowski, K. (2021). “Towards a Concept for Building a Big Data Architecture with Microservices,” in *Proceedings of the 24th International Conference on Business Information Systems,* Hannover, Germany/virtual. 14.06.2021 - 17.06.2021, pp. 83-94 (doi: 10.52825/bis.v1i.67).

Sideras, A., Bougiatiotis, K., Zavitsanos, E., Paliouras, G., and Vouros, G. (2024). “Bankruptcy Prediction: Data Augmentation, LLMs and the Need for Auditor's Opinion,” in *Proceedings of the 5th ACM International Conference on AI in Finance,* Brooklyn NY USA. 14.11.2024 - 17.11.2024, New York, NY, USA: ACM, pp. 453-460 (doi: 10.1145/3677052.3698627).

Simons, W., Turrini, A., and Vivian, L. (2024). “Artificial Intelligence: Economic Impact, Opportunities, Challenges, Implications for Policy,” *European Economy Discussion Papers* 210, European Union.

Simran, T., and Geetha, J. (2024). “Enhancing Graph Database Interaction through Generative AI-Driven Natural Language Interface for Financial Fraud Detection,” in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT),* Kamand, India. 24.06.2024 - 28.06.2024, IEEE, pp. 1-8 (doi: 10.1109/ICCCNT61001.2024.10725408).

Staegemann, D., Pohl, M., Haertel, C., Daase, C., Abdallah, M., and Turowski, K. (2023). “An Overview of the Approaches for Generating Test Data in the Context of the Quality Assurance of Big Data Applications,” in *2023 17th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS),* Bangkok, Thailand. 08.11.2023 - 10.11.2023, IEEE, pp. 30-37 (doi: 10.1109/SITIS61268.2023.00015).

Teixeira, A. C., Marar, V., Yazdanpanah, H., Pezente, A., and Ghassemi, M. (2023). “Enhancing Credit Risk Reports Generation using LLMs: An Integration of Bayesian Networks and Labeled Guide Prompting,” in *4th ACM International Conference on AI in Finance,* Brooklyn NY USA. 27.11.2023 - 29.11.2023, New York, NY, USA: ACM, pp. 340-348 (doi: 10.1145/3604237.3626902).

vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., and Cleven, A. (2009). “Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process,” in *Proceedings of the ECIS 2009,* Verona, Italy. 08.06.2009 - 10.06.2009.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. (2015). “Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research,” *Communications of the Association for Information Systems* (37) (doi: 10.17705/1CAIS.03709).

Webster, J., and Watson, R. T. (2002). “Analyzing the Past to Prepare for the Future: Writing a Literature Review,” *MIS Quarterly* (26:2), pp. xiii-xxiii.

Xia, Y., Han, Z., Li, Y., and He, L. (2024). “Credit scoring model for fintech lending: An integration of large language models and FocalPoly loss,” *International Journal of Forecasting* (doi: 10.1016/j.ijforecast.2024.07.005).

Zhang, X., Xiang, R., Yuan, C., Feng, D., Han, W., Lopez-Lira, A., Liu, X.-Y., Qiu, M., Ananiadou, S., Peng, M., Huang, J., and Xie, Q. (2024). “Dólares or Dollars? Unraveling the Bilingual Prowess of Financial LLMs Between Spanish and English,” in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, R. Baeza-Yates and F. Bonchi (eds.), Barcelona Spain. 25.08.2024 - 29.08.2024, New York, NY, USA: ACM, pp. 6236-6246 (doi: 10.1145/3637528.3671554).

Zheng, X., Li, J., Lu, M., and Wang, F.-Y. (2024). “New Paradigm for Economic and Financial Research With Generative AI: Impact and Perspective,” *IEEE Transactions on Computational Social Systems* (11:3), pp. 3457-3467 (doi: 10.1109/TCSS.2023.3334306).