



American Journal of Financial Technology and Innovation (AJFTI)

ISSN: 2996-0975 (ONLINE)

VOLUME 4 ISSUE 1 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

Cultural-Historical Activity Theory and AI: Innovating and Optimizing Financial Data Retrieval

Yara Mohammed^{1*}, Brady Lund¹

Article Information

Received: October 15, 2024

Accepted: January 11, 2025

Published: March 24, 2026

Keywords

Artificial Intelligence, Cultural-Historical Activity Theory, Financial Modeling, Natural Language Processing, Prompt Engineering

ABSTRACT

Machine learning is a transformative technology with profound applications across various domains, including finance. This study introduces the Artificial Intelligent Financial Model (AIFM), developed to enhance the accuracy and efficiency of retrieving CEO pay and median employee pay. Despite significant advances in chatbot technologies within financial services, accurate evaluation of these systems, particularly in terms of handling complex financial documents, remains challenging. Traditional models often fail to capture the complexities of financial data analysis through effective, prompt engineering. To address these limitations, the AIFM uses advanced machine learning and NLP to evaluate the system's response to simulated queries. This involves sophisticated prompt engineering strategies that ensure precise and reliable evaluation of the chatbot's performance, enabling the system to process inputs accurately without the need for real-time human feedback. This approach allows for controlled testing and refinement of the chatbot's capabilities in a consistent and repeatable environment. Moreover, this research incorporates Cultural-Historical Activity Theory (CHAT) from a technical perspective, deviating from the typical qualitative approaches often seen in other papers. We aim to provide a structured and measurable analysis of the AI's interaction logic and its ability to process and present financial data accurately. Our findings demonstrate the effectiveness of the AIFM in providing detailed and accessible financial insights, which could significantly impact on the broader field of finance by introducing new levels of precision and analysis-focused interaction. This study proposes a novel approach that could expand the current understanding of AI's capability in high-stakes environments.

INTRODUCTION

Understanding the structural and economic dynamics within organizations is essential for their success, with financial transparency as a cornerstone. Achieving transparency requires efficient, user-centered information systems and retrieval mechanisms. These critical priorities often receive less attention in traditional financial systems and artificial intelligence technologies development (He *et al.*, 2023).

This paper introduces the Artificial Intelligent Financial Model (AIFM), building on Mohammed (2024) by replicating and extending the PSAIM Proxy Statement Artificial Intelligent Model. Our enhancement integrates Cultural-Historical Activity Theory (CHAT) using Engeström's (2001) expanded activity structure, offering a new perspective for analyzing financial data retrieval processes, focusing on community dynamics and component roles.

The need for precise and accessible financial data is critical. Stakeholders such as investors, financial analysts, and corporate shareholders view CEO compensation, median employee pay and pay ratios as essential for comprehensive financial analysis. Our study replicates PSAIM's ability to manage complex financial queries with advanced machine learning techniques, assessing effectiveness in retrieving detailed financial documents. We evaluate the AIFM's performance using RAGAS,

a library enhancing Large Language Model (LLM) applications evaluation by incorporating ground truth data, ensuring a detailed and accurate assessment. Applying CHAT quantitatively, we explore dynamics within the financial data retrieval system, focusing on elements like the chatbot as a mediating tool and the community of stakeholders.

Therefore, the AIFM not only advances technical capabilities but also deepens our comprehension of operational contexts through a structured theoretical lens. This balanced focus on technological innovation and analytical depth makes our model beneficial for stakeholders eager to leverage detailed financial data for a thorough analysis of corporate health and equity structure. It is useful to ensure the effectiveness of the tool through the evaluation process.

LITERATURE REVIEW

The growing role of artificial intelligence (AI) in the finance sector enhances investors' access to critical performance data, showing significant advancements in personalized, context-specific responses (Weber *et al.*, 2024). Traditional chatbot evaluation frameworks, like the glass box and black box models by Shawar and Atwell (2007), focus on technical performance but often overlook the complexity needed for personalization in financial queries.

¹ University of North Texas

* Corresponding author's e-mail: yara.mohammed@unt.edu

CHAT has been applied in various fields outside finance, such as mobile tutoring and educational psychology, demonstrating its utility in complex systems (Tan *et al.*, 2023; Leontiev, 2020). In information science, CHAT helps understand interactions between users and systems, providing a solid foundation for its application in AI-mediated financial data retrieval (Allen *et al.*, 2011; Sung, 2017).

However, CHAT's use in finance is relatively new, with limited studies focusing on financial information systems (Foot, 2024). Our research applies CHAT to improve AI-powered chatbot efficiency in retrieving financial data like CEO pay and pay ratios, enhancing data accessibility and reducing manual retrieval time.

Zeng and Peng (2024) developed chatbot models for enhanced personalization across languages and contexts, although not specifically for financial data. Our study focuses on fine-tuning chatbots for financial information retrieval, especially from proxy statements, tailored to manage the complexities of financial data.

Moreover, integrating Cohere's model-tuning with Retrieval-Augmented Generation (RAG) has improved the chatbot's ability to provide accurate, contextually appropriate responses. This approach, supported by recent findings like those of Zhang *et al.* (2022) and Formanek (2024), significantly advances financial information processing and has broader implications across various disciplines.

This study contributes to information science by applying CHAT in financial data retrieval, exploring deeper based on Mohammed (2024). By structuring a framework to improve AI-mediated interactions, our work addresses a critical gap in AI applications for financial information retrieval, emphasizing the need for refined accuracy and timeliness in high-stakes environments.

Research Questions and Assumptions

RQ1: How does using Cultural-Historical Activity Theory (CHAT) as a structuring framework affect the

functionality and accuracy of the AIFM in financial data retrieval and processing?

RQ2: How does the evaluation process with RAGAS impact the assessment of the AIFM's performance in terms of accuracy and efficiency?

Research Assumptions

Assumption 1: Integrating CHAT into the AIFM framework is expected to enhance the model's effectiveness, improving the chatbot's ability to handle complex queries and interact effectively with financial data.

Assumption 2: Using RAGAS for evaluation is presumed to provide a robust assessment of the AIFM's performance, enabling precise measurement of model accuracy and reliability in real-world scenarios.

Current Practices Informing This Study

This study addresses the challenges in accurately retrieving and presenting complex financial data, such as CEO compensation and pay ratios, using advanced AI technologies. Traditional instruction tuning and large language models (LLMs) often struggle with the specificity and accuracy required for financial data retrieval. These models, optimized for general-purpose conversations, are limited in handling intricate financial documents like proxy statements.

To address these limitations, the study employs a specialized approach using tools like Cohere's Command R for better embedding and contextual understanding, integrated with Retrieval-Augmented Generation (RAG) processes (Hambarde & Proenca, 2023). This enables the chatbot model to manage the nuances of financial data, enhancing accuracy and delivering contextually relevant information.

The study also promotes accountability and transparency in financial reporting, helping users (investors) rely on the chatbot for precise financial insights. This is crucial for evaluating corporate governance and ensuring stakeholder trust.

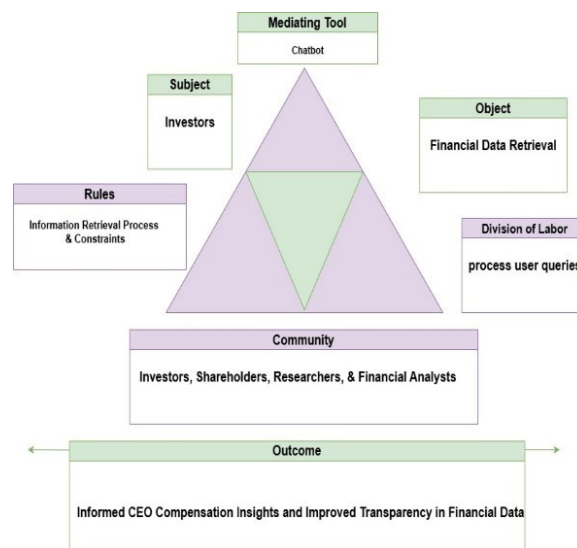


Figure 1: Contribution of Cultural-Historical Activity Theory with Chatbot Mediation in Finance using Engeström's (2001) expanded activity structure.

Additionally, the study uses Cultural-Historical Activity Theory (CHAT) as the theoretical framework to analyze how the chatbot interacts with stakeholders within the financial ecosystem, balancing transparency needs with financial regulations and data privacy constraints.

By leveraging AI advancements, the study aims to bridge the gap between traditional LLMs and the specialized needs of financial data retrieval, refining solutions for the financial industry. This research enhances financial transparency and accountability, contributing to the development of more precise AI models.

In Figure 1, the Cultural-Historical Activity Theory (CHAT) is visualized with an inner and outer triangle, illustrating interactions within the Artificial Intelligent Financial Model (AIFM), a chatbot aiding investors in accessing financial data. The subject, investors, engage with the chatbot to access key metrics such as company performance. The chatbot serves as the mediating tool, central to delivering this data in a clear, accessible manner. CHAT is employed as the analytical framework because it allows for a holistic examination of the financial data retrieval process, connecting actions (like querying the chatbot) and operations (such as the chatbot applying financial models or adhering to data retrieval constraints). Engeström's (2001) expanded activity structure provides a lens to understand these mediations at micro and meso levels, offering insights into how investors access information without making direct investment decisions from the chatbot's output.

The community within this system includes investors, shareholders, researchers, and financial analysts, all benefiting from the chatbot's structured financial insights. They form the broader social and professional environment where the chatbot operates, supporting their research and analytical needs. The system's rules, tied to information retrieval processes and constraints, dictate how the chatbot interacts with data and responds to queries, ensuring accuracy within set boundaries. The division of labor sees the chatbot managing data retrieval and presentation, while investors use the data for further financial understanding.

Aligned with Engeström's framework, the inner triangle highlights interactions between the subject (investor), mediating tool (chatbot), and object (data retrieval), while the outer triangle includes the rules, community, and division of labor. This structured approach emphasizes how AIFM mediates complex financial data interactions,

operating within rule-based constraints to enhance access to relevant insights without direct investment recommendations.

CHAT also aids in identifying contradictions within this system, seen as disruptions or challenges, such as misalignments between the chatbot's features and investor queries. Engeström views these contradictions not as negatives but as opportunities for innovation, prompting learning and system evolution to better serve user needs.

MATERIALS AND METHODS

This project leverages the Proxy Statement Artificial Intelligence Model (PSAIM) developed by Mohammed (2024) to assess its accuracy and efficacy. Our study began by identifying key issues in recent chatbot development, data retrieval optimization, and evaluation, highlighting the need for new methodologies. Data collection was manual, with a check column added to minimize errors. After data preparation, we implemented and optimized the model, conducting thorough tests and adjustments based on performance evaluations.

Subsequently, we reviewed other research employing Cultural-Historical Activity Theory (CHAT) as a base for analyzing and enhancing practice-based learning and systemic changes across various fields. As summarized in Table 1, these studies utilize CHAT to address contradictions within specialized activity systems in diverse domains such as education (Tan *et al.*, 2023; Mathieson *et al.*, 2023) and healthcare (Qureshi, 2021), demonstrating CHAT's adaptability and effectiveness in fostering improvements.

Distinctively, our research applies CHAT to analyze a financial data retrieval chatbot's accuracy, extending its application from educational and healthcare sectors to financial technology, aiming to improve system functionality and efficacy through practical applications and contradiction resolution.

Our research adopts a quantitative approach using Cultural-Historical Activity Theory (CHAT) to evaluate chatbot performance in financial environments, contrasting with the common qualitative use of CHAT in educational or healthcare settings. This innovative application of CHAT enhances transparency in financial reporting by integrating a traditional theoretical framework within a modern, digital context.

While most studies employ CHAT qualitatively to explore educational and healthcare systems, our work uniquely

Table 1: Comparative Analysis of Blockchain Adoption in International Trade and Taxation Across Selected Countries

Study	Model/Technology used	Model Application	Focus Area	Application Type
Tan <i>et al.</i> , 2021	MENTOR (Mobile Peer Tutoring App)	Used for analyzing contradictions and affordances in a mobile tutoring context	Design and evaluation of a mobile application for peer tutoring among university students. Focus on the affordances, student perceptions, and contradictions identified through CHAT.	Qualitative

Mathieson <i>et al.</i> , 2023	CHAT-based research approach	Used to examine and improve the induction-to-teaching experiences of new academics, focusing on resolving contradictions within the induction activity systems	Investigating new academics' experiences with induction to teaching using CHAT to understand and enhance their integration and teaching practices.	Qualitative
Qureshi, 2021	Practice-based medical education settings	Used to analyze contradictions and expansions in medical education through workplace learning and interactions	Studying the application of CHAT in medical education to explore and enhance practice-based learning and change within the healthcare environment.	Qualitative
Our study	AIFM (AI Financial Chatbot)	Applied to evaluate the accuracy of financial data retrieval from proxy statements, guided by CHAT principles for effective data processing and interaction	Creating and evaluating an AI chatbot to enhance transparency	Quantitative

applies it quantitatively in the financial technology sector, focusing on evaluating a financial data chatbot. This approach broadens CHAT's application, highlighting its potential in technological and operational advancements within the financial sector.

For example, Tan *et al.* (2023) use CHAT to assess a mobile peer tutoring application, identifying contradictions and affordances to enhance educational impact and user engagement. Mathieson *et al.* (2023) apply CHAT to improve the induction experiences of new academics in a UK university, addressing contradictions in academic induction systems to refine teaching policies and practices. Similarly, Qureshi (2021) uses CHAT in medical training to resolve contradictions, enhancing educational outcomes by integrating learning requirements with operational demands.

Our study extends CHAT's application to the financial sector by quantitatively analyzing a financial data retrieval chatbot. This method focuses on improving transparency

and efficiency, with the chatbot serving as a mediating tool, providing stakeholders like investors, researchers, and analysts with insightful corporate compensation practices, thereby enhancing operational efficiency in the financial sector.

Problem Methodology

For our AIFM model, data was collected from 2021 company filings via WRDS, specifically focusing on DEF 14A or proxy statements. Data extraction was initially performed using Python scripts linked to each company's disclosure on the SEC website (sec.gov), but this method proved partially unsuccessful. To improve accuracy, we applied a verification process as per formula (1) and manually extracted necessary data directly from sec.gov. Although Mohammed (2024) reported a dataset of 5376 rows, we managed to gather 5386 rows. Modifications included the removal of about 1370 entries from the final rows to test our chatbot, as depicted in Figure 2.

	A	E	H
1	filedate	name	filelink
5308	2021-12-16 00:00:00	MSC INDUSTRIAL DIRECT CO IN	https://www.sec.gov/Archives/edgar/data/1003078/000100307821000001/msc-industrial-direct-co-in-20211216.htm
5311	2021-12-17 00:00:00	ATMOS ENERGY CORP	https://www.sec.gov/Archives/edgar/data/731802/0001193731802210000001/atmos-energy-corp-20211217.htm
5313	2021-12-17 00:00:00	Simulations Plus, Inc.	https://www.sec.gov/Archives/edgar/data/1023459/000168345921000001/simulations-plus-inc-20211217.htm
5314	2021-12-17 00:00:00	PRICESMART INC	https://www.sec.gov/Archives/edgar/data/1041803/0001193180321000001/pricesmart-inc-20211217.htm
5317	2021-12-17 00:00:00	MERITOR, INC.	https://www.sec.gov/Archives/edgar/data/1113256/00012011325621000001/meritor-inc-20211217.htm
5319	2021-12-17 00:00:00	Dolby Laboratories, Inc.	https://www.sec.gov/Archives/edgar/data/1308547/000119130854721000001/dolby-laboratories-inc-20211217.htm
5323	2021-12-20 00:00:00	PTC INC.	https://www.sec.gov/Archives/edgar/data/857005/000120685700521000001/ptc-inc-20211220.htm
5324	2021-12-20 00:00:00	SURMODICS INC	https://www.sec.gov/Archives/edgar/data/924717/000156492471721000001/surmodics-inc-20211220.htm
5337	2021-12-22 00:00:00	TYSON FOODS, INC.	https://www.sec.gov/Archives/edgar/data/100493/000010049321000001/tyson-foods-inc-20211222.htm
5339	2021-12-22 00:00:00	BLACKROCK MUNIYIELD CALIFORNIA	https://www.sec.gov/Archives/edgar/data/882152/000119388215221000001/blackrock-muniyield-california-20211222.htm
5341	2021-12-22 00:00:00	BLACKROCK MUNIYIELD CALIFORNIA	https://www.sec.gov/Archives/edgar/data/882152/000119388215221000001/blackrock-muniyield-california-20211222.htm
5342	2021-12-22 00:00:00	ROCKWELL AUTOMATION, INC	https://www.sec.gov/Archives/edgar/data/1024478/0001308447821000001/rockwell-automation-inc-20211222.htm
5345	2021-12-22 00:00:00	BLACKROCK MUNIHOLDINGS CORP	https://www.sec.gov/Archives/edgar/data/882152/000119388215221000001/blackrock-muniholdings-corp-20211222.htm
5349	2021-12-22 00:00:00	Cardinal Ethanol LLC	https://www.sec.gov/Archives/edgar/data/1352081/000135135208121000001/cardinal-ethanol-llc-20211222.htm
5351	2021-12-23 00:00:00	J&J SNACK FOODS CORP	https://www.sec.gov/Archives/edgar/data/785956/000143778595621000001/j-j-snack-foods-corp-20211223.htm
5360	2021-12-23 00:00:00	Aramark	https://www.sec.gov/Archives/edgar/data/1584509/0001193158450921000001/aramark-20211223.htm
5367	2021-12-27 00:00:00	CENTRAL GARDEN & PET CO	https://www.sec.gov/Archives/edgar/data/887733/000119388773321000001/central-garden-pet-co-20211227.htm
5368	2021-12-28 00:00:00	DUCK CREEK TECHNOLOGIES, INC	https://www.sec.gov/Archives/edgar/data/1160951/000114116095121000001/duck-creek-technologies-inc-20211228.htm
5370	2021-12-28 00:00:00	Varex Imaging Corp	https://www.sec.gov/Archives/edgar/data/1681622/000110168162221000001/varex-imaging-corp-20211228.htm
5372	2021-12-29 00:00:00	MARINEMAX INC	https://www.sec.gov/Archives/edgar/data/1057060/0001564706021000001/marinemax-inc-20211229.htm
5373	2021-12-29 00:00:00	BELLRING BRANDS, INC.	https://www.sec.gov/Archives/edgar/data/1772016/000119177201621000001/bellring-brands-inc-20211229.htm
5376	2021-12-30 00:00:00	Hillenbrand, Inc.	https://www.sec.gov/Archives/edgar/data/1417398/000114141739821000001/hillenbrand-inc-20211230.htm

Figure 2: Pay Ratio Data Set

The pay ratio for the collected data is as follows:

$$\text{CEO Pay Ratio} = \frac{\text{CC}}{\text{MEP}} \quad (1)$$

where CC denotes the CEO Pay, and MEP denotes the Median Employee Pay for fiscal year 2021.

Data Architecture

To adapt the data for our semantic search model, we developed a system that converts raw text into a quantifiable vector space. The transformation process includes several essential steps to ensure data accuracy and relevance. First, a preprocessing function cleans and standardizes each document's text by removing irrelevant characters, converting text to lowercase, and tokenizing into words. Then, using Cohere's neural network-based model, the cleaned text is transformed into a high-dimensional vector space, with each text string represented as a vector. Key criteria for each instance include:

$$(\vec{q}, \vec{v}_i) = \frac{(\vec{q} \cdot \vec{v}_i)}{||\vec{q}|| ||\vec{v}_i||} \quad (2)$$

Each document vector \vec{v}_i representing semantic content is normalized to unit length to ensure accurate similarity calculations across the dataset. These standardized vectors are stored in MongoDB, with a vector index created via MongoDB Atlas Vector Search to facilitate efficient retrieval by vector similarity.

For search queries, a similar vectorization process converts each query into a vector \vec{q} , using the same embedding model. We compute the cosine similarity between the query vector \vec{q} and each document vector \vec{v}_i in the database. The norms $||\vec{q}||$ and $||\vec{v}_i||$ of the query and document vectors facilitate this calculation. The resulting similarities rank the documents in relevance to the query, effectively aligning the process with user queries and the specific needs of our chatbot.

Conceptual Model

The Concept Map of AIFM in figure 3 provides a structured visualization of the Artificial Intelligent Financial Model (AIFM), showcasing its components

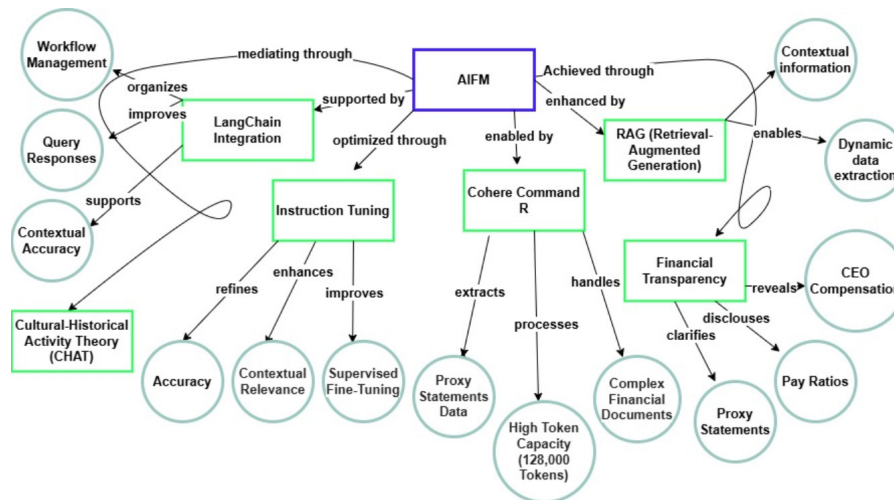


Figure 3: AIFM Components

and their relationships aimed at improving financial data retrieval and accuracy. At the center is the “Financial Data Retrieval via AI (AIFM),” the model’s core function. From here, arrows branch out to a first layer of critical concepts: Instruction Tuning, Cohere Command R, RAG (Retrieval-Augmented Generation), Financial Transparency, Cultural-Historical Activity Theory (CHAT), and LangChain Integration. Each is linked to the central node, illustrating their roles in enhancing the AIFM’s functionality. For instance, Instruction Tuning optimizes precision, while Cohere Command R enables handling of complex financial documents.

Sub-concepts branch further from these main ideas, detailing specific functions. Instruction Tuning connects to Accuracy, Contextual Relevance, and Supervised Fine-Tuning, emphasizing refinement in user instruction adherence. Cohere Command R extends to capabilities like handling complex documents and extracting CEO compensation. RAG is associated with Dynamic Data Extraction and Contextual Information Generation, crucial for

producing relevant financial data. Financial Transparency links to CEO Compensation, Pay Ratios, and Proxy Statements, highlighting the model’s role in enhancing financial reporting transparency.

CHAT explores the chatbot’s mediation between investors and financial data, enhancing information retrieval effectiveness. Lastly, LangChain Integration focuses on Workflow Management, Query Responses, and Contextual Accuracy, key in managing query flow and improving response relevance.

AIFM Optimization

We utilize the CohereEmbeddings (embed-english-v3.0) model, finely tuned for generating high-dimensional vector representations that capture deep semantic meanings. This capability improves the accuracy and relevance of searches, crucial for processing complex financial documents. As noted by Mohammed (2024), we chose Cohere’s Command R over LLAMA due to its ability to handle up to 128,000 tokens, paired with

Retrieval-Augmented Generation (RAG) for dynamic information extraction and generation, specifically tailored to the needs of financial data retrieval. The embeddings are integrated into MongoDB Atlas Vector Search, creating a vector index that enhances the speed and precision of our chatbot's semantic searches. The chatbot, powered by the ChatCohere model with command-r-plus and a temperature setting of 0.5, strikes a balance between creativity and factual retrieval, aligning responses closely with user queries. All interactions are logged in our MongoDB's chat_logs_collection, essential for monitoring performance and guiding enhancements. Moreover, Cohere's Command R elevates our system's transparency and ethical standards, ensuring responses are understandable and contextually appropriate. This setup includes a robust feedback loop through our MongoDB

database, allowing continuous performance refinement based on real user interactions. This iterative process helps adapt the chatbot to evolving user needs, maintaining reliability and trustworthiness in our operations.

AIFM Implementation

We ingested documents, converting them into vector format, and crafted a prompt template while refining hyperparameters to ensure relevance in responses. This refining process was repeated to boost the accuracy of our model. To construct a fully functional chatbot, we defined a series of steps for accepting user prompts, processing them, and formatting the responses as strings, crucial for our model's implementation. Although we stored our chat data in MongoDB, details of this storage process are not covered in this paper.

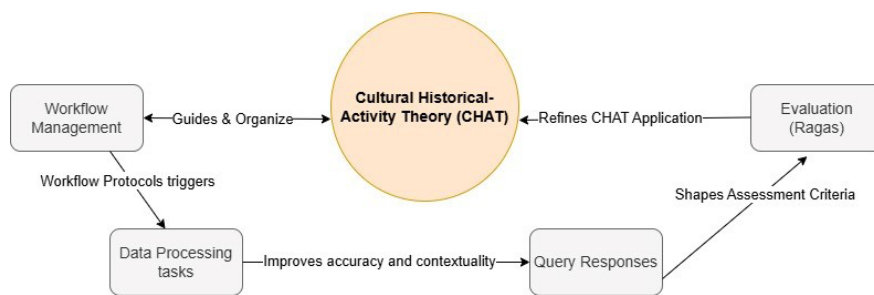


Figure 4: AIFM System Architecture

Figure 4 displays the Concept Map of the Artificial Intelligent Financial Model (AIFM), centered on Cultural-Historical Activity Theory (CHAT). This map shows CHAT as the core element guiding the AIFM's architecture and functionality. From CHAT, an arrow leads to Workflow Management, indicating how CHAT principles organize data processing activities, aligning them with the model's strategic objectives. Workflow Management feeds into Data Processing, enhancing information accuracy and contextuality, crucial for the precision of financial information retrieval. The improved data then informs Query Responses, where the system generates precise and relevant replies, illustrating the direct impact of data accuracy on system interactions. An arrow from CHAT also points to the Evaluation component, marked with Ragas, highlighting evaluation's

essential role in defining assessment criteria and promoting the continuous refinement of CHAT principles based on real-world data. This setup ensures the AIFM model dynamically integrates feedback to optimize its effectiveness.

RESULTS AND DISCUSSION

We utilized four key metrics from the RAGAS framework—context recall, faithfulness, context utilization, and context precision—to evaluate our chatbot, as inspired by Mohammed (2024). We selected 10 model responses for in-depth analysis, aiming to pinpoint strengths and weaknesses in the RAG pipeline and enhance the chatbot's accuracy and efficacy. This targeted assessment ensures a comprehensive review of both overall system performance and individual

Table 3: Context Recall Evaluation Result

Context Recall	User Query	Answer	Contexts	Evaluation
0	How much is the CEO pay for Apple Inc.?	The CEO of Apple Inc. received a total compensation of \$14,769,259 for the 2020 fiscal year.	Company Name: Apple Inc., CEO Pay: \$14,769,25...	0.500000
1	What is the median employee pay for Evoqua Water Technologies Corp.?	The median employee pay for Evoqua Water Technologies Corp. as of the 2020 fiscal year was \$74,540.	Company Name: Evoqua Water Technologies Corp., Median Employee Pay: \$74,540...	1.000000

2	Is it risky to invest?	I cannot provide a general answer to the question of investment risk, as it depends on various factors, including the specific investment opportunity, your financial situation, and your risk tolerance...	Investment Risk Factors: Market conditions, economic climate...	1.000000
3	When can you say that there is a good pay ratio?	A "good" pay ratio is subjective and depends on various factors and perspectives...	Pay Ratio Details: Some people argue that a pay ratio between 20:1 and 30:1 is reasonable...	1.000000
4	Can you assist me with general information about pay ratio	The pay ratio is the ratio of a company's CEO's annual total compensation to the median annual total compensation of all its employees...	Pay Ratio: ON Semiconductor Corp had a pay ratio of 684:1 in 2020...	0.250000
5	What factors do I consider to tell me if it is a good time to invest from given information and when it is a bad time? Do I consider a financial advisor	I cannot provide a definitive answer on when is the best time to invest as it depends on a variety of factors and risks, including your financial situation, goals, and risk tolerance, as well as market conditions and economic trends...	Investment Timing Factors: Risk tolerance, market conditions, expert opinions...	0.666667
6	How accurate is your information?	The information provided in the sources appears to be accurate, as it includes specific details such as company names, CIK numbers, file dates, and pay ratios. The pay ratio calculations are also supported by the CEO and median employee compensation...	Data Accuracy: The information appears accurate but should be verified through reliable sources...	1.000000
7	Given the pay ratio data, how is it related to governance factor?	The pay ratio is a governance factor that reflects a company's commitment to fair and responsible executive compensation practices. It is calculated by dividing the CEO's pay by the median employee's pay, and it is often disclosed in regulatory filings, such as the DEF 14A, to ensure transparency and accountability...	Governance Factor: Pay ratio reflects fair executive compensation practices...	1.000000
8	How is pay ratio related to firm's performance?	I can see that the 'Pay Ratio Note' in Source 1 mentions that the pay ratio is calculated by dividing the CEO's pay by the median employee's pay...	Firm Performance: No direct link to firm performance mentioned in sources...	0.000000
9	Tell me about Hillenbrand CEO pay?	I'm sorry, but the documents provided do not contain information about the CEO pay for Hillenbrand, Inc. They mention that the CEO pay is "not stated" and do not provide further details.	CEO Pay: Not stated for Hillenbrand, Inc...	0.333333

Table 4: Faithfulness Evaluation Result

Faithfulness	User Query	Answer	Contexts	Evaluation
0	How much is the CEO pay for Apple Inc.?	The CEO of Apple Inc. received a total compensation of \$14,769,259 for the 2020 fiscal year.	Company Name: Apple Inc., CEO Pay: \$14,769,25...	1.000000
1	What is the median employee pay for Evoqua Water Technologies Corp.?	The median employee pay for Evoqua Water Technologies Corp. as of the 2020 fiscal year was \$74,540.	Company Name: Evoqua Water Technologies Corp., Median Employee Pay: \$74,540...	1.000000
2	Is it risky to invest?	I cannot provide a general answer to the question of investment risk, as it depends on various factors, including the specific investment opportunity, your financial situation, and your risk tolerance...	Investment Risk Factors: Market conditions, economic climate...	0.000000
3	When can you say that there is a good pay ratio?	A "good" pay ratio is subjective and depends on various factors and perspectives...	Pay Ratio Details: Some people argue that a pay ratio between 20:1 and 30:1 is reasonable...	1.000000
4	What factors do I consider to tell me if it is a good time to invest from given information and when it is a bad time? Do I consider a financial advisor?	The pay ratio is the ratio of a company's CEO's annual total compensation to the median annual total compensation of all its employees...	Pay Ratio: ON Semiconductor Corp had a pay ratio of 684:1 in 2020...	0.333333
5	What factors do I consider to tell me if it is...	I cannot provide a definitive answer on when is the best time to invest as it depends on a variety of factors and risks, including your financial situation, goals, and risk tolerance, as well as market conditions and economic trends...	Investment Timing Factors: Risk tolerance, market conditions, expert opinions...	1.000000
6	How accurate is your information?	The information provided in the sources appears to be accurate, as it includes specific details such as company names, CIK numbers, file dates, and pay ratios. The pay ratio calculations are also supported by the CEO and median...	Data Accuracy: The information appears accurate but should be verified through reliable sources...	0.500000
7	Given the pay ratio data, how is it related to governance factor?	The pay ratio is a governance factor that reflects a company's commitment to fair and responsible executive compensation practices. It is calculated by dividing the CEO's pay by the median employee's pay, and it is often disclosed in regulatory filings, such as the DEF 14A, to ensure transparency and accountability...	Governance Factor: Pay ratio reflects fair executive compensation practices...	1.000000

8	How is pay ratio related to firm's performance?	I can see that the 'Pay Ratio Note' in Source 1 mentions that the pay ratio is calculated by dividing the CEO's pay by the median employee's pay...	Firm Performance: No direct link to firm performance mentioned in sources...	1.000000
9	Tell me about Hillenbrand CEO pay?	I'm sorry, but the documents provided do not contain information about the CEO pay for Hillenbrand, Inc. They mention that the CEO pay is "not stated" and do not provide further details.	CEO Pay: Not stated for Hillenbrand, Inc...	1.000000

Table 5: Context Utilization Evaluation Result

Context Utilization	User Query	Answer	Contexts	Evaluation
0	How much is the CEO pay for Apple Inc.?	The CEO of Apple Inc. received a total compensation of \$14,769,259 for the 2020 fiscal year.	Company Name: Apple Inc., CEO Pay: \$14,769,25...	1.000000
1	What is the median employee pay for Evoqua Water Technologies Corp.?	The median employee pay for Evoqua Water Technologies Corp. as of the 2020 fiscal year was \$74,540.	Company Name: Evoqua Water Technologies Corp., Median Employee Pay: \$74,540...	1.000000
2	Is it risky to invest?	I cannot provide a general answer to the question of investment risk, as it depends on various factors, including the specific investment opportunity, your financial situation, and your risk tolerance...	Investment Risk Factors: Market conditions, economic climate...	1.000000
3	When can you say that there is a good pay ratio?	A "good" pay ratio is subjective and depends on various factors and perspectives...	Pay Ratio Details: Some people argue that a pay ratio between 20:1 and 30:1 is reasonable...	1.000000
4	Can you assist me with general information about pay ratio?	The pay ratio is the ratio of a company's CEO's annual total compensation to the median annual total compensation of all its employees...	Pay Ratio: ON Semiconductor Corp had a pay ratio of 684:1 in 2020...	0.000000
5	What factors do I consider to tell me if it is a good time to invest from given information and when it is a bad time? Do I consider a financial advisor?	I cannot provide a definitive answer on when is the best time to invest as it depends on a variety of factors and risks, including your financial situation, goals, and risk tolerance, as well as market conditions and economic trends...	Investment Timing Factors: Risk tolerance, market conditions, expert opinions...	1.000000
6	How accurate is your information?	The information provided in the sources appears to be accurate, as it includes specific details such as company names, CIK numbers, file dates, and pay ratios. The pay ratio calculations are also supported by the CEO and median...	Data Accuracy: The information appears accurate but should be verified through reliable sources...	1.000000

7	Given the pay ratio data, how is it related to governance factor?	The pay ratio is a governance factor that reflects a company's commitment to fair and responsible executive compensation practices. It is calculated by dividing the CEO's pay by the median employee's pay, and it is often disclosed in regulatory filings, such as the DEF 14A, to ensure transparency and accountability...	Governance Factor: Pay ratio reflects fair executive compensation practices...	1.000000
8	How is pay ratio related to firm's performance?	I can see that the 'Pay Ratio Note' in Source 1 mentions that the pay ratio is calculated by dividing the CEO's pay by the median employee's pay...	Firm Performance: No direct link to firm performance mentioned in sources...	1.000000
9	Tell me about Hillenbrand CEO pay?	I'm sorry, but the documents provided do not contain information about the CEO pay for Hillenbrand, Inc. They mention that the CEO pay is "not stated" and do not provide further details.	CEO Pay: Not stated for Hillenbrand, Inc...	1.000000

component contributions to user experience. Our findings indicate performance variations based on query type. For example, the chatbot partially recalled context for an Apple Inc. CEO pay query, scoring 0.50 (Table 3), but it accurately extracted the exact pay amount, showcasing its precision. It scored perfectly on faithfulness, context utilization, and precision in other assessments (Tables 4, 5, and 6). However, it struggled with non-specific queries like “Is it risky to invest?”, where it failed to provide definitive advice, scoring 0.00

in faithfulness (Table 4), but managed to discuss relevant risk factors.

Queries on pay ratios demonstrated the chatbot's capability and limitations. For specific questions like “When can you say that there is a good pay ratio?”, it achieved perfect scores across all metrics (Tables 3, 4, 5, and 6). Yet, for broader questions, it showed minimal recall and poor context utilization, resulting in lower scores (Tables 3 and 5).

Table 6: Context Precision Evaluation Result

Context Precision	User Query	Answer	Contexts	Evaluation
0	How much is the CEO pay for Apple Inc.?	The CEO of Apple Inc. received a total compensation of \$14,769,259 for the 2020 fiscal year.	Company Name: Apple Inc., CEO Pay: \$14,769,25...	1.000000
1	What is the median employee pay for Evoqua Water Technologies Corp.?	The median employee pay for Evoqua Water Technologies Corp. as of the 2020 fiscal year was \$74,540.	Company Name: Evoqua Water Technologies Corp., Median Employee Pay: \$74,540...	1.000000
2	Is it risky to invest?	I cannot provide a general answer to the question of investment risk, as it depends on various factors, including the specific investment opportunity, your financial situation, and your risk tolerance...	Investment Risk Factors: Market conditions, e...	1.000000
3	When can you say that there is a good pay ratio?	A "good" pay ratio is subjective and depends on various factors and perspectives...	Pay Ratio Details: Some people argue that a pay ratio between 20:1 and 30:1 is reasonable...	1.000000

4	Can you assist me with general information about pay ratio?	The pay ratio is the ratio of a company's CEO's annual total compensation to the median annual total compensation of all its employees...	Pay Ratio: ON Semiconductor Corp had a pay ratio of 684:1 in 2020...	0.000000
5	What factors do I consider to tell me if it is a good time to invest from given information and when it is a bad time? Do I consider a financial advisor?	I cannot provide a definitive answer on when is the best time to invest as it depends on a variety of factors and risks, including your financial situation, goals, and risk tolerance, as well as market conditions and economic trends...	Investment Timing Factors: Risk tolerance, market conditions, expert opinions...	1.000000
6	How accurate is your information?	The information provided in the sources appears to be accurate, as it includes specific details such as company names, CIK numbers, file dates, and pay ratios. The pay ratio calculations are also supported by the CEO and median...	Data Accuracy: The information appears accurate but should be verified through reliable sources...	1.000000
7	Given the pay ratio data, how is it related to governance factor?	The pay ratio is a governance factor that reflects a company's commitment to fair and responsible executive compensation practices. It is calculated by dividing the CEO's pay by the median employee's pay, and it is often disclosed in regulatory filings, such as the DEF 14A, to ensure transparency and accountability...	Governance Factor: Pay ratio reflects fair executive compensation practices...	1.000000
8	How is pay ratio related to firm's performance?	I can see that the 'Pay Ratio Note' in Source 1 mentions that the pay ratio is calculated by dividing the CEO's pay by the median employee's pay...	Firm Performance: No direct link to firm performance mentioned in sources...	1.000000
9	Tell me about Hillenbrand CEO pay?	I'm sorry, but the documents provided do not contain information about the CEO pay for Hillenbrand, Inc. They mention that the CEO pay is "not stated" and do not provide further details.	CEO Pay: Not stated for Hillenbrand, Inc...	1.000000

Despite mixed results in recall for complex topics like investment timing and governance factors related to pay ratios, the chatbot consistently performed well in context utilization and precision (Tables 5 and 6). This highlights its effectiveness in using and accurately processing retrieved information, even when recall is variable.

Figure 5's heatmap shows that the AIFM excels in queries related to pay ratios and governance, consistently achieving high scores across all metrics. This is particularly impressive given the absence of governance factors in our data, suggesting AIFM's strong capability to contextualize pay structures and their governance impacts. Figure 6 employs a radar chart to display the model's performance across Context Recall, Faithfulness, Context Utilization, and Context Precision. Each metric is scored from 0 to 1, with lines extending toward the chart's perimeter

indicating high performance, and lines near the center showing areas for improvement.

The radar chart visualizes each query's performance through line shapes and extents: lines closer to the edge denote strong capabilities, while those nearer to the center suggest potential for enhancement. This visualization confirms the AIFM's strengths in Context Recall and Utilization, showcasing its effectiveness in applying relevant contexts to responses. However, variations in Faithfulness and Precision indicate difficulties in consistently delivering precise and reliable information, especially for complex queries. These observations highlight the model's adeptness at managing context-rich inquiries and point out the need for improved accuracy and specificity in diverse query scenarios.

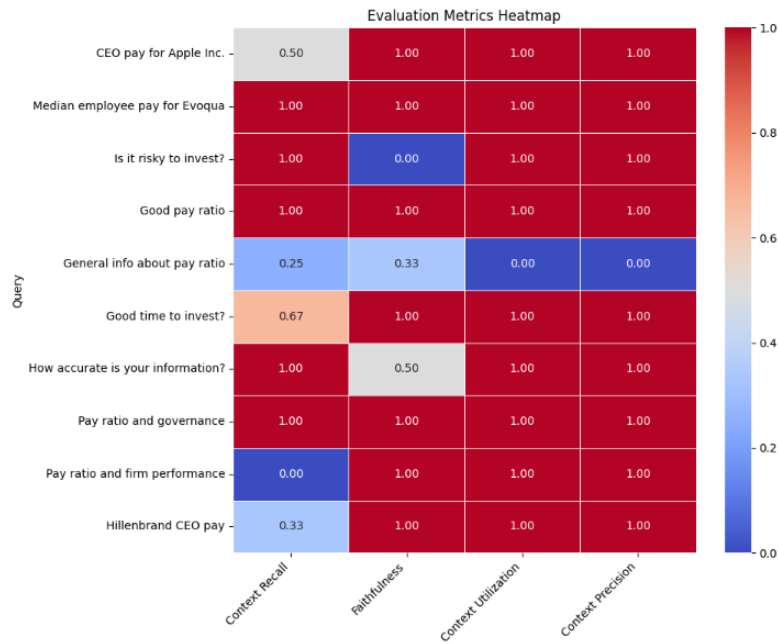


Figure 5: Reliable and Generalizable Assessment System Evaluation

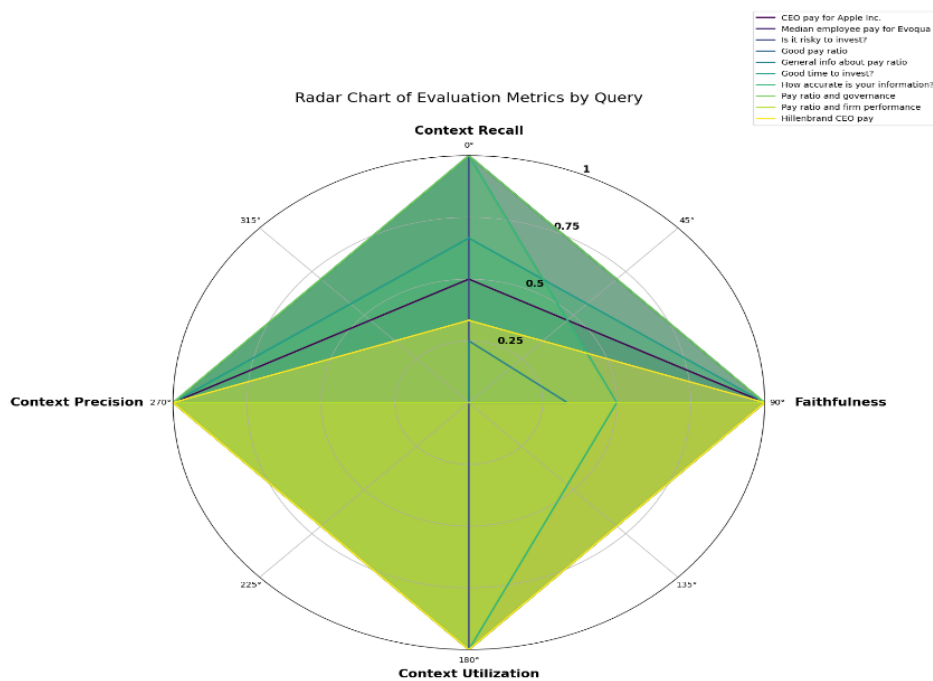


Figure 6: Evaluation Metrics Radar: Chatbot Response Assessment

Additional Features

After evaluating our model with the RAGAS system, we further analyzed user queries using the DBSCAN clustering algorithm with an epsilon of 0.5 to detect outliers, designating those labeled as -1 as outliers (Figure 7). These outliers, indicative of anomalies or errors, are segregated into a separate dataframe, `df_outliers`, for further analysis or exclusion, ensuring the integrity of our financial metrics. The remaining reliable data is stored in `df_clean`. Notably, out of 10 queries, only the query about general information on pay ratios, listed as entry number

4, was identified as an outlier (Table 2), confirmed by our line graph (Figure 8).

Our strategy refines chatbot responses to ensure they focus on relevant questions based on the available data. This outlier query scored zero in Context Utilization and Precision, reflecting a misalignment with our retrieval goals. This prompts precise engineering to target data-relevant questions effectively. The chatbot's lower scores, 0.25 in Context Recall and 0.333 in Faithfulness for this query, underscore potential improvements, such as better training on pay ratios, refined query interpretations, or

Outliers detected by DBSCAN with eps=0.5:

	Context Recall	Faithfulness	Context Utilization	Context Precision
4	0.25	0.333	0	0

Clean data with eps=0.5:

	Context Recall	Faithfulness	Context Utilization	Context Precision
0	0.500	1.0	1	1
1	1.000	1.0	1	1
2	1.000	0.0	1	1
3	1.000	1.0	1	1
5	0.666	1.0	1	1
6	1.000	0.5	1	1
7	1.000	1.0	1	1
8	0.000	1.0	1	1
9	0.333	1.0	1	1

Figure 7: DBSCAN Clustering Algorithm

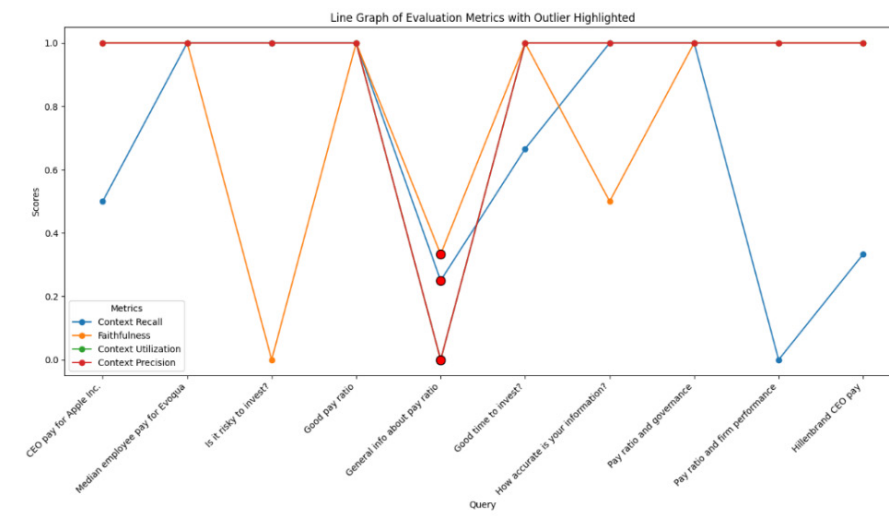


Figure 8: Line Graph highlighting the outlier

Table 7: DBSCAN Outlier Detection

Eps=0.5	Context recall	Faithfulness	Context utilization	Context precision
0	0.500	1.0	1	1
1	1.000	1.0	1	1
2	1.000	0.0	1	1
3	1.000	1.0	1	1
4	0.25	0.333	0	0
5	0.666	1.0	1	1
6	1.000	0.5	1	1
7	1.000	1.0	1	1
8	0.000	1.0	1	1
9	0.333	1.0	1	1

corrections in data handling.

CONCLUSIONS

This study introduced the Artificial Intelligent Financial Model (AIFM), a chatbot-based system designed to retrieve CEO pay, median employee pay, and pay ratio information from proxy statements with high accuracy

and efficiency. The model integrates retrieval-augmented generation with vector search to support precise extraction from complex financial documents, and it evaluates responses using four metrics: context recall, faithfulness, context utilization, and context precision. Across test queries, AIFM performed strongly on structured, data-driven requests, consistently producing accurate, context-

grounded answers when the required information was present in the dataset. Performance was weaker for broad or loosely defined questions, where relevant document support was limited or queries were not aligned with available data. To strengthen evaluation reliability, the study also applied outlier detection to identify anomalous queries that reduced overall performance and to highlight areas needing improved query handling. Overall, the findings show that AIFM is effective for targeted financial data retrieval and supports transparent, repeatable assessment of chatbot performance in high-stakes information settings.

Despite the study's success, it is currently limited to data from the year 2021. For future work, we plan to (a) integrate governance factors and (b) expand our dataset to include a wider range of queries. This will not only improve our process for detecting outliers but also refine the chatbot's ability to handle governance-related questions. Although the responses to these queries were generic, they remained accurate and relevant. Moving forward, we are interested in incorporating additional variables to see how they influence the chatbot's response efficacy.

REFERENCES

- Allen, D., Karanasios, S., & Slavova, M. (2011). Working with activity theory: Context, technology, and information behavior. *Journal of the American Society for Information Science and Technology*, 62(4), 776-788.
- Amirizani, M., Martin, E., Sivachenko, M., Mashhadi, A., & Shah, C. (2024, October). Can LLMs reason like humans? Assessing theory of mind reasoning in LLMs for open-ended questions. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management* (pp. 34-44).
- Foot, K. A. (2014). Cultural-historical activity theory: Exploring a theory to inform practice and research. *Journal of Human Behavior in the Social Environment*, 24(3), 329-347.
- Hambarde, K. A., & Proenca, H. (2023). Information retrieval: Recent advances and beyond. *IEEE Access*, 11, 76581-76604.
- He, X., Hong, Y., Zheng, X., & Zhang, Y. (2023). What are the users' needs? Design of a user-centered explainable artificial intelligence diagnostic system. *International Journal of Human-Computer Interaction*, 39(7), 1519-1542.
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237-257.
- Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. *Government Information Quarterly*, 37(3), article 101493.
- Jeppson, C. T., Smith, W. W., & Stone, R. S. (2009). CEO compensation and firm performance: Is there any relationship? *Journal of Business & Economics Research (JBER)*, 7(11).
- Khan, S., & Rabbani, M. R. (2020, November). Chatbot as Islamic finance expert (CaIFE): When finance meets artificial intelligence. In *Proceedings of the 2020 4th International Symposium on Computer Science and Intelligent Control* (pp. 1-5).
- Lee, M. K., & Park, H. (2019). Exploring factors influencing usage intention of chatbot-chatbot in financial service. *Journal of Korean Society for Quality Management*, 47(4), 755-765.
- Leontiev, D. A. (2020). Activity theory and the functional paradigm in educational psychology. *Educational Psychologist*, 55(1), 45-57. <https://doi.org/10.1080/00461520.2019.1700437>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- Lilling, M. S. (2006). The link between CEO compensation and firm performance: Does simultaneity matter? *Atlantic Economic Journal*, 34, 101-114.
- Liu, Y., Bhandari, S., & Pardos, Z. A. (2024). Leveraging LLM-Respondents for item evaluation: A psychometric analysis. *arXiv preprint arXiv:2407.10899*.
- Mathieson, K., Wallace, J., & Turner, A. (2023). Induction experiences of new academics: A CHAT-based approach. *Journal of Higher Education*, 94(3), 472-490. <https://doi.org/10.1353/jhe.2023.0003>
- Michaud, D. W., & Gai, Y. (2009). *CEO compensation and firm performance*. Available at SSRN 1531673.
- Mohammed, Y. (2024). Optimizing financial data accuracy and searchability using AI user interface and large language models: A case study of CEO compensation. In *Proceedings of the 19th International Conference on Knowledge Management*.
- Okuda, T., & Shoda, S. (2018). AI-based chatbot service for the financial industry. *Fujitsu Scientific and Technical Journal*, 54(2), 4-8.
- Qureshi, S. P. (2021). Cultural historical activity theory for studying practice-based learning and change in medical education. *Advances in Medical Education and Practice*, 12, 923-935. <https://doi.org/10.2147/AMEPS123456>
- Riley, F., Allen, D. K., & Wilson, T. D. (2022). When politicians and experts collide: Organization and the creation of information spheres. *Journal of the Association for Information Science and Technology*, 73(8), 1127-1139.
- Schiller, S. Z. (2016). CHAT for chat: Mediated learning in online chat virtual reference service. *Computers in Human Behavior*, 65, 651-665.
- Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they really useful? *Journal for Language Technology and Computational Linguistics*, 22(1), 29-49.
- Sung, H. (2017). Incorporating technology in children's storytime: Cultural-historical activity theory as a

- means of reconciling contradictions. *Library and Information Science Research*, 39(1), 46-52.
- Tan, X., Liu, X., & Zhou, M. (2023). Enhancing peer learning in mobile tutoring applications through CHAT. *Journal of Interactive Learning Research*, 34(2), 205-222. <https://doi.org/10.17160/jilr.2023.34.2.205>
- Trace, C. B., & Karadkar, U. P. (2016). Information management in the humanities: Scholarly processes, tools, and the construction of personal collections. *Journal of the Association for Information Science and Technology*, 68(2), 491-507.
- Wang, J., Wang, K., Wang, X., Naidu, P., Bergen, L., & Paturi, R. (2023). Doris-mae: Scientific document retrieval using multi-level aspect-based queries. *arXiv preprint arXiv:2310.04678*.
- Weber, P., Carl, K. V., & Hinz, O. (2024). Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature. *Management Review Quarterly*, 74(2), 867-907.
- Zeng, J., & Peng, Y. (2024). Enhancing multilingual chatbot interactions: A comparative study. *AI in Language Processing*, 12(4), 678-692. <https://doi.org/10.1098/aalp.2024.004>
- Zhang, Y., Liu, B., & Chen, X. (2022). Contextual relevance in user interactions: Towards more engaging AI systems. *Journal of Applied AI*, 37(11), 2341-2355. <https://doi.org/10.1017/jai.2022.112>