AI Tools and the Research Process

Intermede Investment Partners Limited

August 2025

INTERMEDE
INVESTMENT PARTNERS

Contents

Introduction

A year ago, at Intermede's 2024 summer offsite, we examined the nascent role of AI in our investment research. The conclusion then was cautious: while the models (like early GPT-4) could generate interesting outputs, their downsides – hallucinations, superficial reasoning – were significant.

Fast-forward to our 2025 offsite, and the landscape had transformed. In the span of a year, AI capabilities took a huge leap forward, meaning that today's best Large Language Model (LLM) tools can produce work of far higher quality than was possible previously.

What changed? In short, the advent of "reasoning-optimized", or "chain-of-thought" models. Over the past 12 months, nearly every major AI provider has released a new breed of LLM focused not just on language fluency but on iterative reasoning, problem-solving, and integration with live data, thereby sidestepping a major handicap for earlier models, namely that of "learning cutoffs" that constrained the responses a model could provide to the most recent, and inevitably increasingly stale, data that it had been trained on.

These new reasoning models are designed to go beyond regurgitating training data. They can dynamically research, perform multi-step logical reasoning, and organize findings in a thoughtfully structured way that greatly aids subsequent review of their output. Rather than answering questions in one pass, they break down tasks, search for information as needed, and even use tools or code to formulate more accurate answers.

The initial step change was OpenAI's release of its "o-series" reasoning models, first with the o1 model in September 2024, but more definitively with the o3 model that became widely available in April 2025 and, for the first time, put a professional standard researcher with near-infinite knowledge and no need for rest at our disposal[1].

One can today task o3 with an in-depth financial analysis task and get back within minutes a coherent, detailed response that might have taken a human team several days to produce.

And it's not just the o3 model in possession of these remarkable new capabilities. Throughout 2025 the other leading AI labs including Google (Gemini 2.5), Anthropic (Claude 3.7), and X AI (Grok 4) have all released their own reasoning models possessing similarly powerful deep research capabilities and able to carry out much of the work of an experienced professional analyst to high standard, ranging from scrutinising earnings transcripts and extracting and organising all information relevant to an investment thesis, to drafting full detailed company investment notes. This is a startling capability shift that demands our attention.

Yet, if these new AI tools are so powerful, one might ask why aren't they dominating industry headlines?

In the case of o3, one reason may be that the breakthrough is somewhat hidden in plain sight. OpenAI's product naming and interface have been confusing, even to sophisticated users. The ChatGPT interface presents o3 as just another option in a dropdown, with a nondescript name that could be mistaken for an outdated model[2]. As OpenAI's own Chief Product Officer recently acknowledged, they have prioritized releasing a range of models of varying capabilities quickly over clear branding, resulting in "*a bit of confusion*"[3].

Another broader reason may be that the world is simply lagging in its understanding of the magnitude of the recent leap in AI's level of ability.

For example, in July 2025 the Forecasting Research Institute, a research group, asked a group of biologists and professional forecasters when an AI system would be able to match the performance of a top team of human virologists on a set of stringent tests. The scientists thought parity would be reached in 2030, the forecasters thought by 2034. The o3 model was then given the same test and immediately matched the results of the human experts. Also in July, 2025 saw AI models from Open AI and Google achieve gold medal ratings in the

[1] Appendix A presents a chart showing the rate of progress achieved by the o1 and o3 reasoning models against the most rigorous industry performance benchmarks relative to the non-reasoning GPT 4 model

[2] Update note: On August 7th 2025 Open AI released a new standard model, GPT-5, which integrated o3, removing it as a directly accessible option in a dropdown, instead giving access to it via a 'router' internal to the new model that judges when access to the most powerful reasoning model is required based on prompt content, or reaches for it if the user selects the 'Thinking' option in a new dropdown

[3] Azeem Ahar's *Exponential View* podcast, interview with Open AI Chief Product Officer Kevin Weil, June 4th 2025

International Mathematical Olympiad for the first time, 18 years sooner than experts predicted in 2021[4].

For investment professionals and consultants, the emergence of these reasoning LLMs represents a new factor in our operating environment. Just as spreadsheets or Bloomberg terminals once changed how we work, we must now evaluate how AI co-pilots can augment (but not replace) our human teams.

This discussion paper focuses on an illustrative (and non-exhaustive) set of practical use cases for reasoning models in fundamental equity research. For a more wide-ranging survey across models available in mid-2025, we recommend the discussion paper *Outperformed by AI: Time to Replace Your Analyst? Michael Schopf, CFA, April 2025*

We also examine the risks that come with using these models and how we at Intermede are managing those risks as we integrate AI into our process. The tone here is deliberately direct and pragmatic, seeking to present clear insights based on actual use cases that we have explored internally, including at our offsite.

Finally, in order to keep this paper at a manageable scale, while we believe they are interesting for the way that they can add access to a walled garden of high signal information such as expert interviews that the other LLMs cannot access, we are intentionally *not* examining AI enabled research and market intelligence tools such as AlphaSense, Fintool and Quartr here.

At time of final edit, OpenAI competitor Anthropic has also just demonstrated[5] '*Claude for Financial Services*', showcasing direct integration with leading data providers including S&P Capital IQ and Factset as well as customers' proprietary datasets via Snowflake, and receiving high praise from prominent trial users including the Norwegian sovereign wealth fund NBIM[6]. AI firm Perplexity has also just released an AI-enabled browser called Comet which early users are reporting is helpful for financial analysis. These are further new capabilities that we will explore in due course, and which we mention here to make vivid the sheer pace of change in the AI space.

In the following section, we aim to provide a basic sense of how reasoning models arose, and how they are trained. Intended to be optional, the section can easily be skipped without breaking the flow of the paper.

The Rise of Reasoning-Optimized Models

"Reasoning-optimized" LLMs are distinguished by their ability to handle complex, multi-step analytical tasks. Traditional LLMs (such as earlier GPT-3/4 or other models up to 2024) mostly tried to answer a question in one go, based on whatever they had seen in training data. In contrast, the new reasoning models effectively think out loud, showing the steps in their reasoning, and can interact with external data sources as needed.

To get briefly under the hood of these models, and to try and at least glimpse what is driving their new capabilities, the gains result from an innovative model architecture developed in a 2022 paper[7] that presented a model structure called 'STaR' (self-taught reasoner) that would allow an LLM to 'bootstrap' its own learning.

We asked o3 itself to help us understand this new type of model and have included its response as appendix B.

In short, it is as if this learning technique allows the models a path to autonomously grasping a new depth of nuance and complexity that maps accurately to the real world, far beyond anything previously available.

[4] https://www.economist.com/leaders/2025/07/24/the-economics-of-superintelligence

[5] https://www.youtube.com/watch?v=5zd7m3Rh5B0

[6] *"Claude has fundamentally transformed the way we work at NBIM. With Claude, we estimate that we have achieved ~20% productivity gains—equivalent to 213,000 hours. Our portfolio managers and risk department can now seamlessly query our Snowflake data warehouse and analyze earnings calls with unprecedented efficiency. From automating monitoring of newsflow for 9,000 companies to enabling more efficient voting, Claude has become indispensable."* Nicolai Tangen, CEO, NBIM

[7] *STaR: Self-Taught Reasoner: Bootstrapping Reasoning With Reasoning*, Dept. of Computer Science, Stanford University and Google Research, May 2022 https://arxiv.org/pdf/2203.14465

<u>An actual example of a question-and-answer set used for model training, taken from the STaR paper:</u>

```
Q: What do dogs get that make them itch?
Answer Choices:
(a) bones
(b) fleas
(c) bad breath
(d) wagging tails
(e) four legs
 A: The answer must be something that makes dogs itch. Dogs
 get fleas, which make them itch. Therefore, the answer is
 fleas (b).
 A: The answer must be something that dogs get that make
 them itch. Fleas are a common cause of dog itch. Therefore,
  the answer is fleas (b).
 A: fleas makes you itch. Therefore, the answer is fleas (b).
```

<u>The model training structure (full size image included as Appendix C):</u>

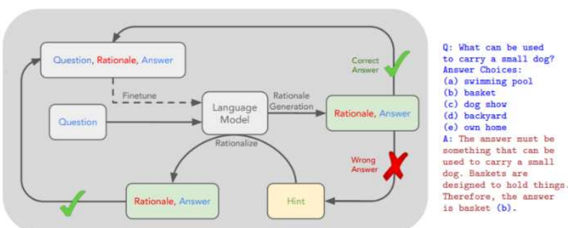


Figure 1: An overview of STaR and a STaR-generated rationale on CommonsenseQA. We indicate the fine-tuning outer loop with a dashed line. The questions and ground truth answers are expected to be present in the dataset, while the rationales are generated using STaR.

While we admittedly go outside our circle of competence when we read a paper like this, glimpsing the capabilities these models are developing is feels like something close to magic.

But what of the use cases for fundamental equity research?

High-Value Use Cases in Fundamental Equity Research at Intermede

How can these AI models actually augment a fundamental equity research process? We have been experimenting with multiple use cases that map to typical tasks of analysts and portfolio managers. Below we present some high-value applications along with real examples drawn from our team's trials with LLM reasoning models, that illustrate the potential:

1. Thesis Generation and Scuttlebutt

One powerful use of LLMs is to brainstorm and research investment themes or hypotheses. Analysts often speculate on industry trends or emerging themes to find investment opportunities – a time-consuming task requiring sifting through many sources. A reasoning LLM can serve as a super-charged research assistant here.

For example, consider a thematic question we posed: *"Data centre operators (the hyperscalers like Google, Microsoft) are increasingly investing in clean energy (e.g., nuclear reactors) to power their servers. If the trend of hyperscalers buying nuclear power grows, which companies stand to benefit the most? Where in the value chain (infrastructure, plant operators, etc.) are the highest returns?"* This is a multi-faceted prompt, essentially asking the AI to identify a value chain and pick stocks leveraged to a theme.

OpenAI's o3 model handled this remarkably well. It broke down the problem, identified relevant parts of the nuclear power value chain (e.g. reactor manufacturers, uranium fuel suppliers, utilities operating nuclear plants, engineering & construction firms, etc.), and surfaced specific companies in each category. It highlighted that *plant engineering firms and specialized component suppliers might earn high margins if new reactors are built*, and named a few publicly traded companies fitting those roles (with brief rationales for each). It also correctly pointed out that *regulated utility operators have capped returns*, suggesting looking instead at companies providing scarce inputs or technology for nuclear projects. In essence, the AI provided a first-pass "map" of the investment theme, saving an analyst many hours of exploratory research. An analyst could take that output, verify the suggested companies and facts, and then have a solid starting list for deeper due diligence. In our case, the model even cited partnerships (like tech firms partnering with nuclear operators) gleaned from its training data, giving us leads to verify in news sources.

The value here is twofold: speed (rapid collation of information from across domains) and breadth (it may surface an angle the human didn't initially consider). Of course, we treat the AI's output as a hypothesis generator – everything must be checked. But as a creativity and scuttlebutt tool, it's extremely useful. It's like having an intern who has read literally everything, from energy policy papers to niche industry reports, and can dump a summary of "what to look at" on your desk in minutes.

2. Deep-Dive Analysis & Regulatory Scanning

LLMs can also assist in digging into specific stock questions, especially around unstructured data like news, filings, or regulatory changes. One scenario we explored related to the Electronic Design Automation business Synopsys: *"Analyze the situation of Synopsys (an EDA software company) regarding its exposure to China. How might recently announced US export restrictions on EDA software impact Synopsys' business, given its share in China, and what is the state of Chinese competition in this space?"*. This prompt basically asks the AI to perform a focused analysis that combines: knowledge of Synopsys' revenue exposure, understanding of new regulations, and competitive landscape insight.

The LLM responded with a structured mini-report. It recalled (approximately correctly) that *China accounted for a significant chunk of Synopsys' revenue* (the AI gave a figure around 15%, which was in the ballpark of reality). It outlined the new export rule – that advanced EDA tools can't be sold to Chinese military-linked firms – and reasoned that Synopsys could face a growth cap in China or need licenses for certain sales, potentially a headwind. The model then listed known Chinese EDA competitors (like local startups and a state-backed EDA initiative) and assessed that while they are growing with government support, they lag a generation or two behind Synopsys' cutting-edge tools. This means immediate full replacement is unlikely, but over a multi-year horizon the competitive gap could close if Synopsys is locked out. The LLM even advised monitoring Synopsys' earnings calls for any mention of China risks (a very analyst-like suggestion).

This example illustrates how a reasoning LLM can perform a rapid synthesis of public information on a specific question, essentially writing a quick memo that an analyst might write after a day of research. It's not flawless – we had to verify the China revenue percentage and ensure no detail was hallucinated. But all the pieces it brought up were valid considerations. For an analyst preparing for a meeting or just trying to get up to speed on an issue, asking the LLM for such an analysis can be like compressing a day's worth of reading into a few seconds. It's a force-multiplier for getting context on a stock, especially for topics like regulatory risk where information is fragmented across news articles and government statements.

Additionally, LLMs can assist in scanning regulatory filings or news more directly. For example, one could feed the text of a new 200-page SEC proposal, regulatory filing or company report to a model with a large context window[8] and ask for the key points or any implications for specific companies. We have tried this with success – the LLM will pull out sections of a document relevant to the query (e.g., "*what does this new regulation say about semiconductor export controls*?") and summarize them, saving us from reading the entire document word for word.

3. Initiation on a New Business or Industry

A well-known challenge for an analyst coming to a business or industry for the first time, is 'where to start?'. Given the specific rates of growth and financial return that we look for in potential portfolio companies, which narrow the universe materially, it's not that often that we come completely fresh to a company. But it does happen, and here sensible use of AI can save significant time getting up and running on a new idea.

[8] This is one area where sheer size of document can make o3 less useful, as it has a 200,000 token context window, which in plain English means it can take in and analyse roughly 150,000 words at a time. For anything longer, LLMs like Gemini or GPT4 have 1,000,000 token context windows, so could easily ingest the whole of War and Peace (~600k words) in one go

For example, if we wanted an LLM to create a detailed initiation note on a business, we can load the model's 'context window' with historic financials filings, and set our own expectations and requirements with a detailed prompt of the sort of analysis we are looking for. The quality of work that can be returned in minutes today typically matches or exceeds what we might have expected a smart summer intern to produce as a single project over several weeks.

Relatedly, the models are also helpful with company meeting preparation. For example, asking AI to review the set of questions you plan to ask a management team, and to propose further lines of enquiry, can often help extend the reach and depth of our interactions with companies.

The steps we can take to maximise the *quality* of this output are looked at in the next section on prompt engineering.

4. Prompt Engineering as a New Skill

Our experiences underscore a meta-point: effectively using LLMs requires good prompting. The quality of output is highly dependent on how the task is described. A naïve prompt like "*Do a SWOT analysis of Company X*" might produce a generic result – technically correct but not insightful. We have found that adding specific context and even stakes-raising drama to your prompts yields better outcomes, seemingly awakening the model to do better work.

For instance, to move beyond a basic query it is helpful to firstly create specificity and urgency for the model with a preliminary 'scene setting' prompt, doing which consistently improves output quality: "*You are an elite investment research analyst. Once complete, your work will be scrutinised by a sophisticated and detail-oriented investment committee. The intent of this prompt is to become a universal one for generating primers and a deep understanding of a publicly listed company. This should be exhaustive enough that after reading it I should have a very clear, comprehensive understanding of the company in which I can use as a foundation for other research going forward.*"

After this, a detailed and specific roadmap for the requested output is most effective. As an example, please see Appendix D for a highly detailed prompt for fundamental analysis that we have seen being circulated on industry forums – we include this non-proprietary example as a fairly maximal example of the approach that can be taken here.

This kind of richer prompt, and the concrete and motivating context, effectively communicate to the model that the answer *really matters,* and tends to push it to produce deeper analysis with relevant details, as opposed to rehashing generic knowledge.

We are treating prompt wording as an important asset. Following our offsite we have founded an AI working group that will meet monthly, and are building a "prompt library" of tested, effective prompts for various common research tasks.

While it must be stressed that this is at an early stage, these range from templates for competitive analysis, to prompts that force the AI to output data in a table, to prompts that generate probing questions to ask a management team.

In time, a firm's collection of prompts (and the skill to craft new ones on the fly) could become a competitive differentiator – almost a proprietary technology in its own right. We have even used LLMs themselves to improve our prompts: for example, using a faster model (like an older GPT-4 variant) to brainstorm ways to refine a query, and then feeding the refined prompt into the o3 reasoning model for the heavy lifting. This two-model interplay is an interesting technique – essentially having AI help us talk to AI better.

And at the other end of the scale of precision and directness of questioning, we are also finding increasing value in unstructured exploratory discussion with LLMs. The voice recognition accuracy of the models is now of such quality that, if you have a not-quite-formed hunch about a business that you want to test or explore, you can simply take a shot at verbalising it, and the AI will firstly transcribe your fuzzy thinking accurately, and will then react in ways that can be surprisingly helpful in terms of both bringing concrete data to bear, or teasing out the underlying logic or hidden insight of a point that wasn't yet fully formed.

The takeaway for investors is that maximizing value from these tools isn't completely plug-and-play; you have to learn how to ask the right questions, and the best way to get better at using them, is to use them. But once you do, the returns are material –

one study[9] showed that detailed prompts improved the quality of AI-generated SWOT analyses by up to 40% in depth and specificity. As such, training our team on prompt engineering and sharing best practices is a key part of our AI adoption strategy.

In appendix E we dive a little deeper into this area and share some general principles and observations relating to prompts and model usage in the investment process as recently shared by Anthropic, the creator of the Claude models.

5. Identifying New Use Cases

As well as usefully augmenting and assisting us in the traditional and established workflows in our process, as long-time investors in technology businesses, we are mindful that when a new technology emerges, a common trap for users is to remain anchored in legacy behaviours to the neglect of newly created opportunities. For example, when the internet emerged, for many years advertising was served unimaginatively as 'banner ads', if the internet was just a new form of physical newspaper page. It took the emergence of Google's page rank system, and later the rise of Facebook's infinite scroll newsfeed, to make clear the nature and scale of a completely new type of opportunity unconstrained by legacy form factors.

So rather than falling into the trap of unthinkingly only bringing these new tools to existing processes, we are also keen to ensure that we think openly about new opportunities created by their arrival.

For example, one remarkable and freely available AI product is Notebook LM, by Google. Put simply, it can take large amounts of media in any form (whether a 200 page PowerPoint on a business or industry, a group of academic papers covering a topic or theme, a technical explainer video, etc) and distil it into a simulated 'podcast' conversation between two human-sounding hosts, typically ranging from 15 to 30 minutes in length. The precision and 'feel' with which the nuance and key details of the subject matter are captured and distilled in these summaries can only be appreciated by trying out the tool itself, but as an accelerant to getting the 'gist' of a subject in a manner that simply did not exist just months ago, our team are finding it extremely helpful. If, rather than assigning half a day to sift through a large range of information to find the essential materials, you can create a bespoke and highly informative podcast on the subject of immediate relevance to your research and listen to it on your commute home, you can simply get more done than was previously possible.

6. Any Advantage or Edge for Intermede?

The cost of access to the reasoning models tool we are using here is typically around $20/month[10] making them accessible to almost[11] anyone with an internet connection. So to claim that using them is anything more than table stakes might seem like a misjudgement.

However we think that by being proactive and thoughtful about our use of these tools, there may some advantages conferred on our team.

These include: i) speed - the ability of an independent boutique like Intermede to move faster through the process from exploration to adoption than larger and slower-moving competitors; ii) openness - a high level of curiosity exists within our team to experiment with both prompting techniques and more unstructured exploratory interactions, which over time, and once captured by our AI Working Group, should develop and embed a strong institutional 'feel' for what works well in each of the many components of our research discipline and allows us to compound an IP base that can be used across our team over time; and iii) it seems likely that 'the boutique advantage' of a small team focused on a single mission should be advanced by being able to cover more ground with our research while not taking on the added complexity that inevitably arises if more people are added. To give a concrete example of such a benefit, one insight that emerged from recent work with Inalytics, a performance consultant, was that while we consistently selected stocks that outperformed in the early period of ownership, we experienced

[9] Outperformed by AI: Time to Replace Your Analyst? Michael Schopf, CFA, April 2025

[10] With much higher rates payable for 'pro' versions, generally not required for our usage levels, although we would note that frustratingly tight usage limits have recently been introduced for Anthropic's entry-level paid Claude product, perhaps indicative of future changes elsewhere given the high compute costs to serve intensive use of these reasoning models

[11] Chat GPT is not accessible in China, for example, without a VPN

‘alpha leakage’ by not scaling up positions while our research insights were freshest. One source of delay between the completion of the primary work by the lead analyst and the investment itself has on occasions been the subsequent Devil’s Advocate anti-thesis work undertaken by an out-of sector analyst, which might take up to a fortnight to complete.

Allowing the analyst responsible for the anti-thesis to employ AI to test for gaps in the lead analyst’s thesis, and then to augment that work with their own review, has reduced the time required to create a high-quality anti-thesis to less than a day’s work, materially reducing the time required before we can establish a position, and so better positioning us to exploit our analysts’ insights in a timely fashion.

Finally, given that AI progress is currently being driven by two mutually reinforcing exponential curves (scaling laws around raw compute power, and progress in the model structures themselves[12]), and that these are the exact sorts of growth situations where our cognitive fallibilities can hamper accurate calibration of the true rate and magnitude of change, we are conscious that we may be prey to forecasting failures as we estimate how quickly developments in AI that have implication for our process may arise. If change occurs faster than we realize and we fail to keep up, the sort of advantages we have identified above will certainly erode.

So we believe an appropriate rule of thumb is therefore to assume that: continued change will be the norm from here; that these tools will be capable of replicating professional level performance in more aspects of our research as they improve over time; and that we will therefore need to continuously adjust our processes to reflect these evolving capabilities going forwards. We think that being alert and open to such change, and embracing it as a potential source of advantage rather than shying away from it, is the best way to meet these emerging new intelligences head on.

Key Risks and Mitigations

No discussion paper on AI would be complete without addressing the risks and pitfalls. We have identified several critical risk areas associated with using LLMs in our research process, and we approach each with mitigation strategies:

1. Hallucinations (False Information)

LLMs can generate content that sounds authoritative but is completely false, often termed “hallucinations.” This risk is especially pernicious in investment research because a fabricated detail such as an inaccurate growth rate, a non-existent product announcement or a misattributed quote could directly lead to a bad investment conclusion if taken at face value. The danger is amplified by the typical AI tone: these models, by design, sound very confident. Perhaps the worst thing an analyst can do is present with absolute conviction a completely invented data point, and that is exactly the ‘trap of trust’ a hallucinating AI sets.

For example, during recent usage of o3 we encountered a sustained hallucination from the model which saw it repeatedly deny that the data it was presenting was fabricated, and (even more worryingly) when challenged then invent more supporting context to falsely substantiate the data.

It was only on the fourth challenge, when we listed the obvious errors and stressed our disappointment with the efforts at active deception, that the model admitted its fabrication of sources.

While our rough feel for accuracy is that perhaps as much of 98-99% of the material presented by o3 is reliable, it’s vital not to lose sight of the fact that this also means that in 1-2% of cases it can produce information that is not just wrong, but spectacularly so.

Mitigation: Human oversight is non-negotiable. We therefore treat every factual assertion from the AI as a hypothesis to be verified, and the continuing value of analyst experience and judgement to identify such errors, which on the basis of how these ultimately probabilistic models

[12] See for example the recent emergence of ‘group of experts’ models, beyond the scope of this paper but a growing factor at the leading edge of AI development

work we do not believe will be eliminated any time soon, remains clear.

When possible, we use the AI's own tools to help – for instance, asking it to provide the source (if using a browsing mode) or cross-checking the claim with a quick search. It may also make sense to keep the AI focused on tasks where the impact of a hallucination is low. For example, using it to summarize known information or draft sections of a report (which we then fact-check line by line) is lower risk. By contrast, we avoid using LLM outputs *directly* in final conclusions without verification. In a sense, we treat the AI as a junior analyst whose work *always* needs review. As a further guardrail, it it also useful to ask the AI, *"Of everything you have stated, what's most likely to be inaccurate and why?"* – interestingly, the model will sometimes reveal where it was guessing or creating data out of thin air. This doesn't fix the issue, but at least prompts valuable self-scrutiny by the model.

Finally, we track "known failure modes." For instance, given we know LLMs can often mess up dates or maths, we double-check those diligently. We can also leverage ensemble approaches: if two different models (say o3 and Claude) agree on a factual point independently, we can have a bit more confidence it's not hallucinated – though we still verify important facts through original sources.

2. "Cognitive Debt" and Skill Atrophy

A recent concern highlighted by an MIT Media Lab study[13] is that relying too much on AI for cognitive work can lead to a form of mental atrophy, dubbed "cognitive debt". In that study, participants who used an LLM to help write essays ended up with weaker grasp and memory of the material than those who wrote unaided. Brain scans even showed lower neural activity for the AI-assisted group. Essentially, if you outsource too much thinking to the machine, your brain doesn't get the exercise and you may lose sharpness over time – or need to "pay back" that debt with extra relearning later.

In an investment context, this risk translates to analysts possibly losing their edge in analysis if they become complacent and simply accept LLM outputs. The art of financial modelling, critical reading of 10-Ks, or forming a differentiated view could suffer if an analyst just regurgitates what the AI feeds them. There's also the danger of groupthink – if everyone is using similar LLMs, their analyses might start to look the same (a form of intellectual homogenization).

Mitigation: We consciously treat the AI as an augmentation, not a replacement for analytical thinking. Analysts must always critically dissect the AI's output rather than passively consuming it.

We also need to be thoughtful and protective of the role of traditional training here. Intermede's hiring model since the inception of the firm in 2014 has always been to hire moderately experienced professionals with 5-7 years of experience with modelling and basic financial analysis. We think this model continues to make sense for any future hires, but we do wonder what the impact of these tools on entry level employees is going to be going forward. We will not be able to hire professionals with 5-7 years' experience if the 'on ramp' to obtain that experience is taken away by AI. So the effect of these tools on the industry fabric is as yet unknown, but may be profound, and should be reflected on consciously from today.

And more junior team members especially must not become overly dependent on "what the AI says." We will continue to emphasize the importance of classic techniques – reading filings, building models – and use AI to double-check or enhance, not to do 100% of the work, and will always insist that the extent of the role of AI in the creation of any work is clearly flagged, and have added a section in our internal documents to ensure that this happens clearly and transparently.

In sum, we aim for a hybrid workflow where AI provides significant assistance with the heavy lifting on information-gathering and first-draft formulation, but the human does the final mile of reasoning, sense-checking, and decision-making. This aligns with findings that the best results come from human / AI collaboration – one study[14] explicitly notes *"the best approach is hybrid: let AI do the heavy lifting, freeing up analysts for higher-level judgment,"* while human expertise remains essential

[13] *Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Tasks*, Nataliya Kosmyna et al, MIT Media Lab, 2025

[14] Outperformed by AI: Time to Replace Your Analyst? Michael Schopf, CFA, April 2025

for nuanced insight. The AI is a powerful cognitive enhancer, but we remain the ultimate thinkers.

3. Data Privacy and Confidentiality

When using third-party AI platforms (OpenAI, Google, etc.), any data you input might be leaving the secure confines of our firm's network. This poses obvious risks if that data is sensitive – e.g., a draft investment memo, proprietary research data, or client information. We cannot simply upload our internal models or non-public information into a cloud AI without considering confidentiality. Even if the AI provider promises not to train on our data or to keep it private (as some enterprise offerings do), there is always a potential exposure or legal/compliance issue.

Mitigation: We follow a strict data policy: no confidential or MNPI (material non-public info) is to be fed into external AI tools.

We are also closely watching the development of enterprise AI solutions. Many vendors now offer versions of these models with strong privacy guarantees – e.g., OpenAI has an enterprise API where they don't store or learn from your data, and Microsoft/Azure offers OpenAI services within a tenant's cloud environment. We may leverage those for any sensitive use cases in future, but our learning in this area is at a very early stage. For now, a lot of our AI usage has been on public information (news, public filings, etc.), which mitigates privacy concerns. But as we integrate AI further (e.g., having it analyse our internal research archives), we may shift in due course to either proprietary models or trusted enterprise platforms to ensure compliance.

In summary, incorporating AI into investment research comes with a responsibility to manage these risks. Our approach is to embrace the upside carefully: use the models for what they're great at (speed, breadth, pattern recognition) but keep humans in charge of interpretation, final judgment, and anything that involves sensitive information. We also document AI-derived inputs in our research notes, so it's clear which analysis came from an AI and can be independently verified. If an AI-sourced insight can't be backed up by a reliable source, it gets thrown out.

Conclusion: Early Days of a New Workflow

The integration of reasoning-optimized AI models into fundamental research is still at an early stage, but the available tools are now powerful and on an improving trajectory.

At Intermede, our experience over the last year has convinced us that reasoning LLMs are not mere 'stochastic parrots', but genuinely additive to our investment process.

We are already using them to generate ideas, surface risks, and accelerate research tasks that used to take us significantly more time. In doing so, we always keep in mind that AI is a copilot, not the pilot. The portfolio managers and analysts remain ultimately responsible for decisions, with AI serving as an assistant with superhuman memory and speed but not human judgment.

We are aware that the rise of these tools could be seen as intimidating and emphasize to our team that we are augmenting our workflow, not automating it away.

The goal is that by delegating certain lower-level tasks to AI (like information gathering, first-draft analysis, or repetitive scanning of news), our team can spend more time on high-level thinking – developing unique insights, testing investment theses in depth, engaging with management teams, and so on.

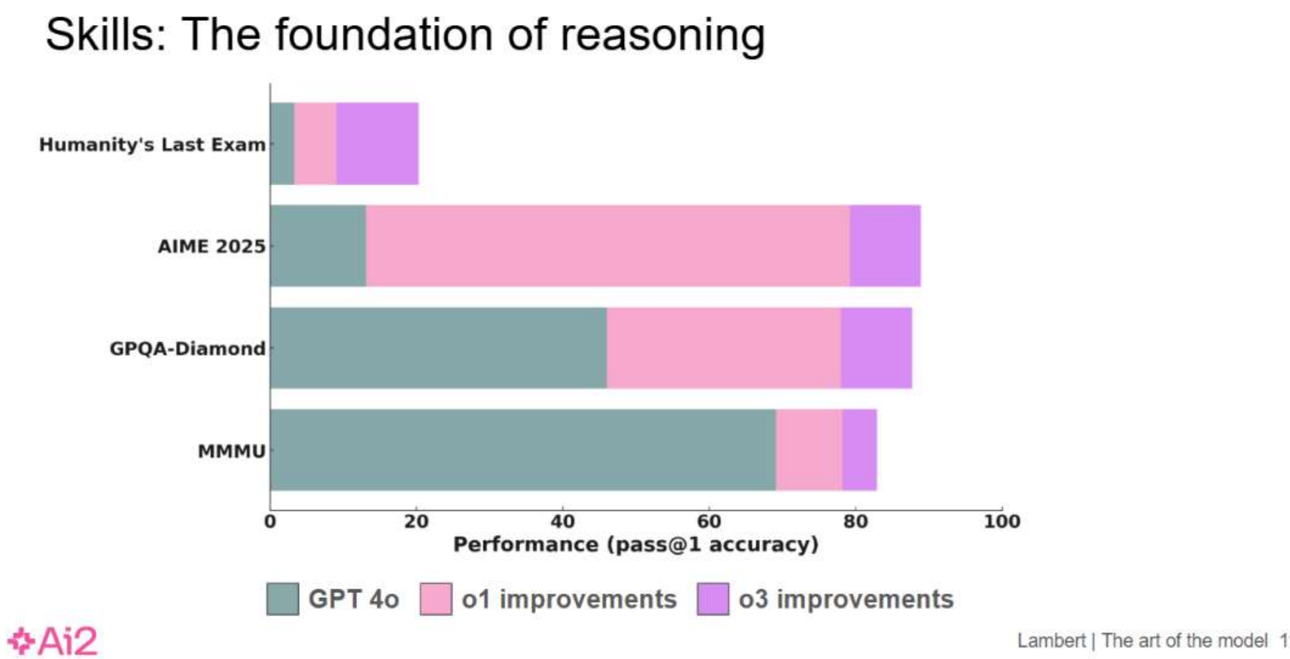
Looking ahead, we plan to expand our use of AI carefully but steadily. As models get even better (and they will), we expect to trust them with more tasks.

But for now, we remain cautious and will continuously monitor for new risks. The regulatory environment around AI is evolving, and as fiduciaries we must ensure that using these tools do not introduce compliance issues or hidden biases in our decisions. Each new model or tool will be evaluated (much like we'd evaluate a new data vendor) before adoption.

In conclusion, we are convinced that we have reached a turning point: AI co-pilots like o3 are now powerful enough that not using them would put us at a competitive disadvantage. At the same time, using them wisely – with robust oversight and clear understanding of their limitations – can give us a significant edge in research productivity and insight generation.

We believe the ultimate takeaway is this: **the AI tools have arrived, and they work** – not infallibly, but as extremely powerful extensions of our analytical capabilities. Those who learn to harness them early, while avoiding the pitfalls, will be better positioned to process information and make informed decisions in an increasingly data-saturated market. In a world where informational edge is fleeting, the reasoning edge – the ability to ask the right questions and synthesize answers quickly – appears ever more crucial. These new AI models, used properly, appear to be a means to sharpen that edge.

Appendix A



Skills: The foundation of reasoning
Humanity's Last Exam
AIME 2025
GPQA-Diamond
MMMU
0
20
40
60
80
100
Performance (pass@1 accuracy)
GPT 4o
o1 improvements
o3 improvements
Ai2
Lambert | The art of the model 11

Appendix B

🧠 **The Breakthrough: Bootstrapping Reasoning**

At the heart of this approach is a **shift from learning fixed outputs to learning reasoning *procedures***—the *how* rather than just the *what*. Here's a step-by-step abstraction of what's happening:

---

1. Initial Training with Correct Answers

At first, the model is given:

- A question.
- A set of possible answers, where one is correct (and ideally more "deeply" correct).
- Sometimes, an explanation or reasoning chain that leads to the correct answer.

This phase resembles traditional supervised fine-tuning. The model learns:

- Which final answers are likely.
- How to construct paths (chains of reasoning) that correlate with correctness.

---

2. Identifying & Reinforcing Reasoning Paths

Once the model gets better at **generating intermediate steps** (sometimes called *scratchpads*, *rationales*, or *thought traces*), training focuses on:

- Selecting *which reasoning paths* correlate with good final answers.
- Scoring or ranking these paths (sometimes via another model or a reward function).
- Encouraging the model to *self-consistently* choose and follow high-quality chains.

Think of this as training the model not to just "guess the answer," but to *learn the process that gets there*.

---

3. Bootstrapping Without External Labels

Here's where the key breakthrough of the *StAR* method kicks in:

- Once the model has internalized some good reasoning habits, it is then **challenged with questions *without* known correct answers**.
- It generates multiple candidate reasoning chains → evaluates them (internally or via a helper model) → selects the most promising one → adopts that as the "teacher."
- In this way, the model **starts training itself**—bootstrapping from its own increasingly robust reasoning capabilities.

This is the *STaR* insight: the model learns *to reason* by teaching itself *how to reason better*. Crucially, it's not relying on the final answer anymore, but on internal consistency, plausibility, and generalization of the reasoning steps.

---

4. Outcome: Generalized Reasoning Capacity

Now the model can:

- Tackle questions it's never seen before.
- Evaluate and refine reasoning chains on the fly.
- Improve its own outputs during inference, not just via training.

So **the correct answer is no longer needed as an external crutch**—the model has internalized a *search-and-verify process* that makes it more powerful at independently finding the right answers.

Analogy

Think of it like this: Early on, a math student is shown the steps to solve problems. Later, even if the answer key is taken away, they can check their work by reviewing their own steps. Eventually, they become confident and flexible problem solvers who don't just memorize answers—they *know how to reason toward them.*

Summary:

This is how reasoning models become **self-improving problem solvers**, not just answer predictors

Appendix C

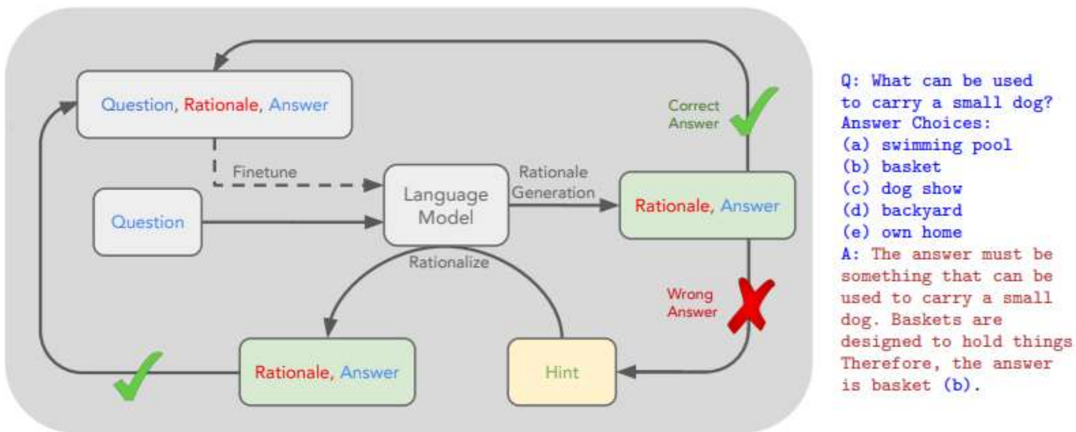


Figure 1: An overview of STaR and a STaR-generated rationale on CommonsenseQA. We indicate the fine-tuning outer loop with a dashed line. The questions and ground truth answers are expected to be present in the dataset, while the rationales are generated using STaR.

Appendix D

An Example Prompt for Company Specific Research

When you review stock X provide me with the following, and follow the instructions I have provided

· Information on operations. Products, geographies, segments, place in the supply chain.

· Competitive information – industry size/dynamics, competitive advantages and positioning.

· A general industry primer- level set the players. Do this as if you were Gartner providing an industry report.

· Size of their particular end markets, growth, and emerging opportunities.

· What issues have been topical to the company in the last couple years, what comes up with analysts and on earnings calls?

· Financial information including last 3 years growth, general margin / profitability profile, returns of the share price.

· Any information regarding unit economics and scalability of the business.

· Nuanced qualitative analysis around the business – what competitive dynamics does it exhibit? Network effects, switching costs, platform power, physical assets, unique capabilities, etc

· Analysis of risks/headwinds to the business. Regulatory, Competition, industry specific cycles.

· Information about management and their track record. Detail out capital allocation decisions, including M&A returns if applicable from the last 5 years, but also dividends that are paid, buybacks, capex, etc.

· Highlight any notable catalysts that are foreseen on the horizon. Also analyze and scrutinize the stock specific trading nature. How does it perform in up and down markets?

· Comment at length on the success of capital deployment – dividends, buybacks, capex, M&A, etc. Be detailed and link them back to performance of the underlying shares. Find a way to benchmark the success against peers.

· Identify, access, and demonstrate data sources that are out-of-consensus that could be leading indicators of performance for the company.

· Where possible, provide charts that show revenue progress, gross margin progress, and 2 others that you think are relevant and interesting, but necessarily direct company financials.

In terms of materials, do your best to draw and download official filings if I do not provide them. This includes investor day transcripts, earnings transcripts, 10K, 10Qs, whatever else you can find.

If I do upload them, I want you to use them exhaustively. Pull out real numbers, excerpts, tables, etc. Use these as your first sources, but make sure to do external diligence as well on the web before you present your findings.

I want you to integrate from a novel range of sources on top of those official filings to craft your total narrative. In the presentation of your information, try not to be redundant.

Overall, I want this to be a prompt that saves me a lot of time in getting spun up/familiar on a new way. In some ways it should be a replacement for a high-quality industry analyst. I want the output to be structured, concise, and dense with information. It should include insightful nuance and interesting commentary on competitive dynamics and standing of the company. You should think of / or even access creative data sources or angles if you think of them and integrate that as well. If you do not know something, do not hallucinate and state clearly what you do not have information wise. Please let me know you understand and then I will provide a stock symbol.

Appendix E

Prompt Engineering – General Principles

Prompt Engineering is a new and fast evolving technique that arose as a result of the growing realisation that the deepest powers of LLMs can only be accessed by question structures that have inbuilt carrot and stick that can stir the LLM to reach more deeply into its powers.

Leading LLM lab Anthropic, the creator of the Claude series of models, has a useful Prompt Engineering Guide[15], some key insights from which include:

1. Be clear and specific - state task upfront, provide context.
2. Use examples - show the format/style you want.
3. Encourage thinking - ask Claude to "think step-by-step."
4. Leverage Claude's knowledge - include relevant context.
5. Use role-playing - "A"
6. Specify your audience - tell Claude who content is for.
7. Define tone and style - describe desired voice.
8. Define output structure - provide outlines/lists to cover.
9. Be specific about summaries - ask for specific aspects.
10. Use document names - refer to attachments by name (ex: <Style Guide>)
11. Ask for citations - request specific sections/pages.
12. Specify desired format - tables, bullets, etc.
13. Allow uncertainty
    - tell Claude it's okay to say “I don't know.”
14. Include all context - Claude doesn't retain info between conversations.
15. Iterative refinement - give specific feedback for improvements.

[15] https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/overview

GENERAL INFORMATION

This document has been produced by Intermede Investment Partners Limited ("Intermede"), a limited company registered in England and Wales (Reg. No. 8827481), having its registered office at 6 Warwick Street, London W1B 5LU, United Kingdom.

Intermede is authorised and regulated by the UK Financial Conduct Authority with firm reference number: 624278. Intermede is registered as an Investment Adviser with the Securities and Exchange Commission in the United States.

This document has been provided for information purposes only and has not been provided as either a solicitation or recommendation with respect to the purchase or sale of any securities or an offer to provide investment services. The information contained in this document does not constitute an offer to purchase securities to any person in the United States or to any U.S. Person as such term is defined under the United States Securities Act of 1933, as amended.

The information provided is believed to be accurate at the time of compilation and is provided by Intermede in good faith. However, no representation is made or implied as to the quality, accuracy, reliability, timeliness or completeness of the material. Forward-looking statements reflect our views and assumptions with respect to future events as of the date of this document and are subject to risks and uncertainties. Given such uncertainties, undue reliance should not be placed on any forward-looking statements. Any opinions or estimates expressed are our current opinion as of the date of this document and are subject to change without notice. Except in so far as liability under any statute cannot be excluded, Intermede, its directors, employees, representatives and consultants do not accept and disclaim any liability (whether arising in contract, tort, negligence or otherwise) for any error or omission in the material or for any loss of damage (whether direct, indirect or consequential) that may be suffered by you, the recipient of the information, an investor or any other person. Certain information contained herein has been obtained from third-party sources and such information has not been independently verified by Intermede. No representation is made as to the accuracy or completeness of such information. Index performance and statistics are provided for comparison purposes only.They are not meant to be indicative of Intermede portfolio performance, asset composition, volatility, or other attributes, which will differ significantly from those of the index. Index returns do not take into account transaction costs or other fees and expenses. It is not possible to invest directly in an index.

The information contained in this document is not intended to be and should not be taken as investment, tax, legal of any other type of advice. It does not take into account the objectives, financial, tax or legal situation or needs of any investor or purport to be comprehensive and should not be relied upon as such. You must obtain professional or specialist advice before taking, or refraining from, any action on the basis of the content on this document.

This document is not intended for and is not to be distributed to retail investors.

Intermede Investment Partners Ltd claims compliance with the Global Investment Performance Standards (GIPS®). Copies of all GIPS® composites is available on request. GIPS® is a registered trademark of CFA Institute. CFA Institute does not endorse or promote this organization, nor does it warrant the accuracy or quality of the content contained herein.

Rates of return are for illustrative purposes only, with dividends reinvested and are net of fees and other expenses. Past performance is not a reliable indication of future performance, does not guarantee future results and all investments are subject to the risk of loss. As with all investments, income may fluctuate, values may fall as well as rise with changes in the market and an investor may not get back the amount invested. Income from investments may fluctuate. An investment with Intermede is not a deposit or liability and is not guaranteed by Intermede, its shareholders or employees or any affiliate. All information is to be treated as confidential and may not be reproduced or redistributed in whole or in part in any manner without the prior written consent of Intermede. This document is not intended for public use or distribution.

This information is issued and approved by Intermede, part of the Insignia Financial Ltd group of companies (comprising Insignia Financial Ltd ABN 49 100 103 722 and its related bodies corporate) ("Insignia Group").

The capital value, payment of income and performance of any investment with Intermede is not guaranteed. An investment with Intermede is subject to investment risk, including possible delays in repayment of capital and loss of income and principal invested.

UNITED KINGDOM

This document is directed only at persons in the United Kingdom who are professional clients within the meaning of Directive 2004/39/EC.

Persons who do not have professional experience in matters relating to investments should not rely on this document.

AUSTRALIA

In Australia, this information is solely for the use of persons who are "wholesale clients" (as defined in the Corporations Act 2001 (Cth))', which include (but are not limited to): (a) Australian financial services licensees, including but not limited to research report providers and financial advisers; (b) registrable superannuation entities;

(c) a person or entity who controls gross assets of at least $10 million; or (d) a person or entity that has obtained a qualified accountant's certificate (within the last 6 months) stating they have net assets of at least $2.5 million, or a gross income for each of the last two financial years of at least $250,000. This document must not be distributed to "retail clients" (as defined in the Corporations Act 2001 (Cth)). This document is not intended to be distributed or passed on, directly or indirectly, to any other class of persons in Australia.

Intermede Investment Partners Limited is exempt from the requirement to hold an Australian financial services licence under the Corporations Act in respect of the financial services it provides. Intermede Investment Partners Limited is regulated by the FCA under United Kingdom laws, which differ from Australian laws.

This document is approved and distributed in Australia by MLC Asset Management Pty Limited ABN 44 106 427 472, AFSL No. 308953.

This document contains general information and may constitute general advice. Any advice in this communication has been prepared without taking account of individual objectives, financial situation or needs. It should not be relied upon as a substitute for financial or other specialist advice.

Before making any decisions on the basis of this document, you should consider the appropriateness of its content having regard to your particular investment objectives, financial situation or individual needs.

None of MLC Asset Management Pty Limited, MLC Asset Management Holdings Limited ABN 16 134 158 517 nor any other member company in the Insignia Group accepts any liability for or provides any guarantee in respect of the activities of Intermede.

The capital value, payment of income and performance of any financial product referred to in this document or an investment in any product offered by Intermede or any other member of the Insignia Group is not guaranteed. An investment in any such financial product is subject to investment risk, including possible delays in repayment of capital and loss of income and principal invested. Portfolio information provided in this presentation relates to a composite portfolio made up of all client portfolios managed by Intermede over the period or at the date specified and is not indicative or a guarantee of the composition of any portfolios managed by Intermede.

NEW ZEALAND

In New Zealand, this information is solely for the use of researchers, financial advisers and wholesale clients.This material has been prepared without taking into account a potential investor's objectives, financial situation or needs and is not intended to constitute personal financial advice and must not be relied on as such. Recipients of this document, who are not wholesale investors (in accordance with Schedule 1, Clause 3 Financial Markets Conduct Act 2013), or their duly appointed agent, should consult an Authorised Financial Adviser and the relevant Product Disclosure Statement where applicable.

UNITED STATES

This document is intended only for an investor that is a "qualified purchaser" as defined in the US Investment Company Act of 1940 and is not intended for public use or distribution. It is not to be construed as investment advice.

This material has been prepared as, and is furnished solely in connection with, a one-on-one presentation by Intermede and is not for public distribution. It is not to be construed as investment advice. This document is directed only at persons "in the United States" as defined by the United States Securities Act of 1933, as amended (the "US Securities Act"). This document is not to be distributed to any investor that is not a "qualified purchaser" as defined by the United States Investment Company Act of 1940 ("US Company Act").

Rates of return and other assumptions are for illustrative purposes only, with dividends reinvested and are net of fees and expenses. As an example, an account with an initial

$1,000,000 investment on January 1st, 2019, earning a recurring 2.4125% quarterly gross return (10.00% annualized), and paying a 0.20% quarterly management fee (0.80% annual fee) based on end of periods assets, would have grown to $1,610,510 on a gross of fees basis and $1,547,640 on a net of fees basis by December 31st, 2023 (5 years). Actual return and performance may vary significantly from the illustration and examples given.

This material has also included the application of potential market return based in part by historical market data.Thus, the potential effect of future market risk is not taken into account.There is no guarantee that the intended projections or results will be achieved. Past performance does not guarantee future results and may differ for different time periods. All investments are subject to the risk of loss.

Values may fall as well as rise and you may not get back the amount you invested. Income from investments may fluctuate. Certain information contained herein has been obtained from third-party sources and such information has not been independently verified by Intermede. No representation is made as to the accuracy or completeness of such information. Index performance and statistics are provided for comparison purposes only and relate to the MSCI ACWI, a market capitalization weighted index designed to provide a broad measure of equity-market performance throughout the world.The index is maintained by Morgan Stanley Capital International (MSCI) and is comprised of stocks from both developed and emerging markets.They are not meant to be indicative of Intermede portfolio performance, asset composition, volatility, or other attributes, which will differ significantly from those of the index. Index returns do not take into account transaction costs or other fees and expenses. It is not possible to invest directly in an index.

OTHER COUNTRIES

The distribution of this document into any other jurisdictions may be restricted and accordingly any persons in such jurisdictions into whose possession this document comes are required to inform themselves about and to observe such restrictions.