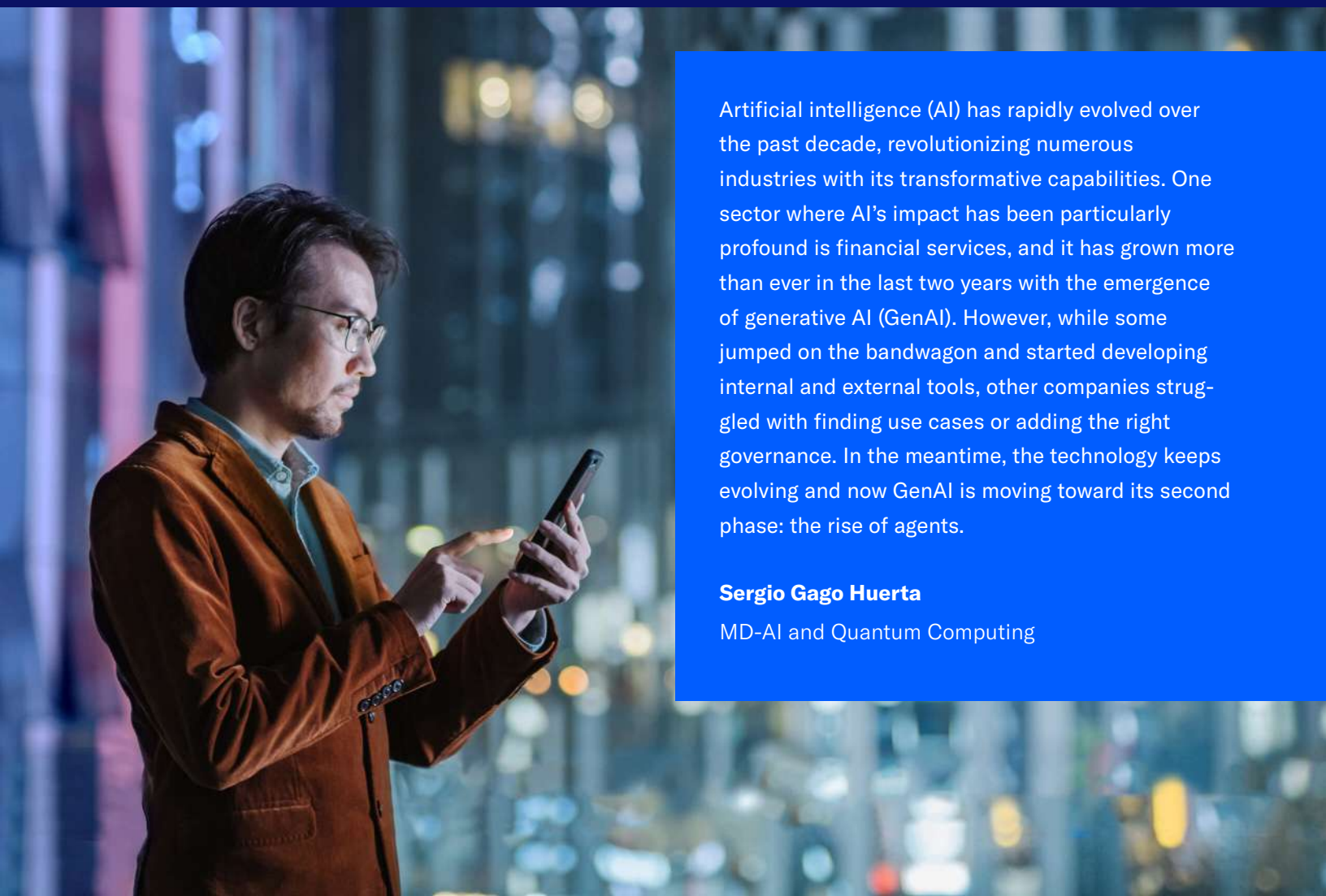
MOODY'S

GenAI's transformative potential in the financial sector: the evolution of agents



Artificial intelligence (AI) has rapidly evolved over the past decade, revolutionizing numerous industries with its transformative capabilities. One sector where AI's impact has been particularly profound is financial services, and it has grown more than ever in the last two years with the emergence of generative AI (GenAI). However, while some jumped on the bandwagon and started developing internal and external tools, other companies struggled with finding use cases or adding the right governance. In the meantime, the technology keeps evolving and now GenAI is moving toward its second phase: the rise of agents.

Sergio Gago Huerta
MD-AI and Quantum Computing

MOODY'S

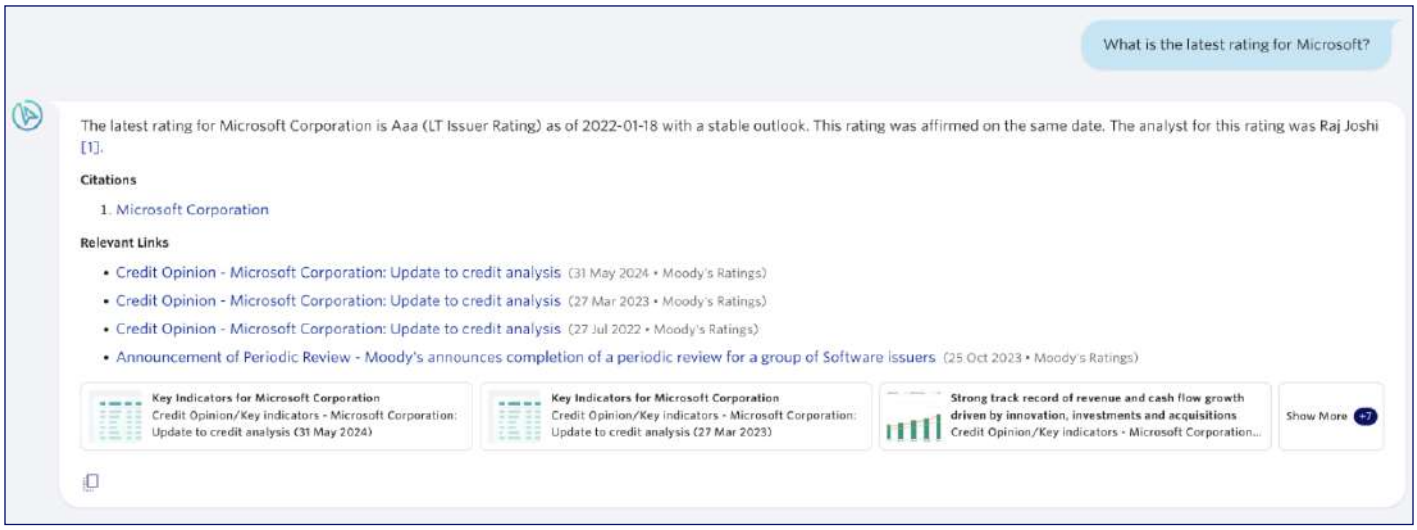
Moody's saw very early on that this technology would be a game-changer for the financial industry, and since then we have been on a GenAI journey involving creating a set of skills, assistants, and navigators and being the first financial services company to launch a GenAI tool in the market, the Moody's Research Assistant. We have also been busy researching AI agents since we believe they could potentially create a new standard for excellence in financial analytics.

But before diving into Moody's achievements, it's essential to grasp what AI agents are and how they differ from conventional chatbots.

Traditional chatbots operate based on predefined rules and scripts. They follow a set of programmed instructions to respond to user queries, often resulting in limited interactions. For instance, if a user asks a rule-based chatbot about a given company rating or financial performance, the chatbot might respond with a predetermined set of responses or direct the user to a specific webpage. These cases leverage the retrieval-augmented generation (RAG) framework that combines the power of large language models (LLMs) with individual companies' datasets.

In Moody's case, our datasets include decades of research, ratings, articles, and macroeconomic information about the world. RAG is effectively a way of telling the LLM, "We don't necessarily trust your internal knowledge and the data you have been trained on, so please use this knowledge base I am giving you access to — and by the way, please make sure you provide citations and sources to every claim you make."

Figure 1 Moody's Research Assistant provides direct access to the sources used for a given answer — for example, the last rating action for a given company as well as additional details such as financials.



MOODY'S

We have defined four stages of GenAI maturity: assisted intelligence, augmented intelligence, automated intelligence, and autonomous intelligence.

1. Initially, everyone begins with basic chatbots and rudimentary RAG frameworks, which fall under the assisted intelligence stage. At this level, AI is basic, prone to hallucinations, and requires extensive human review.

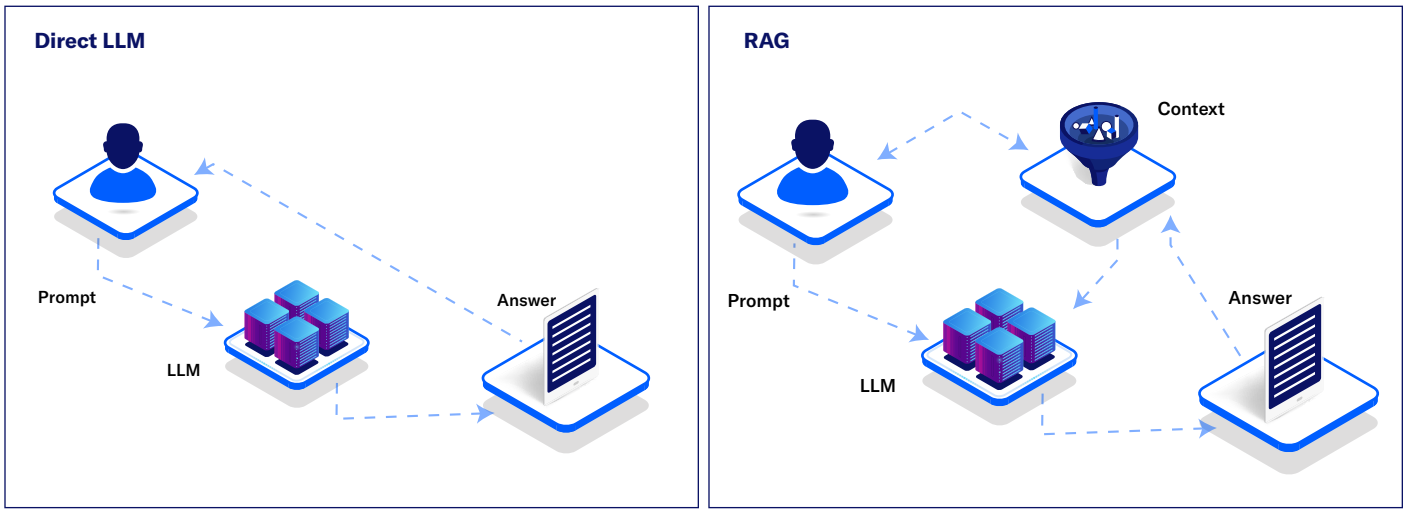
2. Augmented intelligence brings context-specific capabilities, allowing models to assess implications and determine the appropriate tools for each situation. This stage can be considered advanced RAG and tool usage, but human review and decision-making are still necessary.

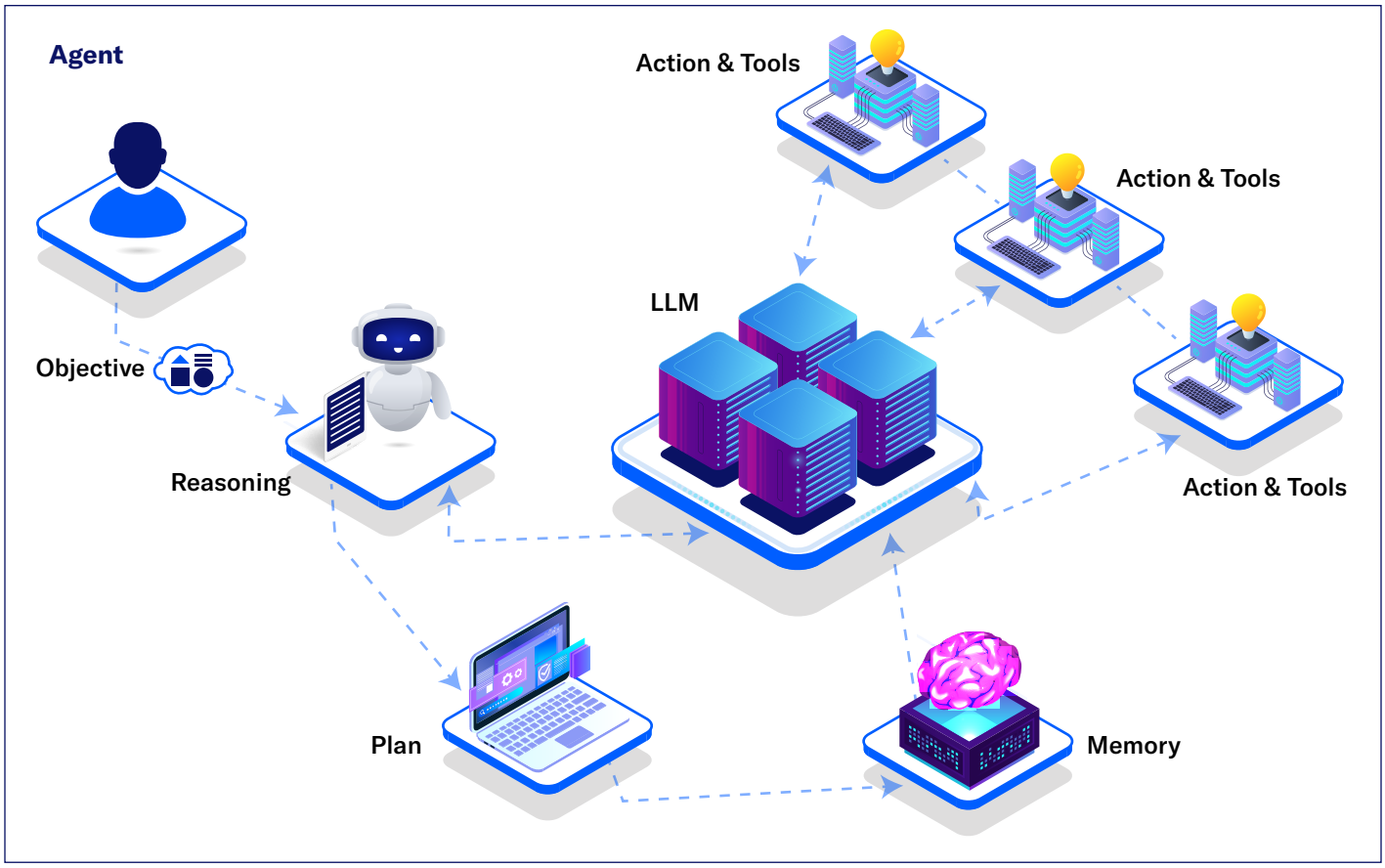
3. The next step is automated intelligence, where AI can execute small tasks or make basic recommendations. Here, models are provided with extensive context, agency, and rules, along with benchmarks and quality controls. Examples of such automation include entity or address-matching, data hydration, and anomaly detection. This type of automation is similar to what classical machine learning has achieved but is now applicable to a broader domain.

4. Lastly, autonomous intelligence involves AI's ability to plan, execute those plans, evaluate outcomes, and adapt accordingly. This may involve groups of specialized agents contributing their unique perspectives and using different skills and multi-modal interaction (vision, voice, and even robotics). At this stage, we are exploring which tasks and projects are suitable. Examples include software development and analyzing a small business's financials with the same level of scrutiny applied to a Fortune 500 company.

AI agents are designed to bring us to the last stage of autonomous intelligence. They can handle more complex tasks where a simple question and answer may not be enough. Some examples could be analyzing financial reports, generating insights, and even making predictions. The agents can understand nuanced queries, learn from interactions, and provide tailored responses that evolve over time.

Figure 2 How agentic workflows really differ from RAG or pure LLM question and answer





For example, imagine an AI agent designed to assist investors. Unlike a traditional chatbot that might only provide predefined sets of information, this AI agent can analyze market trends, interpret financial statements, and offer personalized investment recommendations based on the user's portfolio and risk tolerance; it can even generate a full report in the user's desired format. The agent can do that by reasoning and planning in an iterative way, whereas the chatbot would try to write the report in one go from top to bottom. Just like with human workers, the process is more efficient and yields better results when you first create a structure, research the topics, create a first draft, and have a colleague review it and provide ideas or suggestions until you have a final version. That is what agents do.

Now, imagine several LLM-powered agents that can interact with each other and have different specialties. For example, we could have a credit risk specialist agent; an environmental, social, and governance one; a news analyzer; several business analysts; and a team of coders that could create software on demand to fetch information or generate insights. These agents are the equivalent of having an army of assistants working together as a team to provide a solution for a business goal rather than just answering a question.

Moody's has been at the forefront of deploying AI agents, leveraging several advanced frameworks to enhance their functionality. During 2023 we extensively experimented with a multi-agent simulation framework. By the end of 2023 and going into 2024, we saw fundamental improvements in the technology that could lead us to build enterprise-grade production applications. On one hand, LLMs are getting better, faster, and more accurate.

On the other hand, agent frameworks are now more efficient and easier to run. Below are three examples of frameworks we currently use and have based our research on:

- **Autogen:** Autogen is known for its advanced natural language processing capabilities. It allows Moody's AI agents to generate high-quality financial reports and insights by interpreting large volumes of data. For example, Autogen can automatically draft comprehensive credit reports by analyzing financial statements, market conditions, and other relevant datapoints.

- **CrewAI:** CrewAI focuses on collaborative AI applications, enhancing interactions between human analysts and AI agents. This framework is particularly useful in scenarios where human expertise and AI capabilities need to complement each other. For instance, CrewAI can assist analysts in identifying emerging market trends by analyzing data patterns and suggesting actionable insights.

- **Langgraph:** Langgraph is part of the Langchain ecosystem, employs a graph-based system to represent agent workflows, and can be used to analyze complex data relationships. Langgraph can visualize various financial indicators' interconnectedness and predict potential market shifts.

Some potential use cases where multi-agentic workflows could excel include:

Automated credit analysis: Traditionally, credit analysis involves a meticulous review of financial statements, market conditions, and qualitative factors. Powered by Autogen, Moody's AI agents streamline this process by automatically generating detailed credit reports. For instance, when assessing a company's creditworthiness, the AI agent analyzes its financial history, industry trends, and economic conditions to produce a comprehensive credit assessment. This is especially useful when you want to evaluate smaller private companies for which there is not a lot of available data.

Enhanced risk management: Managing financial risk requires constant monitoring and analysis of market conditions. Moody's uses AI agents to predict potential risks by analyzing historical data and market trends. For example, an AI agent might forecast potential economic downturns by evaluating indicators such as unemployment rates, interest rates, and geopolitical events. This requires several agents discovering the outliers and how they are correlated as well as creating different scenarios and outcomes.

Software development/service as a software: Agents can behave like authentic software development squads. One agent can be a project manager, another a quality assurance specialist, another a back-end developer, and another a front-end developer. Examples like Devin are proliferating around the world, and while it is still early days for fully fledged applications, automated agents can safely wield smaller tools. For example, queries like "Find the daily stock prices for the last two years for each ticker in my portfolio" can now be fully automated with a very easy and simple query.

Figure 3 The Autogen framework automatically created a plot for stock market price evolution.

Below we have an agent automatically plotting a trading graph with the last two years of stock prices with just one simple, unoptimized prompt — all in about 10 seconds:

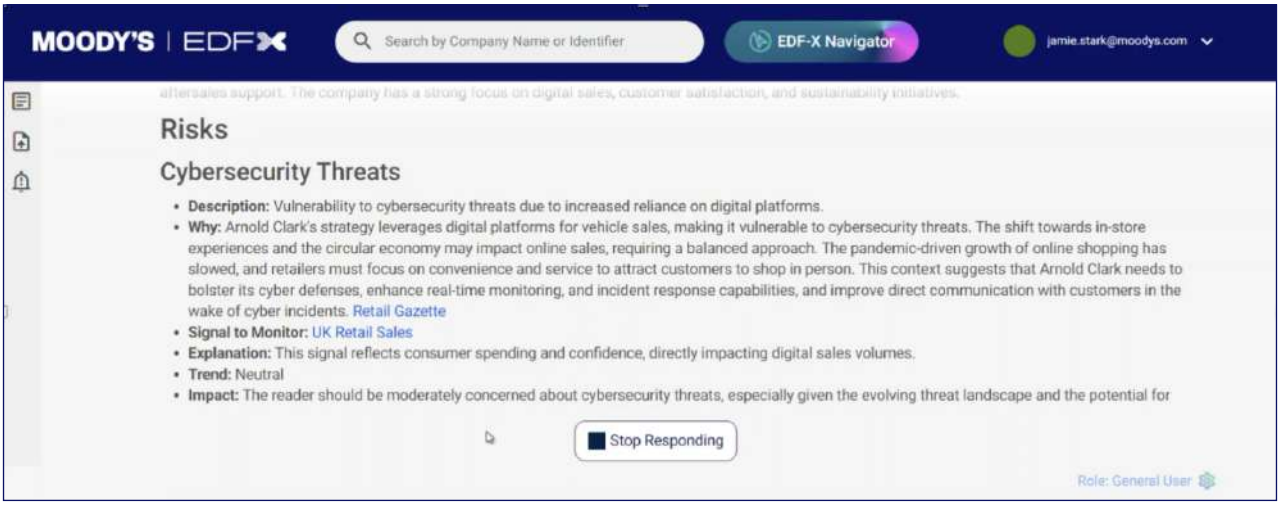
Beneath the surface, the agent built software that connected to the internet, installed the right libraries, downloaded the stock prices, and printed the plot. The industry calls this "service as a software," as opposed to software as a service (SAAS).

Figure 4 The actual source code that an agent creates and executes to produce the desired output for the user

(These screen captures took a few seconds to produce using Autogen Studio, one of the available user interfaces for the Autogen framework.)

However, for more extensive cases in production, agents need to be tamed and their abilities constrained, from avoiding hallucinations to maintaining alignment with user requirements. Moody's has put this concept to work in one of our latest products, ReconAI, which is incorporated in our EDF-X platform. In this product, a collection of agents evaluates all the potential metrics that could trigger Early Warning Signals for any company of any size in a dynamic and customized way. For example, since we can trigger infinite agents, we can evaluate those Early Warning Signals from any public or private company and not just the Fortune 500 organizations that typically get analyzed. Another collection of agents can monitor everything that happens in the world, from news events to controversies to macroeconomic indicators to regulatory filings, and extract the signal on which companies will be affected — all in real time.

Figure 5 A screenshot of a ReconAI agent evaluating which signals to monitor for specific cybersecurity threats to a given company



MOODY'S

As a company integrates AI agents into its operations, evaluating their performance and effectiveness is crucial. There are several methods companies can take to help improve agentic-driven solutions' accuracy and reliability.

- **Retrieval-augmented generation evaluation:** As previously mentioned, RAG is a framework that combines information retrieval with generative models. Moody's uses Ragas, which is an automated evaluation of RAG with LLM as a judge — that is, another LLM is "asked" for the correctness of the answer following an extremely specific scorecard and criteria) among other frameworks to evaluate our AI agents' performance. This approach involves using LLM judges to assess the quality of generated content when reference data is unavailable. BERTScore, which measures the contextual similarity between two passages, is also applied when reference data is present to provide additional metrics on responses' relevance and accuracy.

- **Search engine/retrieval performance:** AI agents are also evaluated based on their performance in search-related tasks. Moody's leverages LLM judges to assess search results' relevance and AI agents' effectiveness in retrieving relevant information.

- **Multiple agent voting/mixture of experts:** The last method we use and experiment with is a mixture of agents and voting. Foundational models may have biases, system prompts, and guardrails that influence their answers. In some cases, they may be less confrontational. In several of our use cases, we execute the same task through several models (for example, GPT-4, Claude 2, Llama 3, Gemini, and a Moody's fine-tuned model) and hold a majority voting decision. Each answer provided by each of the underlying models gets scored and evaluated; this way we increase the accuracy and reduce the hallucination possibilities.

This is just the beginning, though. Thanks to having worked on classical machine learning, we had many of the basics covered. But this industry and technology move very fast, especially in benchmarking, which includes some of our following research lines:

- **Specialized models:** Research is underway to develop specialized models for LLM-as-a-judge, such as Google's Cappy and KAIST's Prometheus 2. These models aim to improve AI judgments' accuracy and consistency.

- **Financial benchmark datasets:** Moody's is working on creating benchmark datasets to measure LLM performance across different applications. These datasets will help detect performance drift and maintain LLM API providers' quality of service.

- **Comparative studies:** Future research will focus on comparing specialized versus general-purpose LLMs for evaluation purposes. Developing consistent scoring mechanisms for LLM judges is also a priority to enhance evaluation reliability.

As Moody's continues to lead the way in AI agent technology, several trends and developments are shaping the future of financial services, including increasing automation and depth (i.e., applying the same analysis to every company regardless of its size), enhanced personalization and contextualization, and greater integration across financial systems.

MOODY'S

Critical technical components are in place to be leveraged by RAG and Tuning approaches

01	02	03	04	05
Data interoperability	**Entity resolution**	**Data lineage and trust**	**API architecture**	**Permissions and entitlements**
What do you have on Tesla in all your databases?	That's not what I meant. MCO is not the Orlando Airport.	Where did this data come from? Can I trust it?	Can I easily plug Moody's data into my existing workflow tools?	Who is authorized to access this data?

AI safety considerations: As we advance through the stages of GenAI maturity, it is crucial to comprehensively address AI safety. It is paramount that AI systems operate reliably and ethically.

This involves implementing robust safety protocols, continuous monitoring, and rigorous testing to prevent unintended consequences. By prioritizing AI safety, we can build trust and confidence in AI technologies and make sure they are used responsibly and beneficially.

7

Moody's journey in adopting and deploying AI agents exemplifies GenAI's transformative potential in the financial sector. By leveraging advanced frameworks, addressing evaluation challenges, and exploring future research directions, Moody's continues to set new benchmarks in financial analytics and risk management. As AI technology evolves, AI agents' role in enhancing financial services will become increasingly pivotal, driving innovation and delivering greater value to stakeholders.

Even though the technology had been around for several years, 2022 saw the birth of GenAI, while 2023 was the year of RAG. In 2024, we're seeing GenAI 2.0's evolution with agents. While it is still in its infancy and we have a lot of ground left to cover, the development speed is rapid and the impact is potentially exponential. Being at the forefront of applied research yields great benefits on how our industry is evolving. Moody's was the first financial institution with a GenAI product in the market with paying customers, and this year we will be the first with an agentic platform under our EDF-X suite.

By taking these steps, enterprises can deepen their understanding of AI and truly embrace the opportunities that unfold as the technology continues to evolve.

MOODY'S

© 2024 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "MOODY'S"). All rights reserved.

CREDIT RATINGS ISSUED BY MOODY'S CREDIT RATINGS AFFILIATES ARE THEIR CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES, AND MATERIALS, PRODUCTS, SERVICES AND INFORMATION PUBLISHED OR OTHERWISE MADE AVAILABLE BY MOODY'S (COLLECTIVELY, "MATERIALS") MAY INCLUDE SUCH CURRENT OPINIONS. MOODY'S DEFINES CREDIT RISK AS THE RISK THAT AN ENTITY MAY NOT MEET ITS CONTRACTUAL FINANCIAL OBLIGATIONS AS THEY COME DUE AND ANY ESTIMATED FINANCIAL LOSS IN THE EVENT OF DEFAULT OR IMPAIRMENT. SEE APPLICABLE MOODY'S RATING SYMBOLS AND DEFINITIONS PUBLICATION FOR INFORMATION ON THE TYPES OF CONTRACTUAL FINANCIAL OBLIGATIONS ADDRESSED BY MOODY'S CREDIT RATINGS. CREDIT RATINGS DO NOT ADDRESS ANY OTHER RISK, INCLUDING BUT NOT LIMITED TO: LIQUIDITY RISK, MARKET VALUE RISK, OR PRICE VOLATILITY. CREDIT RATINGS, NON-CREDIT ASSESSMENTS ("ASSESSMENTS"), AND OTHER OPINIONS INCLUDED IN MOODY'S MATERIALS ARE NOT STATEMENTS OF CURRENT OR HISTORICAL FACT. MOODY'S MATERIALS MAY ALSO INCLUDE QUANTITATIVE MODEL-BASED ESTIMATES OF CREDIT RISK AND RELATED OPINIONS OR COMMENTARY PUBLISHED BY MOODY'S ANALYTICS, INC. AND/OR ITS AFFILIATES. MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND MATERIALS DO NOT CONSTITUTE OR PROVIDE INVESTMENT OR FINANCIAL ADVICE, AND MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND MATERIALS ARE NOT AND DO NOT PROVIDE RECOMMENDATIONS TO PURCHASE, SELL, OR HOLD PARTICULAR SECURITIES. MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND MATERIALS DO NOT COMMENT ON THE SUITABILITY OF AN INVESTMENT FOR ANY PARTICULAR INVESTOR. MOODY'S ISSUES ITS CREDIT RATINGS, ASSESSMENTS AND OTHER OPINIONS AND PUBLISHES OR OTHERWISE MAKES AVAILABLE ITS MATERIALS WITH THE EXPECTATION AND UNDERSTANDING THAT EACH INVESTOR WILL, WITH DUE CARE, MAKE ITS OWN STUDY AND EVALUATION OF EACH SECURITY THAT IS UNDER CONSIDERATION FOR PURCHASE, HOLDING, OR SALE.

MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS, AND MATERIALS ARE NOT INTENDED FOR USE BY RETAIL INVESTORS AND IT WOULD BE RECKLESS AND INAPPROPRIATE FOR RETAIL INVESTORS TO USE MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS OR MATERIALS WHEN MAKING AN INVESTMENT DECISION. IF IN DOUBT YOU SHOULD CONTACT YOUR FINANCIAL OR OTHER PROFESSIONAL ADVISER.

ALL INFORMATION CONTAINED HEREIN IS PROTECTED BY LAW, INCLUDING BUT NOT LIMITED TO, COPYRIGHT LAW, AND NONE OF SUCH INFORMATION MAY BE COPIED OR OTHERWISE REPRODUCED, REPACKAGED, FURTHER TRANSMITTED, TRANSFERRED, DISSEMINATED, REDISTRIBUTED OR RESOLD, OR STORED FOR SUBSEQUENT USE FOR ANY SUCH PURPOSE, IN WHOLE OR IN PART, IN ANY FORM OR MANNER OR BY ANY MEANS WHATSOEVER, BY ANY PERSON WITHOUT MOODY'S PRIOR WRITTEN CONSENT. FOR CLARITY, NO INFORMATION CONTAINED HEREIN MAY BE USED TO DEVELOP, IMPROVE, TRAIN OR RETRAIN ANY SOFTWARE PROGRAM OR DATABASE, INCLUDING, BUT NOT LIMITED TO, FOR ANY ARTIFICIAL INTELLIGENCE, MACHINE LEARNING OR NATURAL LANGUAGE PROCESSING SOFTWARE, ALGORITHM, METHODOLOGY AND/OR MODEL.

MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND MATERIALS ARE NOT INTENDED FOR USE BY ANY PERSON AS A BENCHMARK AS THAT TERM IS DEFINED FOR REGULATORY PURPOSES AND MUST NOT BE USED IN ANY WAY THAT COULD RESULT IN THEM BEING CONSIDERED A BENCHMARK.

All information contained herein is obtained by MOODY'S from sources believed by it to be accurate and reliable. Because of the possibility of human or mechanical error as well as other factors, however, all information contained herein is provided "AS IS" without warranty of any kind. MOODY'S adopts all necessary measures so that the information it uses in assigning a credit rating is of sufficient quality and from sources MOODY'S considers to be reliable including, when appropriate, independent third-party sources. However, MOODY'S is not an auditor and cannot in every instance independently verify or validate information received in the credit rating process or in preparing its Materials.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability to any person or entity for any indirect, special, consequential, or incidental losses or damages whatsoever arising from or in connection with the information contained herein or the use of or inability to use any such information, even if MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers is advised in advance of the possibility of such losses or damages, including but not limited to: (a) any loss of present or prospective profits or (b) any loss or damage arising where the relevant financial instrument is not the subject of a particular credit rating assigned by MOODY'S.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability for any direct or compensatory losses or damages caused to any person or entity, including but not limited to by any negligence (but excluding fraud, willful misconduct or any other type of liability that, for the avoidance of doubt, by law cannot be excluded) on the part of, or any contingency within or beyond the control of, MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers, arising from or in connection with the information contained herein or the use of or inability to use any such information.

NO WARRANTY, EXPRESS OR IMPLIED, AS TO THE ACCURACY, TIMELINESS, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OF ANY CREDIT RATING, ASSESSMENT, OTHER OPINION OR INFORMATION IS GIVEN OR MADE BY MOODY'S IN ANY FORM OR MANNER WHATSOEVER.

Moody's Investors Service, Inc., a wholly-owned credit rating agency subsidiary of Moody's Corporation ("MCO"), hereby discloses that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by Moody's Investors Service, Inc. have, prior to assignment of any credit rating, agreed to pay to Moody's Investors Service, Inc. for credit ratings opinions and services rendered by it. MCO and Moody's Investors Service also maintain policies and procedures to address the independence of Moody's Investors Service credit ratings and credit rating processes. Information regarding certain affiliations that may exist between directors of MCO and rated entities, and between entities who hold credit ratings from Moody's Investors Service, Inc. and have also publicly reported to the SEC an ownership interest in MCO of more than 5%, is posted annually at www.moodys.com under the heading "Investor Relations — Corporate Governance — Charter Documents - Director and Shareholder Affiliation Policy."

Moody's SF Japan K.K., Moody's Local AR Agente de Calificación de Riesgo S.A., Moody's Local BR Agência de Classificação de Risco LTDA, Moody's Local MX S.A. de C.V, I.C.V., Moody's Local PE Clasificadora de Riesgo S.A., and Moody's Local PA Calificadora de Riesgo S.A. (collectively, the "Moody's Non-NRSRO CRAs") are all indirectly wholly-owned credit rating agency subsidiaries of MCO. None of the Moody's Non-NRSRO CRAs is a Nationally Recognized Statistical Rating Organization.

Additional terms for Australia only: Any publication into Australia of this document is pursuant to the Australian Financial Services License of MOODY'S affiliate, Moody's Investors Service Pty Limited ABN 61 003 399 657AFSL 336969 and/or Moody's Analytics Australia Pty Ltd ABN 94 105 136 972 AFSL 383569 (as applicable). This document is intended to be provided only to "wholesale clients" within the meaning of section 761G of the Corporations Act 2001. By continuing to access this document from within Australia, you represent to MOODY'S that you are, or are accessing the document as a representative of, a "wholesale client" and that neither you nor the entity you represent will directly or indirectly disseminate this document or its contents to "retail clients" within the meaning of section 761G of the Corporations Act 2001. MOODY'S credit rating is an opinion as to the creditworthiness of a debt obligation of the issuer, not on the equity securities of the issuer or any form of security that is available to retail investors.

Additional terms for India only: Moody's credit ratings, Assessments, other opinions and Materials are not intended to be and shall not be relied upon or used by any users located in India in relation to securities listed or proposed to be listed on Indian stock exchanges.

Additional terms with respect to Second Party Opinions (as defined in Moody's Investors Service Rating Symbols and Definitions): Please note that a Second Party Opinion ("SPO") is not a "credit rating". The issuance of SPOs is not a regulated activity in many jurisdictions, including Singapore. JAPAN: In Japan, development and provision of SPOs fall under the category of "Ancillary Businesses", not "Credit Rating Business", and are not subject to the regulations applicable to "Credit Rating Business" under the Financial Instruments and Exchange Act of Japan and its relevant regulation. PRC: Any SPO: (1) does not constitute a PRC Green Bond Assessment as defined under any relevant PRC laws or regulations; (2) cannot be included in any registration statement, offering circular, prospectus or any other documents submitted to the PRC regulatory authorities or otherwise used to satisfy any PRC regulatory disclosure requirement; and (3) cannot be used within the PRC for any regulatory purpose or for any other purpose which is not permitted under relevant PRC laws or regulations. For the purposes of this disclaimer, "PRC" refers to the mainland of the People's Republic of China, excluding Hong Kong, Macau and Taiwan.