BACKGROUND INFO CTO-A

###TEMPLATE [Fixed Knowledge for CTO-A]

SCOPE This summary focuses on the current abilities and limitations of frontier LLM models used in chatbots (excluding additional tools and systems), in the context of financial analysis related activities.

INTRODUCTION Large language models (LLMs) like GPT-4 and Claude 2.1 represent revolutionary AI systems, trained on vast data, power chatbots to converse via text with humans. Frontier LLMs like GPT-4 and Claude 2.1 demonstrate advanced reasoning and language capabilities, but still have limitations compared to human cognition and abilities.

RESULTS AT EXAMS In specialized exams, frontier LLMs reveal remarkable language understanding and reasoning that meets and sometimes exceeds human performance. GPT-4 scored 90% on the Uniform Bar Exam, while Claude ranked in the 93rd percentile on the Graduate Record Examination test (Analytical Writing, Verbal Reasoning, Quantitative Reasoning). However, GPT-4 failed the CFA exam so far: it showed potential in certain areas but struggled with complex finance topics, especially in the Level II exam. Its performance was close to the estimated passing threshold under certain conditions but was not consistently above the passing mark.

FUNDAMENTAL LIMITATIONS However, frontier LLMs used via chatbots demonstrate clear constraints versus humans: • No Real-Time Data: They cannot process or analyze live data, limiting usefulness in fast-changing finance settings. • No Quantitative Analysis: They lack the mathematical capabilities for statistical analysis expected in finance. • Context Size Constraints: even with 'large' context widows (128k tokens for GPT-4, 200k tokens for Claude 2.1), performance drops when context exceeds 70k tokens (about 100 pages or 50k words), risking losing crucial details. • Factual Inaccuracies: Reality gaps mean they generate plausible but incorrect information much more frequently than humans. • Difficulty to follow complex multi-steps directions compared to humans. • Overconfidence: They attempt to answer questions beyond their competencies, rather than admitting ignorance. • Reading limitations: Interpreting complex tables and charts continues to challenge them. • While chatbots can produce text, they cannot execute actions (e.g. trading, etc.) without additional tools.

ABILITIES IN PRACTICE For finance-related use cases, they show strengths in: • Reading text (from reports, news, research etc) almost immediately, in multiple languages, and a low cost ($0.1 to $1 per 150 pages). • Analysis: Assessing sentiment, pros & cons, applying scoring systems, extracting information, calculating basic ratios, comparing, summarizing key takeaways, etc. • Generating text: Processing volumes of text in multiple languages quickly and with no spelling mistakes, at low cost ($0.2 to $3 per 150 pages).

CONCLUSION These abilities can be mobilized to augment (speed, scope) or replace tasks in several workflows related to financial analysis, such as consuming research to spotlight connections or drafting summaries, analyzing large volume of documents to identify relevant insights, generating first draft of reports, supporting research, etc."