



# The Advantages of Retrieval Augmented Generation [RAG AI] Over Generative AI for Financial Services

**KNOW MORE. MISS NOTHING. OUTPERFORM**

[auquan.com](https://auquan.com)

SEPTEMBER 2023



# CONTENTS



i. Executive Summary	2
ii. Introduction	3
iii. <b>Section ONE:</b> The challenges with using Generative AI in the financial sector	4
iv. <b>Section TWO:</b> RAG AI: Bridging the critical gap between static knowledge of LLMs and the dynamic nature of the financial industry	6
v. <b>Section THREE:</b> The case for RAG in financial services	8
vi. <b>Section FOUR:</b> Implementing RAG AI for financial services: Three use cases	10
vii. <b>Section FIVE:</b> Future Implications	18
viii Conclusion	20
ix. References	21
x. About AUQUAN	22



# Executive Summary

In an era where artificial intelligence (AI) influences numerous facets of business, the capabilities of Large Language Models (LLMs) like OpenAI's ChatGPT are truly groundbreaking and are being integrated across diverse industry sectors. Yet, for one as regulated, dynamic, and nuanced as financial services, LLMs present significant limitations that impede their full utility. Choosing the right AI framework will be instrumental in harnessing its true potential for financial services.

This white paper explores the application of retrieval augmented generation AI (RAG) for financial services. RAG is a cutting edge enhancement to Generative AI pioneered by Meta that addresses common pitfalls of using Generative AI, particularly for financial services use cases. RAG combines the strengths of retrieval-based models, such as the ability to pull real-time, accurate information from vast databases, with the ability of generative models to craft natural, coherent responses using that information.

In the financial sector, this enhancement to Generative AI is nothing short of transformative because it enables AI systems to produce timely insights for investment decisions, enhances compliance by verifying data sources, and mitigates risks associated with outdated or generalized information.

# Introduction

In today's digitized world, Generative AI has shown tremendous promise for dramatically reshaping industries and operations. A recent McKinsey report called Generative AI the next productivity frontier, with the potential to automate work activities that absorb 60 to 70 percent of employees' time today. McKinsey estimates that work involving analyzing natural language text accounts for an average of 25 percent of total work time in enterprise.<sup>1</sup>

In the financial sector, with its vast volumes of textual content, such as company reports, regulatory documents, 10K's and 10Qs, broker research and other reports, this number could be much higher. Traditionally, highly-compensated financial professionals spend a significant amount of time sifting through lengthy documents and extensive reports in order to gather, understand, summarize, and communicate insights. Hence the sector stands to benefit immensely from Generative AI's increased ability to understand natural language and simplify this labor-intensive task, by summarizing vast amounts of data into relevant insights.

These innovations promise to empower finance professionals, including equity research analysts, risk managers, private equity associates and sustainability researchers, to redirect their focus from mundane data processing to more critical tasks, such as analyzing meaningful insights, drawing conclusions, and making informed and strategic decisions faster.

Unfortunately, as promising as Generative AI may seem, it has severe limitations for complex enterprise financial services use cases. For any information-retrieval or search system to be worthy of advanced enterprise use cases, it needs to possess three core criteria:

- **Comprehensiveness and timeliness**
- **Transparency and trustworthiness**
- **Credibility and accuracy**

In the next section, we evaluate current LLMs against these three criteria.

## SECTION ONE

## The challenges with using Generative AI in the financial sector

Generative AI, in its standard, out-of-the-box format, fails on all three criteria.

### 1. The data lacks timeliness

Current LLMs are, in essence, snapshots of the past. Their inability to incorporate real-time data means they are "frozen in time", depriving users of output that incorporates up-to-date information. Any developments past their training cutoff date, be they in financial markets, global economic shifts, or even regulatory changes, remain outside its knowledge. The static knowledge base is akin to having a financial advisor who knows the history of stock markets but isn't aware of today's macro conditions or recent company news.

### 2. Their knowledge base lacks comprehensiveness

LLMs are designed for generalized tasks. The vastness of their training means they lack domain-specific knowledge, such as niche yet impactful datasets such as sanctions lists, records of regulatory fines, or detailed shipping data. This renders them unsuitable for tasks demanding in-depth expertise.

This is a known problem of Generative AI models employed in the banking industry, they encounter difficulties in fully understanding the intricate financial context, individual circumstances, and nuanced scenarios.<sup>3</sup> Picture an investment banker in need of information regarding a potential regulatory fine in order to make an investment or compliance decision. The ramifications of making this decision without considering recent sanctions or regulatory updates could be both legally and financially disastrous.

### 3. They lack transparency, making them difficult to trust

Generative AI tools such as ChatGPT do not provide sources for the information they generate. This "black box" nature raises further concerns, as discerning the exact sources or basis for their conclusions remains challenging and requires additional work. Without transparency, it's hard to assess the reliability or bias in an AI's recommendations. If biases

exist, if conclusions are derived from outdated, non-credible, or inaccurate data, there's no way to verify or correct it. Moreover, in cases of errors or regulatory scrutiny, being unable to trace back the AI's decision-making process can result in severe compliance violations, legal repercussions, and erosion of trust among clients and regulators.

From an investment perspective, unverified guidance puts capital and reputations at risk. For compliance use cases, untraceable data sources could contravene standards. Bain placed trust is a fundamental issue with Generative AI's adoption in financial services, with hallucinations, toxicity, privacy, bias, and data governance concerns creating a trust gap.<sup>2</sup>

#### **4. The responses lack credibility or accuracy because they frequently hallucinate**

Add to this the now infamous problem of "hallucinations" with LLMs – coherent but misleading or outright fabricated responses – and the pitfalls of Generative AI for financial services become even more clear. An equity research analyst may make a recommendation based on a linguistically impeccable response, which in reality is factually inaccurate. Or consider a private equity (PE) analyst who has to verify every risk surfaced on a deal by LLMs, only to realize that half of them are fabricated. In a field where every detail counts, having to always second-guess or verify what an LLM says can slow things down and increase the chance of costly mistakes.

In high-stakes sectors like finance, where source credibility, accuracy, and timeliness are paramount, LLMs significantly underperform. Consider, for example, a financial analyst AI-copilot, powered by a generic LLM, designed to assist risk managers in assessing risks of their investments. The lack of real-time updates on the macro environment, such as recent interest rate movements, regulatory changes such as a new liquidity requirement or company-specific insights such as latest earnings release could result in catastrophic losses. Similarly, any decision made on hallucinated or inaccurate responses to queries such as "does this company have any previous accusations of fraud or bribery" or "does this company have any potential violations of the modern slavery act" could result in regulatory issues. The response may inaccurately allege wrongdoing, or miss a recent accusation, yielding additional enormous fines.



## SECTION TWO

## RAG AI: Bridging the critical gap between static knowledge of LLMs and the dynamic nature of the financial industry

So, how can financial service industry participants benefit from the capabilities of LLMs while avoiding these pitfalls? Enter Retrieval Augmented Generation (RAG) AI – a new enhancement to generative AI developed by Meta<sup>4</sup> that represents a paradigm shift for the industry

RAG AI represents a significant leap forward because it is specifically designed to tackle knowledge-intensive tasks. At its core, RAG AI merges the strengths of an AI-powered information retrieval system and Generative AI, bridging the gap between static knowledge of LLMs and the need for dynamic, context-aware generation in financial services.

**RETRIEVAL AI + GENERATIVE AI = RETRIEVAL AUGMENTED GENERATION (RAG) AI**

RAG AI can access external, up-to-date knowledge sources, ensuring responses are not only coherent but also rooted in fact. This interplay between retrieval and generation means unlike LLMs, which need extensive retraining to update their knowledge, RAG's internal knowledge can be seamlessly modified without overhauling the entire system.

### How does RAG AI work?

Just like a Generative AI system, RAG AI begins with a user or machine submitted input – like a question or a specific query. The system runs a search across an extensive and continuously updated knowledge base, much like a conventional search or retrieval engine to narrow down petabytes of potentially noisy or irrelevant data to a subset relevant to the input query. RAG then fine-tunes this subset further using semantic search techniques to information that aligns with the query's inherent intent, not just the keywords used. It translates the query into numerical representations, or vectors, using an embedding model. These vectors, which capture the core or semantic meaning of the input, enable machines to identify semantically similar entries to the initial query from the subset of data. This step ensures that the subsequent response is generated from information that is not just accurate but also contextually meaningful. This final data subset, combined with the original query, is then fed into a Large Language Model (LLM).

And this is RAG AI's distinctive edge. This enriched contextual prompt enables the LLM to generate a response that is both linguistically coherent and anchored in the most recent and accurate facts.

Consider the same financial analyst AI that aids risk managers but powered by RAG AI. When now prompted by queries such as “does this company have any previous accusations of fraud or bribery”, the system will first search through its knowledge base, which may include recent analyst notes, broker research, consultancy reports, company filings, government updates, news and legal documents, to identify any data points that directly or contextually refer to the company as well as fraud and bribery. The text from these documents is then fed to the LLM, along with the original question. The response will now include an accurate, trustworthy summary of the information which includes the latest facts.

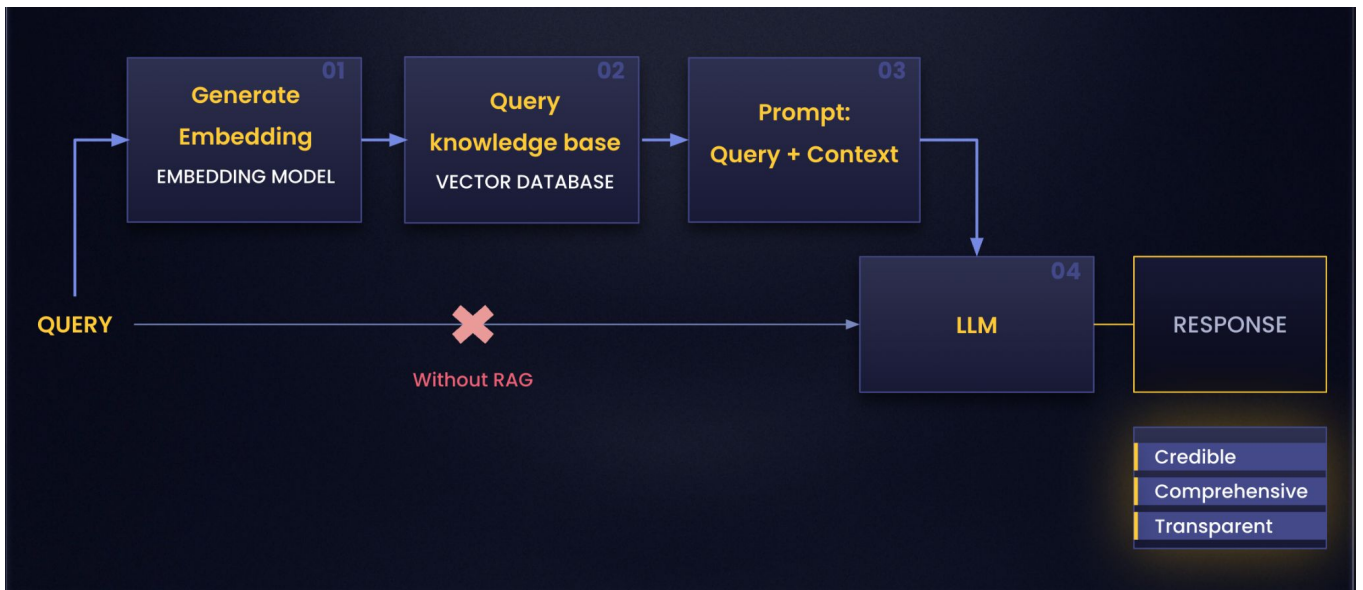


Fig. 1: Query process with and without RAG



## SECTION THREE

## The case for RAG in financial services

For financial services, the implications of upgrading from Generative AI to RAG AI are transformative:

### ■ **TIMELINESS: Real-time access to information**

The retrieval component of RAG means it has real-time access to information. Timeliness of information is paramount in financial services, both from a regulatory requirements perspective as well as the imperative that investment decisions be based on accurate and recent data. Picture the same investment banker as before, in need of information on a potential regulatory fine. Instead of grappling with a static LLM, which might give an outdated or incorrect response, they can leverage RAG AI to quickly retrieve the most recent data on the topic, ensuring their decisions are grounded in real-time insights.

### ■ **COMPREHENSIVENESS: Incorporating domain-specific information in the responses**

RAG AI's context-aware generation cross references and pulls data from datasets pertinent to financial services, such as earnings transcripts, sanctions lists, product releases, partnerships, supply chain information, and shipping data. This capability has the potential to create significant outperformance in investment decision making, enhance risk management, or save many hours in due diligence and know-your-customer (KYC) efforts.

### ■ **TRANSPARENCY AND TRUSTWORTHINESS: Citing sources to overcome the "Black Box" issue**

In an industry where regulatory compliance and accountability are crucial, RAG's ability to cite sources is invaluable. It allows any AI assisting in asset management or risk assessment to not only provide quick insights but also substantiate them with data sources. This isn't merely about enhancing accuracy. RAG brings transparency to AI's inner workings by providing a clearer view into which sources or datasets the model references, which can build trust in the system among users. For professionals in financial services, knowing the basis of an AI's recommendation – be it a market report, a recent financial study, or historical data – is mandatory.

**CREDIBILITY AND ACCURACY: Mitigating the risks of “Hallucinations”**

RAGs ability to generate responses derived only from the information supplied by the retrieval system means the AI remains rooted in relevant and material fact, greatly reducing the probability of “hallucinations”. A PE analyst now can safely use a RAG AI system to instantly surface all risks on a deal, without having to verify each one, because they have been summarized from pre-vetted and credible sources.

To sum it up, RAG AI significantly amplifies the generative capabilities of LLMs by incorporating the retrieval component, on all the core criteria necessary for any information-retrieval or search system to be worthy of advanced enterprise use cases.

**Table 1: Generative AI vs RAG AI**

CRITERIA	GENERATIVE AI	RAG AI
<b>Timeliness</b>	The knowledge base is “frozen in time” and lacks any information after training cut-off date	Retrieval component has real-time access to information
<b>Comprehensiveness</b>	The model is trained for generalized tasks	Retrieval component cross references and pulls data from datasets pertinent to specific use cases
<b>Transparency and trustworthiness</b>	“Black boxes” that do not cite their sources, making them difficult to trust	Cites sources to overcome the “black box” transparency issue
<b>Credibility and accuracy</b>	They frequently hallucinate and produce coherent but misleading or fabricated responses	Generates responses derived only from the information supplied by the retrieval system

## SECTION FOUR

## Implementing RAG AI for financial services: Three use cases

The most obvious financial sector use case for RAG AI is to identify, perform due-diligence on, and monitor investment opportunities in public markets.

Let's explore three additional use cases.

### 1. Enhancing Due Diligence for private companies

The pre-screening process for a potential investment in private equity (PE) follows the deal sourcing stage and serves as an initial evaluation to determine whether an opportunity is worth pursuing in more detail. The exact steps can vary among private equity firms depending on their focus, size, and strategy. However, a typical pre-screening process includes a review of the company's industry, product or service, and competitive position, an evaluation of growth and value-creation opportunities, a determination of the fit with firm's investment criteria, an assessment of the company's management team, and the identification of any potential risks related to market, competition, regulation, and other factors.

The process involves reviewing structured and unstructured data sources. These sources include self-disclosed information from companies, industry research by consulting firms, industry associations, and market research firms. It also includes relevant news, recent events, controversies, and regulatory filings. This process can uncover compliance issues, legal challenges, ESG research, customer reviews from forums and social media, patent and intellectual property documents, and competitor annual reports. It may also involve analyzing conference calls or presentations.

It is evident that applications developed on top of Generative AI can play a role in driving time efficiencies and enabling entirely new ways of conducting company pre screens that render old ways obsolete. The most obvious use cases include summarizing large volumes of data into concise summaries, competitive analysis, risk identification and personalized insights. However, RAG AI offers distinct advantages over traditional Generative AI to ensure accuracy, personalization, and timeliness:



### ■ **Summarizing:**

Generative AI can process voluminous data to generate concise summaries of market and product descriptions, highlighting key points about a potential investment opportunity. Adding the the retrieval component of RAG ensures this data is summarized only from credible information on the company and not hallucinated. The summarization becomes a reliable snapshot that PE analysts can trust for decision-making. For example, from hundreds of verified customer reviews, RAG AI can produce a summary capturing the major advantages as well as pain points of a company's product.

### ■ **Competitive Analysis:**

AI can generate competitive landscapes much faster than humans by comparing data from multiple sources, including but not limited to regulatory filings, online reviews, customer blogs etc. RAG AI can be even more effective at producing robust competitive sets because the retrieval component acts as a company reference database of firmographic information such as a location, size and ownership status while the generative component can act as a cognitive service, comparing different companies and selecting the more relevant ones. It can hone in on the most relevant competitors based on firmographic information and contextual needs of the PE firm, offering a more insightful analysis.

### ■ **Risk Identification:**

Generative AI can sift through large volumes of data to uncover and flag potential risks, from regulatory violations to negative news mentions, providing a holistic risk profile of a potential investment. Adding the retrieval component of RAG ensures these risks are generated from accurate and credible information. Importantly, the "adaptive" nature tailors the risk profile to align with the specific concerns of a PE firm and its exclusion criteria, whether that's regulatory issues or market competition.

### ■ **Customized Presentation of Insights:**

Generative AI can be very useful in formulating the basic building blocks of an IC paper, that can be expanded by a deal team member with more context and insight, especially accelerating the process of drafting repetitive and mundane sections of the IC report. Using RAG, this report can be tailored depending on the specific interests or concerns of a PE firm. For example, one



firm might prioritize ESG considerations while another is more focused on intellectual property strength.

**Table 2: Generative AI vs RAG AI in Private Company Screening**

USECASE	GENERATIVE AI	RAG AI
<b>Summarizing</b>	Processes large datasets to provide summaries of a company's market and product descriptions	Ensures these summaries are derived only from credible, up-to-date sources, eliminating the risk of 'hallucinated' information
<b>Competitive Analysis</b>	Can churn out competitive landscapes and comparative reports by scanning multiple sources	Further refines comparative reports by accessing a rich company reference database via its retrieval component
<b>Risk Identification</b>	Highlights potential risks to investment from a broad set of data.	Adds a layer of trust by sourcing risk data from credible references and adapting them to the firm's needs
<b>Customized Presentation of Insights</b>	Helps in generating the basic components of Investment Committee (IC) paper	Not only assists in drafting the paper but also customizes it to suit the priorities of a specific PE firm

In summary, while Generative AI can offer a base layer of automation and data processing, RAG AI takes it a notch higher by ensuring that the information is accurate, up-to-the-minute, and highly personalized to the unique needs of each PE firm. This focus on context-aware intelligence makes RAG AI a superior technological choice for driving efficiency and precision in the private equity pre-screening process.

## 2. Running Rapid KYC and Background Checks

Conducting Know Your Customer (KYC) and background checks in finance is essential before onboarding a customer as a loan counterparty in an investment

The typical process includes collecting initial identification and verifying documents. Compliance checks like PEP and sanctions list screening are run. Additional due diligence is conducted to assess financial stability, creditworthiness, reputational risks, and past incidents of fraud or money laundering. This information is used to create a risk profile. The compliance team or higher management reviews the profile for approval or denial. It is also used to determine the level of monitoring required post-onboarding for compliance and risk reassessment.

This multi-step process can involve a myriad of documents and unstructured data sources, including:

- **Media articles and reports** to assess reputational risks, and to identify any past controversies or operational risks.
- **Social media** to gauge public perception and reputation.
- **Industry reports** to understand the business's position in the industry.
- **Legal filings** to spot any pending or past lawsuits or regulatory actions.
- **Regulatory notices** to identify any compliance-related issues.
- **Local and foreign watchlists** to look for any flags related to sanctions or other regulatory matters.
- **Internal notes and communications** for insights into previous interactions with the entity, if applicable.

Generative AI-based tooling can deliver value here by summarizing large volumes of structured and unstructured data and offering contextual insights, helping compliance officers to understand the basic risk profile of a corporate client faster. Upgrading from Generative AI to RAG AI, adds another layer of specificity and credibility to this process. RAG AI can ensure that the model's outputs are rooted in verified data and tailored to the contextual needs of the compliance team, thereby enhancing the accuracy, personalization, and timeliness of the KYC and background checks.



### ■ **Ensuring Accuracy and Credibility for Regulatory Compliance:**

Generative AI can produce text based on patterns and examples it has seen during its training but may inadvertently include inaccurate or unverified information. This could lead to regulatory fines or reputational damage for the financial institution. RAG AI, on the other hand, retrieves data from specific, approved sources before generating its responses. This ensures that any flags for potential risks like money laundering, fraud, or sanctions violations are grounded in verified data, enhancing compliance assurance.

### ■ **Personalization to Company's Internal Risk Criteria and Controls:**

Generative AI operates with limited context about a firm's specific risk parameters or compliance protocols. RAG AI can be tailored to retrieve and generate information that aligns with the internal requirements of the specific investment bank doing the onboarding. For instance, one bank might have tighter controls around anti-money laundering, while another may prioritize ESG compliance.

### ■ **Dynamic Risk Profiling and Real-time Alerting:**

Generic LLMs will overlook real-time issues and concerns on a company, as its training data becomes outdated. This can lead to significant gaps in a counterparty's risk profile, potentially causing financial and regulatory repercussions. RAG AI can be set to pull from up-to-date sources in real-time for timely identification of risks, especially in a regulatory landscape that is ever-evolving. It can also be set up for the creation of dynamic risk profiles with immediate alerting if a critical compliance issue is identified to assist in ongoing due diligence and make the post-onboarding monitoring more efficient and responsive.

In summary, while Generative AI has its merits, RAG AI's integration into the risk analysis process ensures that responses are not only accurate and up-to-date but are also aligned precisely with a risk manager's unique needs. This translates into rapid, in-depth KYC and background checks, equipping financial institutions with the tools to make informed decisions swiftly, minimizing exposure to potential risks.

### 3. Supporting ESG Sustainability Research

There are an estimated \$10 trillion worth of investments tied to special ESG criteria, and despite the backlash against it, sustainability linked research is on the rise, both in public and private market investing. KPMG's 2023 ESG Due Diligence study found that more than half (53%) of 200 investors surveyed have canceled M&A deals because of material findings in ESG due diligence.<sup>5</sup> Even in the US, where the anti-ESG rhetoric is loudest, 82% of investment managers use ESG analysis to make decisions, per Russell Investments.<sup>6</sup>

Sustainability research is a complex process involving pre-investment screening and analysis to identify companies that align with ESG fund criteria. It also includes post-investment monitoring to demonstrate active ownership, vote on ESG-related issues, and report on ESG performance.

Investment managers need to analyze corporate sustainability reports, regulatory filings, third-party ESG ratings from agencies like Sustainalytics, MSCI or similar, news articles, patent filings, proxy voting records, and more. The ultimate aim is to ensure that investment decisions align with predefined ESG criteria and offer both financial and ethical returns.

Standard LLMs on their own can be extremely useful in automating both pre-investment and post-investment research. For example, they can be used to sift through large sets of structured and unstructured data including company reports, news articles, regulatory filings to auto-generate a nuanced view of a company's ESG performance and benchmark against industry standards. RAG AI, by combining the contextual reasoning power of Generative AI with the ability to monitor multiple data streams in real-time, offers several advantages that can enhance the depth of ESG evaluations in asset management.

#### ■ **Credibility and Transparency:**

RAG AI is designed to delve into a deeper context, fetching information not just from a predefined training dataset – but from multiple correlated sources. The additional ability to maintain a trail to the original source of the data and cite them directly in the generated reports contributes to transparency and allows for a clear understanding of how different data pieces contributed to an ESG score. This is crucial for asset managers and stakeholders who want to



understand the reasoning behind ESG evaluations, at a moment when ESG scores are facing immense scrutiny.

■ **Timeliness:**

Similar to KYC use case, RAG AI can be set up to continuously monitor specific ESG metrics or news sources in real-time, ensuring that asset managers have the most up-to-date information when making decisions.

■ **Avoiding Greenwashing:**

RAG AI can cross-reference information across multiple, continually updating sources to verify the validity of ESG claims made by companies against industry benchmarks, third-party reports, and actual performance metrics, helping to detect exaggerated ESG claims and call out greenwashing attempts.

By employing RAG AI, funds can achieve a new level of sophistication in their ESG research, combining depth of analysis with credibility, transparency, timeliness and trustworthiness in a way that standard Generative AI models might not be able to match, enabling them to make better-informed investment decisions and foster more responsible stewardship.

---

## QUICK SUMMARY

Primary use cases for RAG AI:

### 1. Enhancing Due Diligence for private companies

■ **Improved Summarizing**

■ **Effective Competitive Analysis**

■ **Improved Risk Identification**

■ **Customized Presentation of Insights**

## QUICK SUMMARY

Primary use cases for RAG AI:

### 2. Running Rapid KYC and Background Checks

- Ensuring Accuracy and Credibility for Regulatory Compliance
- Personalization to Company's Internal Risk Criteria and Controls
- Dynamic Risk Profiling and Real-time Alerting

### 3. Supporting ESG/Sustainability Research

- Improved Credibility and Transparency:
- Ensuring Timeliness
- Avoiding Greenwashing

## SECTION FIVE

## Future Implications

As enterprise AI implementations gain steam, we expect to see the restrictions that organizations are placing on the adoption of AI to give way to policies designed to facilitate the safe adoption of AI, which is now more possible with innovations such as RAG AI. Internal and competitive pressure will be the primary drivers of this shift in policy.

We also anticipate further innovation with RAG AI, particularly in ways that cater automatically to the specific and unique needs of the user in order to deliver more personalized, contextual insights. In financial services, this would mean systems can craft differentiated and custom responses for PE analysts, risk managers, compliance managers, equity research analysts or ESG analysts, focusing on the insights that are the most relevant for those users.

Why is this important? Because one-size-fits-all approaches fall short of meeting the different needs and priorities that different professionals in this industry have. The same underlying data available on a single company could supply different users with distinct, relevant insights tailored to their roles and responsibilities.

For example: an equity research analyst needs to dig deep into the nitty-gritty of a company's current quality of earnings and future revenue potential; a risk manager will be interested in identifying and understanding idiosyncratic risks that might erode investment capital; and a sustainability analyst might focus on reported emissions metrics or potential ESG violations that could impact a company's long-term reputation and sustainability.

By delivering personalized, contextual insights, RAG-based systems can streamline decision-making processes and enhance risk management strategies.

### ■ Deeper domain-specific coverage::

By adaptively selecting the right knowledge base, RAG-based systems could offer even more domain-specific expertise, be it in private company diligence, equity research on a specific sector like healthcare, or any niche area.

### ■ Increased usability

RAG-based systems could further tailor responses based on user persona and

context, ensuring precision in delivering relevant information. By focusing on persona-specific details RAG-based systems can further optimize user productivity.

■ **More hidden insights:**

When working with vast amounts of unstructured data, critical yet hidden often go unnoticed. RAG-based systems have the potential to bring an extreme level of focus by zeroing in on context, user persona, and specific dataset allows it to surface hidden information that humans and Generative AI tools typically overlook.



# Conclusion

Generative AI has not lived up to the expectations of the financial services industry, as it has failed to meet the core criteria of being timely, comprehensive, transparent, trustworthy, and accurate. RAG AI meets these core criteria by combining the strengths of information retrieval systems and generative AI to meet the knowledge-intensive demands of financial services.

RAG-based systems can access external knowledge sources, providing coherent and fact-based responses without requiring retraining. And RAG-based systems have real-time access to current information and comprehensive data specific to financial services. They're transparent with regard to information sourcing, and can focus their output on relevant and factual information.

As adoption of AI in financial services moves apace, expect to see more RAG-based solutions implemented. Innovation will focus on providing deeper domain-specific coverage, increasing usability, and uncovering more hidden insights in unstructured data. And enhanced personalization and contextualization of system output is in demand, therefore we anticipate catering to the specific needs of different professionals in the industry.

RAG AI systems have the potential to transform a variety of financial services use cases and help finance professionals make more informed and strategic decisions faster – and outperform their peers.

# References

1. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction>
2. <https://www.bain.com/insights/five-principles-for-generative-ai-in-financial-services/>
3. <https://masterofcode.com/blog/red-flags-and-best-practices-for-generative-ai-in-telecom-banking-and-finance>
4. <https://research.facebook.com/file/4283170945104179/Retrieval-Augmented-Generation-for-Knowledge-Intensive-NLP-Tasks.pdf>
5. <https://www-consulting-us.cdn.ampproject.org/c/s/www.consulting.us/news/amp/9509/kpmg-more-than-half-of-firms-have-cancelled-ma-deals-because-of-esg-findings>
6. <https://russellinvestments.com/us/insights/articles/2022-esg-manager-survey>

# About AUQUAN

Auquan is an AI innovator transforming vast amounts of unstructured data into actionable intelligence for financial services customers like UBS and Federated Hermes. Professionals in investment banking, private equity, asset management, and other financial institutions use Auquan's SaaS-based Intelligence Engine to discover hidden value in seemingly worthless data and identify financially-material ESG, reputational, and regulatory risks on more than 500k private companies, equities, infrastructure projects, and other entities. Auquan leverages retrieval augmented generation (RAG) – a cutting-edge AI technique that addresses common pitfalls of using generative AI in the enterprise to consistently produce timely, comprehensive, and accurate insights.

---

## DISCLAIMER:

The document herein and the information encapsulated within are provided on a confidential basis, with the sole intent of probing potential business opportunities between the disclosing entity (Auquan Ltd.) and the recipient. This document is not to be shared with any third party, nor used for any other purposes, without the explicit written consent of the disclosing party. The information contained within this report should not be used as a replacement for, or be presumed to be, specific professional, particularly financial, advice. Auquan categorically disavows any responsibility for any loss or damage incurred by any individual or entity concerning any actions or lack thereof pertaining to the content of this report.

Auquan does not provide any express or implied warranty, endorsement, or representation regarding the information contained herein, the potential results deriving from the usage of the information, or any other matter. The role of Auquan is to gather information from public sources and disseminate it in the form of this report.

Auquan expressly rejects, and the reader hereby relinquishes, all implied warranties, inclusive of, but not limited to, warranties of originality, precision, completeness, merchantability, suitability for a particular use, and warranties associated with potential infringements of intellectual property rights, trademark rights, or any other rights belonging to a third party. This report may be cited or employed for business objectives, provided that the report is used in its entirety and Auquan is expressly acknowledged as the source. Auquan reserves all copyrights and rights of origin to the content of this report.