

BACKGROUND INFO CTO-C

###OUTLOOK – TEMPLATE FOR THE OUTPUT [Fixed Knowledge for CTO-C]:



“OUTLOOK OF AI SYSTEMS CAPABILITIES

Six Months Outlook

With **Google’s ‘Gemini Ultra’** model release, we’ll see LLMs leap in capabilities, enhancing the foundational LLM/chatbot layer with sophisticated multimodal input handling: Gemini Ultra’s achieved a higher score than GPT4 in the MMMU benchmark, which assesses a range of capabilities. The main novelty of Gemini Ultra is to natively handle multiple types of data, including images, audio and live video.

This will improve data extraction from reports with better capabilities to read charts and tables, aiding in more comprehensive analysis while still requiring human oversight. It will also enable summarizing live events and TV reports.

Google has also recently unveiled **AlphaCode 2**, powered by the Gemini AI model. AlphaCode 2 demonstrates notable advancements in coding capabilities and is particularly adept at handling complex mathematical and theoretical computer science challenges. It employs dynamic programming, a method of breaking down complex problems into simpler sub-problems, and utilizes a combination of policy models and a scoring model to generate and select the best code solutions.

These new models will drive progress in valuation and risk modeling by LLM on their own. They are also expected to increase the capabilities of multi-agent systems including live data feed and ML models.

However, in 6 months, most complex system using agents will remain at experimental stage, limiting the ability to automate an entire workflow. Besides, the nuanced understanding of market dynamics and design of disruptive scenarios will likely remain within the human analyst's domain.

2 Years Outlook

LLMs. Within 2 years however, more sweeping architectural innovations combining enlarged model sizes exceeding 5 trillion parameters (GPT5, Claude 3), novel training techniques, and high-quality curated datasets could plausibly result in LLMs exceeding average individual human cognitive capabilities across nearly all present professional benchmark tests, including finance.

Fine-tuning. Current frontier models such as GPT4 are expected to be made available for fine-tuning and open-source models which can be trained for specific purposes are expected to reach the capabilities of current frontier LLMs. This is expected to drive down the cost of using LLMs for large scale analysis.

Specifically in financial analysis, this could enable reliable accuracy in most real-world reporting, valuation, forecasting, investment decision-making, risk assessment, and compliance contexts.

Agents and multi-agent systems. Even more critically, the progress in the integration of agents and the deployment of multi-agent systems will enable the automation of complex multi-step processes and the development of 24/7 systems monitoring written and video news, processing real time data and updating models in real time.

But outlier cases needing advanced creative skills or versatile human judgment seem likely to remain the domain of human analysts.

5 Years Outlook

In 5 years, LLMs demonstrating cognitive capabilities comparable to median human financial specialists across sectors like accounting, investments, insurance, banking, and advising appear achievable. Even relatively niche corners like tax optimization, complex securities structuring, and esoteric global regulations could see automation encroaching on human primacy.

Machine advantages in computational speed, information accessibility, and disciplined analytical consistency may surpass individual financial experts across most domains enable the augmentation of most tasks. Current trajectory projections suggest LLMs could excel at certain financially-relevant capabilities like persuasive communication, deception detection, and predictive forecasting beyond levels individually attainable by even top professionals today."

Progress in AI video generation and AI connection with robotics and VR will likely enable 'autonomous agents' to interact with humans in virtual or real environment, closing the gap between reality and science fiction.

Multi-agent systems will incorporate very different skills, enabling the replacement of entire multi-disciplinary teams for standard workflows.

But outlier gaps around the design of new workflows and systems, creative risk-taking, intuiting complex interconnections, establishing emotional bonds to do business, and applying versatile judgment seem likely to keep the most elite human practitioners securely valued."

##END OF THE OUTLOOK

###ADDITIONAL BACKGROUND KNOWLEDGE [Fixed Knowledge for CTO-C]

1) Analytical Framework

The rise of GenAI has given a central role to LLMs powering chatbots. The ability of a chatbot can increase dramatically with the addition of software and robotic layers. To standardize the analysis our framework includes the following layers, each layer enhancing the abilities achieved by the system:

- 1) Layer 1: **Chatbots**. The framework we use for assessing the abilities of AI systems assumes that frontier Large Language Models (LLM) such as GPT4 or Claude constitute the fundamental building block. We therefore start by assessing the abilities of such models as a standalone tool provided via a chatbot such as ChatGPT.
 - a. Example: The LLM can analyze the text in financial statements
- 2) Layer 2: **Tools**. Tools that can be connected to LLMs to automate tasks such as database query, calculation, image analysis and generation, voice-to-text and text-to-voice, etc. In theory, LLMs can be connected to pretty much all the software tools that exist thus greatly enhancing the abilities of the model. In practice however, not all the related extensions or ‘agents’ have been developed to date.
 - a. Example: Tools are necessary to extract data from tables, charts, develop an NPV model and output charts
- 3) Layer 3: **Abilities of a Fine-Tuned LLM**. To better perform on specific tasks and topics an LLM can be fine-tuned on a specific dataset, or even re-trained for open-source models. However, as of today, only less powerful models such as GPT3.5 turbo or open-source models like Mistral can be fine-tuned or trained by users.
 - a. Example: An LLM is fine-tuned to understand complex tables corresponding to a certain format (e.g. total sales on the top line and sales per business line under it).
- 4) Layer 4: **Multi-Agent Systems**. MAS involves several bots collaborating with each other and using various tools to execute multi-step processes. Although there is no example of large-scale commercial application yet, research has demonstrated that the use of these MAS dramatically increase the performance of LLMs.
 - a. Example: The MAS automates the workflow from financial statement download to output of the analysis
- 5) Layer 5: **VR and Robotic Extensions**. AI systems are limited in their ability to interact seamlessly with their physical environment and notably humans. This limitation can be partly addressed with robotic extensions (sensors, actuators) and/or the connection with Virtual Reality environments. Robotic tools such as sensors, actuators, effectors, etc. that enable an interaction of the AI system with the physical environment. Virtual, Augmented and Mixed Reality reducing the need for interactions between the AI system and the physical environment.
 - a. Example: Robotics is used to control the interface with a laptop and bypass all restrictions on websites access by bots
 - b. Example: An AI clone of the financial analyst deliver the presentation in VR.

In the examples above, the task of analyzing financial statements face hurdles when only using the chatbot. However, with agents the system can [read charts](#) and [generate models](#) to process the data. Combining [multiple agents](#) working with each other mimics the skillset of a team with different domains of expertise and tools. Add Virtual Reality and the [avatar of a human analyst](#) can present and discuss the results to a client in a [language](#) he doesn't speak.