EDGE RESEARCH
PONDERING POSSIBILITIES

NUS National University of Singapore | Centre for Governance and Sustainability
NUS Business School

CASE STUDY

Using AI Tools to Research Emerging Technology Trends

December 2025

Tim ZHANG; Lawrence LOH; Joshua ARJANTO; CHEN Xiaojing; HU Shiyou; Asher Ethan KOH; Muhammad Khaizuran Bin MOHAMAD ROSLE

Project Lead: Tim ZHANG, Lawrence LOH,
Project Team: Asher Ethan KOH, CHEN Xiaojing, HU Shiyou, Joshua ARJANTO, Muhammad Khaizuran Bin MOHAMAD ROSLE
Editorial Support: ANG Hui Min, Verity THOI

This report was created as a Field Service Project (FSP) under the National University of Singapore (NUS) Business School, under the supervision of Professor Lawrence Loh, Director, Centre for Governance and Sustainability at the School.

Published December 2025

TABLE OF CONTENTS

1. Abstract

This project, conducted in partnership with Edge Research Pte Ltd, examines how large language models (LLMs), GPT-5, Gemini, and DeepSeek, can systematically augment investment research across five emerging-technology sectors from G20 countries: Artificial Intelligence, Robotics, Quantum Computing, Space, and Fusion. We evaluate LLMs not as end-to-end automation, but as accelerators within a governed, human-in-the-loop workflow. Across sector onboarding, multilingual discovery, numerical extraction, and first-draft synthesis, LLMs reduce mechanical workload by roughly half while broadening source coverage and enabling faster iteration. However, our findings show that human judgment remains indispensable for causal reasoning, credibility assessment, evidence weighting, and final investment interpretation; LLMs shift analyst effort, but do not replace it.

To ensure evidence hygiene at scale, the team developed a Python-based fact-checking pipeline that automates link reachability, canonicalisation, publication-date extraction, and AI-generation risk screening, later extended into a low-code web interface. This prototype reduced verification time from 4.3 hours to 1.25 hours per 50 links, a 71% reduction, with projected savings of over 90% under full integration of headless browsing and GPTZero API scoring. Alongside this, a comprehensive source database with over 250 records of LLM-cited sources, graded by recency and credibility, enabled reproducible evaluation of model outputs and surfaced failure modes such as hallucinations, outdated citations, and contextual drift.

Overall, the project demonstrates that LLMs are most effective as structured force multipliers when paired with rigorous verification tooling and human oversight. With further hardening of the fact-checking pipeline and clearer task-to-model routing, AI-assisted research can achieve significant efficiency gains while preserving the analytical standards required in professional investment workflows.

2. Introduction & Problem Context

This Field Service Project (FSP), titled "Using AI Tools to Research Emerging Technology Trends," is conducted in partnership with Edge Research Pte Ltd, led by CEO Tim Zhang.

2.1 Client Brief and Scope (Emerging Tech: AI, Robotics, Quantum, Space, Fusion)

Our project's central aim was to explore how AI-based research tools can be effectively used to gather information and conduct investment research with a focus on emerging technologies from G20 countries; specifically, AI, Robotics, Quantum Computing, Space Technologies, and Nuclear Fusion. The client's goal was to understand how recent advancements in Large Language Models (LLMs) could improve the efficiency, quality, as well as scalability of investment research across the various sectors, whilst maintaining accuracy, credibility, and recency; qualities which are all expected in professional investment analysis (Defend & Mortier, 2025).

Unlike a methodological study of LLMs themselves, the project primarily focused on using these AI tools to conduct real research on the aforementioned sectors and to evaluate the credibility of AI-assisted findings in practice. Alongside the primary research effort, a secondary objective was to attempt the development of a sustainable fact-checking process; eventually leading to a prototype tool capable of validating AI-generated investment research report outputs through link reachability, recency analysis, and basic source-credibility assessment using AI generation scores.

During the initial phase of the project (June 2025), the team experimented extensively and familiarised ourselves with three of the leading LLM-powered platforms, including Google Gemini, OpenAI ChatGPT, and DeepSeek. This familiarisation period allowed us to understand their individual capabilities, interfaces, and limitations through hands-on use across a variety of self-exploratory research tasks. Each team member tested a variety of different features such as document reading, web extraction, image-reading, numeric table parsing, citation generation, and long-context reasoning, to determine where each tool performed best in a professional research workflow. Eventually, the team then embarked on a 5-month long weekly research journey, where we used these outputs to construct full sector value-chain analyses and business-impact briefings for each emerging-technology vertical, while simultaneously refining the fact-checking workflow that underpinned the second half of the project.

The second component of the project expanded into developing a prototype fact-checking tool to automate parts of the verification process. This tool aimed to address a recurring challenge in our AI-assisted research: ensuring source reachability, recency, and credibility in automatically retrieved references by the LLMs. Together, these two streams of work (comparative research and prototype development) provided the foundation for actionable recommendations on how AI tools can be integrated into investment research workflows in both a sustainable and responsible manner.

2.2 Baseline: The Traditional Investment-Research Workflow

In a traditional investment research framework, analysts typically engage in a manual process of collecting, verifying, and synthesising information from various primary and secondary sources to derive insights about sectors or value chains. This process often involves reviewing market research reports, analyst publications and leveraging third-party data providers such as Bloomberg and S&P Capital IQ for company and market data. Analysts must progress through this series of structured, yet highly manual stages to build domain understanding, construct market views and eventually build actionable insights (Collin, 2025).

The workflow generally follows sequential stages. Analysts start with **sector onboarding**, building foundational literacy through primers and reference materials. They then conduct **market sizing**, gathering quantitative and qualitative data to assess the sector's attractiveness. Next, they map the **value chain** and screen companies to form a structured universe and shortlist. This is followed by **company deep dives**, analysing financials, strategy, and valuation to gauge traction and competitiveness. The process concludes with **report writing and review**, where insights are synthesised into a final report and validated for accuracy and rigour (Twin, 2025).

Across all stages, the friction points are predictable: numerous hours of manual research, fragmented data, conflicting estimates, time-intensive extraction from PDFs and transcripts, and the tedious compilation and reconciliation of materials. Analysts also spend significant effort tracking the recency and credibility of each source. As our client, Tim, noted, completing this full five-stage process for just one sector typically requires *two to three months of full-time analyst work* to reach actionable recommendations. These pain points make the baseline process slow, costly, and repetitive; providing a clear benchmark for comparison against an LLM-assisted research workflow.

2.3 Research Questions (Augment vs Replace)

This project explores how large language models (LLMs) and related AI tools can enhance the traditional investment research workflow, focusing on the balance between augmentation and replacement. While manual processes offer both rigour and traceability, they are also slow, labour-intensive, and limited by human bandwidth; particularly in parsing lengthy documents, reconciling conflicting data points, and maintaining consistency across sources. The key question is whether AI can accelerate these workflows without compromising analytical depth or evidence quality. Below is a list of key research questions we considered and analysed (LexisNexis, 2024).

Research Questions:

This study considers four guiding questions:

1. Where in the investment research process (such as sector onboarding, market sizing, value-chain mapping, and company deep dives) can AI tools most effectively augment analyst work?

2. How do AI-assisted workflows compare with traditional manual research in terms of speed, coverage, and evidence quality?

3. What limitations or risks arise when integrating AI outputs, particularly around hallucinations, citation accuracy, or contextual loss?

4. What is the right balance between automation and human oversight to ensure credible, reproducible insights?

These questions form the analytical foundation for the project. They will guide the comparative assessment in Section 5 (Traditional vs LLM-Assisted Research) and the resulting recommendations in Section 11 (Process Recommendation: Augment vs Replace), where evidence is used to determine where AI can credibly augment (and where it cannot yet replace) the traditional research process.

2.4 Success Criteria (Quality, Recency, Cost/Time, Reproducibility, Hallucination Risk)

To evaluate whether AI tools can credibly augment the investment research process, the project defines a clear set of success criteria across five dimensions. These criteria measure both the efficiency gains and the quality outcomes of LLM-assisted workflows compared to traditional manual research (Niel et al., 2024).

Quality and Accuracy:

Outputs must be factually correct, internally consistent, and supported by credible primary or secondary sources. A successful AI-assisted workflow should produce findings that are at least as accurate as traditional manual research, while making it clear where each fact or number came from. Every statement should be traceable to an identifiable source, and the reasoning behind key conclusions should be easy for another analyst to review and verify (Shah, 2025).

Recency and Coverage:

Given that emerging technology sectors evolve rapidly, the ability to retrieve and synthesise up-to-date information is critical. Success is measured by whether AI tools can surface recent data points (e.g., latest financials, funding rounds, product launches, regulatory actions) and provide comprehensive coverage across sub-sectors and geographies.

Cost and Time Efficiency:

AI augmentation should reduce the time and labour required to complete core research tasks such as data extraction, summarisation, and drafting. Improvement is quantified by comparing average task duration and analyst hours between manual and AI-assisted workflows, without sacrificing quality or rigour.

Reproducibility and Consistency:

A reliable workflow should produce consistent outputs when repeated under similar prompts or datasets. Success is defined by the ability to replicate findings, maintain standardised formatting, and avoid subjective drift across different analysts or model sessions.

Hallucination and Risk Management:

Reducing factual hallucinations and citation errors is essential for credibility. The benchmark for success is a verifiable output where every quantitative figure and factual claim can be traced to a live, reputable source (Moore, 2025). Hallucination mitigation is evaluated through cross-checks, link logging, and manual verification steps.

Together, these criteria form the evaluative framework for subsequent sections. They underpin the analysis in Section 5 (Traditional vs LLM-Assisted Research), where outcomes are compared along each dimension, and inform the process recommendations in Section 11, which propose how to operationalise a hybrid human-AI research model at scale.

3. Methodology Overview (How We Worked)

In order to maintain comparability of our outputs, we instituted a common value-chain framework across all sectors. We segmented the value-chain schema into 'Suppliers' (inputs and tooling), 'Upstream' (core hardware/compute), 'Midstream' (platforms and control layers), and 'Downstream' (end-applications). Our client and CEO, Tim Zhang, determined the parameters, definitions, and standards of evidence while the research ran in weekly cycles from June to October. For those readers that are interested in the full granular details on operations, you may find a comprehensive narrative of weekly meeting cadences, role assignments by track, our five-stage weekly workflow and month-by-month research timeline provided in Appendix A: Full Overview of Research Methodology. Here we retain only the elements needed to interpret findings.

3.1 Tools Used (ChatGPT, Gemini, DeepSeek; Python Notebooks)

We kept tooling lightweight and task-specific. Generally, Gemini was used for first-pass breadth and recency (news, announcements, funding, policy); ChatGPT for document reading and structured extraction from filings/transcripts and for producing clean comparison tables; DeepSeek for Asia/China enrichment and multilingual look-ups. Deeper capability comparisons deferred to Section 4: LLM Capabilities & Cross-Model Comparison. Python notebooks were used to develop and validate the fact-checking prototype as well as for data analysis. Python was chosen for its readable markdown mode, step-by-step structure and tight integration with Python's library ecosystem, enabling rapid iteration and clear inline documentation (Ramos, 2025). The prototype's objectives, architecture, and results are detailed in Section 8: Python Prototype.

3.2 Data Governance: Citations, Link Logging, Source Recency Checks

The data we collectively found and collated was continuously recorded in a source database excel sheet capturing features such as URL, date accessed, value-chain position, and notes. We assigned each source a credibility score and a recency rating to prioritise up-to-date information. Cross-source triangulation was required for key facts, and any unresolved conflicts were explicitly annotated in the deck. This governance ensured traceability, reproducibility, and clear separation between LLM-suggested leads and human-verified evidence.

For a full view of the database (its fields, scoring rubrics, examples, and how records are used in analysis), see Appendix C: Source Database.

4. LLM Capabilities & Cross-Model Comparison

In this section we aim to compare the different nuances in the LLM's usage and how we have learnt to utilise each LLM to the best of our ability.

4.1 Model Snapshots & Versions Used (With Dates)

Over the course of the project, we consistently relied on three systems, each playing a distinct and complementary role.

ChatGPT (GPT-5 Thinking family)

Served as our main analytical environment, used daily for drafting, document ingestion, number extraction, reconciliation, table construction, and code generation. The model's stability and ability to handle long, structured documents made it the backbone of our workflow.

Gemini (Deep Research; Flash 2.5)

Employed during web-grounded research sprints, especially when we needed a broad, up-to-date sweep of sources. Its recency advantage made it well-suited for gathering fast-moving market information, regulatory changes, and technology updates. Its behaviour in recency-focused retrieval is consistent with descriptions in the Gemini technical report (Google DeepMind, 2024).

DeepSeek (latest public release during project window)

Used selectively when Chinese-language materials were central：corporate disclosures, industry standards, policy documents, and Chinese market news. Outputs were then normalised, translated more precisely, and stylistically harmonised in ChatGPT.

Across the project period, this division of labour stayed relatively stable: GPT-5 took on most of the analytical and editorial work; Gemini provided breadth and timeliness for evidence gathering; DeepSeek helped us surface and interpret China-specific materials that often lacked English analogues.

4.2 Comparative Strengths

In practice, ChatGPT proved most dependable for "closed-book" document work: reading PDFs, performing unit-faithful numeric extraction, reconciling conflicting figures across sources, and turning raw notes into clean tables, figures, and code. Gemini excelled when breadth and recency mattered; when seeded with focused prompts, it returned balanced, source-linked overviews and surfaced counter-arguments that improved report neutrality. DeepSeek's advantage was recall and nuance for China-specific materials, mapping Chinese technical terminology to defensible English equivalents before final editorial passes in ChatGPT (Refer to *Table 1*, Appendix F: Tables).

4.3 Cost, Latency, Reliability (Hallucination/Coverage Issues)

We controlled cost and turnaround by specialising tasks across models. ChatGPT handled iterative drafting and document-grounded analysis; it is responsive for short edits but overall moderate in speed and slower on long, structured outputs such as large tables, multi-document synthesis, and code blocks. Gemini ran in scheduled, web-grounded sweeps when new topics or updates were needed; wall-clock latency was batch-oriented and longer, but the returns were well-organised, citation-rich digests. DeepSeek was used on demand for Chinese-language retrieval, delivering moderate–fast responses on targeted lookups. Reliability varied by task: ChatGPT showed the lowest hallucination rate in our doc-grounded tasks yet can over-generalise on breaking topics without fresh sources; Gemini's breadth occasionally introduced over-aggregation, duplicate links, and inconsistent as-of dates; DeepSeek sometimes carried translation/scale drift when converting Chinese language to English language figures (including 亿 to billion mapping), and occasional A/H-ticker mismatches. To mitigate these issues we required source-backed prompts and quick arithmetic/identity checks for ChatGPT; imposed top-N caps, domain de-duplication, and explicit date/provenance constraints for Gemini; and ran a glossary- and style-normalisation pass in ChatGPT for DeepSeek outputs with explicit FX/date capture before integration (Refer to *Slide 1,* Appendix G: Slide Snapshots). Similar reliability patterns have been documented in surveys of LLM hallucination and multilingual alignment (Ji et al., 2023; Zhang et al., 2024).

4.4 Failure Modes & Examples (From Deck Screenshots)

Across models we observed distinct error patterns that required disciplined controls. ChatGPT/GPT-5, which we used for analysis and table construction, occasionally produced plausible-looking numeric hallucinations when parsing filings or scraped tables, such as pairing a quarter-end enterprise value with TTM denominators, reading parentheses in disclosures as negative EBITDA rather than stylistic formatting, and drifting between units or currencies when OCR artefacts were present. Gemini (Deep Research/Flash 2.5) was strongest on recency but its breadth sometimes came with source-hygiene issues: over-aggregation of near-duplicate links, inconsistent as-of dates across citations, link rot, and unfiltered vendor material that could inflate claims unless prompts enforced strict dating and provenance. DeepSeek, used for Chinese-language discovery, showed high recall yet introduced terminology and scale drift during translation (e.g., mapping "亿" imprecisely to "billion"), occasional A/H-ticker mismatches, and RMB-to-USD conversions without explicit FX references, all of which needed normalisation before integration. Similar Chinese language to English language scaling inconsistencies are highlighted in multilingual LLM alignment research (Zhang et al., 2024).

A representative example appears in our satellite-operator comps, where the red cells mark values from an early ChatGPT pass that seemed coherent in-row but failed deterministic checks. We corrected the table by manually verifying each figure against the original source (same as-of date and currency), re-keying the base fields, and recomputing the ratios before inclusion in the deck (Refer to *Slide 2:* Appendix G).

4.5 Our Solution: Workflow Design & Control Framework

As the project progressed, we realised that no single model could realistically handle every part of the workflow. GPT-5, Gemini, and DeepSeek each responded to tasks in their own ways, so we gradually arrived at a process that leaned into their strengths while managing their weaknesses. In practice, the work fell into three phases: we first cast a wide net to collect external sources, then pulled those materials apart to create document-grounded extracts, and finally reconciled everything into a consistent dataset. Gemini and DeepSeek were the most effective for that initial sweep because they surfaced recent and multilingual information quickly, while GPT-5 became the space where we re-read documents, extracted figures, checked units, and shaped the results into usable tables or code. This structure was not something we designed upfront, it emerged as the only stable way to keep the analysis accurate as the dataset grew.

This general structure only worked because we paired it with more targeted solutions for each model. GPT-5 occasionally produced numbers that looked reasonable until we checked them against the filing, so we began locking units inside the prompt and running deterministic arithmetic checks before accepting any derived metrics. Gemini sometimes returned the same link multiple times or mixed evidence from different reporting dates, which we controlled by limiting the number of domains it could pull from and forcing it to state an explicit "as-of" date for each citation. DeepSeek's translations were rich in detail but could drift in scale, especially with terms like "亿", so we ran a normalisation pass in GPT-5 to correct numerical phrasing, verify ticker symbols, and make sure any RMB-to-USD conversions were tied to an explicit FX reference.

This combined approach is a general workflow plus targeted controls, this ended up being the most reliable configuration. It allowed us to draw on GPT-5 for analytical synthesis, use Gemini when recency mattered, and rely on DeepSeek for high-recall Chinese-language discovery, without letting the distinct failure modes of each model leak into the final deliverables. The workflow we established here directly informs the operational recommendations presented under Section 11.

4.6 How Usage Evolved as Newer Models Released

Our model usage shifted as both the project and the available tools evolved. At the start, we relied heavily on GPT-4, which was capable of strong reasoning and drafting but struggled to surface the most recent web materials. Because early-stage research required wide and timely evidence gathering, we leaned more heavily on Gemini (Deep Research/Flash 2.5) for recency-based sweeps and on DeepSeek for Chinese-language sourcing.

This changed mid-project when GPT-5 was released. Its improvements in numeric extraction, table construction, document reconciliation, and code-assisted analysis directly matched the needs of the second half of our workflow, which had shifted from conceptual mapping to quantitative estimation and validation. Once GPT-5 proved more consistent across these tasks, it became the central model for synthesis and final deliverables. Gemini remained valuable for periodic web updates, and DeepSeek continued to provide Chinese market visibility, but the bulk of the analytical effort migrated to GPT-5. This transition reduced rework and accelerated the turnaround time for tables, figures, and appendices.

5. Traditional vs LLM-Assisted Research

This section compares the established, manual investment-research workflow used across industry with the emerging LLM-assisted workflow trialled by the project team. Drawing on sectoral research experiences across AI, Robotics, Quantum, Space, and Fusion, as well as firsthand validation from our client Tim (based on his tenure in investment research), this analysis evaluates how LLMs reshape effort profiles, output quality, and analyst time allocation. While the benefits of LLMs are meaningful, particularly in accelerating early-stage scoping, data processing, and first-draft generation, these gains are counterbalanced by persistent limitations in reasoning, judgment, and credibility verification. The comparison below clarifies where LLMs augment the research process versus where human capabilities remain central and irreplaceable.

5.1 Effort & Time Estimates for Core Tasks (Manual Baseline vs LLM-Assisted)

5.1.1 The Traditional Investment Research Workflow

Industry practice for investment analysts typically relies on a structured six-phase workflow that spans **8-12 weeks** for a comprehensive sector or thematic report. Each phase is sequential and labour-intensive, requiring multi-disciplinary expertise across data sourcing, financial modelling, narrative synthesis, report drafting, and client communication (Refer to *Slide 3,* Appendix G).

1. **Scoping & Design (2-4 days)**
   Analysts begin by defining the research question, establishing scope, and framing hypotheses. This stage is inherently iterative and often shaped by initial conversations with stakeholders, past coverage, and internal knowledge. Manual workflows require analysts to independently explore the opportunity set and identify gaps.

2. **Data Collection (1.5-2 weeks)**
   This phase is the heaviest in terms of raw hours. Analysts manually gather macro statistics, regulatory updates, company filings, earnings transcripts, market-share data, and expert insights, often dispersed across paywalled sources. Data cleaning and organisation are major sub-tasks (Vipond, 2025). Human judgement is required to determine relevance, reconcile conflicting data, and ensure source quality.

3. **Analysis & Modelling (2-3 weeks)**
   Traditional workflows involve building financial models, benchmarking peers, constructing sensitivity scenarios, and validating assumptions. This requires technical modelling expertise and domain understanding. Errors can propagate easily, so analysts spend long hours verifying calculations and cross-checking inputs.

4. **Synthesis & Narrative Development (1.5-2 weeks)**
   Analysts must translate data and models into coherent investment theses, narratives, and visual exhibits. Writing is performed from scratch and often requires significant redrafting

(CFA Institute, 2020). This stage demands an ability to spot patterns, weigh materiality, and connect disparate insights into a persuasive argument.

5. **Review & Publication (3-5 days)**
   Compliance, peer review, fact-checking, formatting, and editorial refinement occur here. The process is slow due to manual cross-referencing and institutional sign-off requirements.

6. **Dissemination & Follow-Up (2-3 days)**
   Finally, analysts present findings to clients and investors, handle Q&A, and update coverage based on market feedback. Follow-up is periodic, reflecting manual monitoring of news, filings, and sector developments (Grata, 2025).

This traditional workflow is thorough and high-confidence but slow, resource-intensive, and constrained by human bandwidth, especially during data-heavy phases.

5.1.2 The LLM-Assisted Investment Research Workflow

The team's experimentation with LLM tools produced a modified workflow with a materially reduced timeline of **3-6 weeks**, achieved not by eliminating steps but by automating and accelerating key tasks. The workflow retains the six phases but redistributes the effort. (Refer to *Slide 4,* Appendix G).

1. **Scoping & Design (1-3 days)**
   LLMs help analysts shape hypotheses, surface initial angles, and identify data gaps quickly. By prompting models with preliminary sector context, analysts can obtain landscape maps, potential sub-themes, and early framing suggestions. This reduces the time spent on exploratory scoping by roughly half.

2. **Automated Data Collection (3-6 days)**
   LLMs significantly cut the time required for data harvesting by enabling rapid summarisation of company filings, web sources, transcripts, and regulatory documents. AI-driven scraping and classification automate much of the tagging and cleaning work that traditionally consumes analyst hours. This is where the team saw the largest efficiency gains (60-70% time reduction).

3. **Assisted Analysis & Modelling (1-1.5 weeks)**
   LLMs generate preliminary ratio analyses, scenario outlines, peer benchmarks, and sensitivity structures, which analysts then validate and refine. Importantly, LLMs do not replace modelling expertise but accelerate the production of starting points (Pop et al., 2024). Analysts still review assumptions, ensure consistency, and adjust logic. Time savings here averaged 40-50%.

4. **AI-Drafted Synthesis (4-6 days)**
   LLMs produce first-draft narratives, structure key messages, and generate tables or charts. This allowed team members to move quickly from raw insights to near-publishable text (Madanchian & Taherdoost, 2025). Analysts then layered domain judgement and sector-specific nuance onto these drafts. This stage saw strong efficiency gains (50-60%).

5. **Review & Compliance (3-5 days)**
   Human review remains essential, but AI assists with tone harmonisation, citation formatting, and consistency checks. LLMs help flag contradictory statements or unclear reasoning. Time savings are more modest here (~30%), partly due to compliance requirements that cannot be automated.

6. **Continuous Update Loop (Real-time)**
   A major improvement introduced by LLMs is automated monitoring of filings, macro releases, and news. Instead of periodic batch updates, analysts can receive real-time alerts, dramatically improving reactivity and enabling ongoing refinement of research outputs.

Across phases, LLMs free analysts from the most repetitive tasks, allowing them to spend more time on higher-order reasoning rather than mechanical processing.

5.1.3 Comparative Time Impact

The combined effect of LLM assistance is a **40-60% reduction in overall turnaround time** (Refer to *Table 2,* Appendix F**)**. This finding is consistent with both the team's experience and emerging industry consensus. Tasks that previously required highly manual effort (data gathering, summarisation, preliminary modelling, and first-draft writing) experience the greatest acceleration.

However, the human-dependent portions of analysis (judgment, causal inference, credibility checks, and final decision-making) see smaller degrees of reduction because they cannot be reliably delegated to LLMs.

5.2 Output Quality & Coverage: What Improved, What Still Needed Human Judgment

5.2.1 Improvements Observed with LLM Assistance

1. Expanded coverage and breadth

LLMs enabled analysts to review more companies, geographies, and thematic sub-topics than would be feasible manually (Cognizant, 2025). For instance, in sectors such as Robotics and Space (where the number of niche players is extremely large), LLM-based summarisation allowed the team to scan dozens of companies within hours. This expansion of the initial universe improves comprehensiveness and reduces blind spots .

2. Faster synthesis and clearer structure

Across all sectors researched, LLMs produced well-organised and logically structured draft narratives. They were able to transform raw notes, tables, and models into coherent prose. This was particularly useful in AI and Quantum (domains with complex value chains and fast-changing technical landscapes) where LLMs could help distill dense information into digestible formats.

3. Pattern recognition and anomaly surfacing

LLMs were effective in identifying KPI shifts, peer-trend divergences, and recurring themes across filings or earnings calls. For example, during Robotics research, LLMs highlighted recurring mentions of sensor-cost declines across multiple companies, something that would have taken much longer to detect manually. Similarly, in Fusion and Quantum, where technical differentiation is subtle, LLMs helped spot thematic clusters in company roadmaps.

4. Real-time, always-on monitoring

LLM-powered alerts strengthened the team's ability to track regulatory updates, new research releases, macro policy shifts, and quarterly announcements. This continuous flow improved situational awareness and reduced the latency of follow-up analysis (AWS, 2024).

5. Improvements in communication quality

LLMs enhanced the polish of written outputs: sentence clarity, flow, and exhibit integration. This helped accelerate the production of presentation-ready documents and ensured consistent tone across different authors.

5.2.2 Areas Where Human Judgment Remained Essential

Despite the improvements above, several core components of research remained outside the reliable scope of LLM automation.

1. Judgment about materiality

LLMs can list many factors, but they struggle to determine which few genuinely move a stock or matter for valuation. Analysts must weigh magnitude, likelihood, and strategic relevance; tasks that require sector experience and intuition (Kang & Liu, 2023).

2. Causal reasoning and inference

LLMs frequently describe correlations but cannot reliably infer causality. For example, when examining why a Quantum company's revenue dipped, LLMs often attribute generic reasons rather than identifying precise operational bottlenecks. Similarly, in AI and Robotics, understanding the business-model implications of technological change requires human interpretation, not pattern repetition.

3. Credibility checks and contextual grounding

Investment research ultimately relies on calls with management, expert interviews, and cross-firm triangulation. LLMs cannot validate claims with lived industry context, nor can they detect when something “feels off.” Credibility checking remains an analyst-driven activity.

4. Handling ambiguity and grey areas

Sector research often requires analysts to form views amid incomplete information, especially in early-stage domains like Fusion and Space. LLMs struggle when data is sparse, contradictory, or ambiguous; they often fill gaps with generalisations. Humans must adjudicate uncertainty and decide how to frame risks and opportunities.

5. Accountability and ethical standards

Analysts are accountable for their calls, valuations, and recommendations. LLMs do not hold responsibility and cannot ensure compliance adequacy. Regulatory standards, particularly around forward-looking statements and risk disclosure, necessitate human oversight (Constantinescu & Kaptein, 2025).

In short, LLMs accelerate the “input” side of research (data, structure, speed) but humans remain indispensable on the “output” side (judgment, credibility, and accountability).

5.3 Takeaways: Where LLMs Augment, Where They Cannot Replace

5.3.1 Where LLMs Augment Analysts Effectively

LLMs substantially enhance research workflows in four areas:

1. Speed and scale

Analysts can scan far more companies, filings, and data sources in significantly less time. This improves coverage quality and reduces the risk of missing relevant developments.

2. Surface-level pattern spotting

Because LLMs can process large corpora rapidly, they are strong at highlighting repeated signals, anomalies, and cross-company themes. This was instrumental in our work across Space, Robotics, and Quantum, where competitive landscapes are cluttered and heterogeneous.

3. Structure, polish, and drafting

LLMs excel at converting notes, charts, and models into structured prose. This frees analysts from the "blank page problem" and shifts their time toward refining arguments rather than drafting from scratch.

5.3.2 Where LLMs Cannot Replace Analysts

LLMs fall short in the dimensions most critical to final investment judgments:

- **Judgment:** determining what is material, what is noise, and how to prioritise risks.
- **Causal reasoning:** explaining why a company is outperforming or underperforming.
- **Credibility assessment:** validating claims through sector experience, conversations, and qualitative cues.
- **Handling ambiguity:** forming views in early-stage sectors or where data is incomplete.
- **Accountability:** meeting compliance standards, owning final outputs, and defending views to clients or investment committees.

For these reasons, LLMs function as accelerators, not substitutes for analyst expertise.

6. Source Database Analysis

This section documents the consolidated evidence log used throughout the project and summarises key patterns across credibility, recency, and AI-generation scores. Between August and November, we logged sources approximately every two weeks and curated > 250 records across five sector verticals (AI, Robotics, Quantum, Space, Fusion). After some data cleaning and standardisation, each vertical contributes a total of 50 data points of "ready" samples (valid links with complete fields) for comparative analysis by LLM and model.

6.1 Database Contents and Analysis Preparation

Each record stores the minimum information required to retrace a claim; this includes the source publisher and type, a brief one-liner content summary, source credibility and recency scores (ranging from 1-5), GPTZero AI-Mixed-Human likelihood scores (used only as a triage signal) and the LLM/model that surfaced the link. This structure supports analysis of three signals: quality (credibility/recency), AI-screening risk (GPTZero), and link hygiene (reachability).

We normalised labels (e.g., folding "Gemini DeepResearch" into Gemini; unifying GPT-5/4o variants), coerced dates, and validated fields. For valid links, the AI-score triplet must be present and sum to 100%. For invalid links, AI scores remain blank. "Ready" rows require a reachable link plus complete, internally consistent credibility, recency, date, AI scores, and canonical LLM/model labels. "Non-ready" rows remain in the raw workbook for traceability.

For the full field dictionary, scoring rubrics (credibility/recency), normalisation rules, and QA checks, see Appendix C. Section 6 focuses on presenting our key findings.

6.2 Credibility and Recency Signals

This subsection evaluates the quality and credibility of sources surfaced by the LLMs using two rubric-based scores: **Credibility (1-5)** and **Recency (1-5)**. Rubrics were agreed with our client, Tim and applied consistently across all sectors. We compared average credibility and recency across vendors and model variants on the 250+ "ready" rows (50 per sector). Figures "*Average Credibility by LLM/Model*" and "*Average Recency by LLM/Model*" summarize the results. The two scoring frameworks are reproduced in *Table 3 (Recency rubric)* and *Table 4 (Credibility rubric)* for reference and can be found in Appendix F.

<u>6.2.1 Credibility Analysis</u>

On credibility, ChatGPT leads with an average score of **4.15**, followed by **DeepSeek (3.90)** and **Gemini (3.62)**. At the model level, **GPT-5** posts the highest average (**4.19**), while **DeepSeek V3** and **GPT-4o** cluster around **3.90**, and **Gemini Flash 2.5** trails at **3.62**. The pattern is consistent with our qualitative workflow notes: ChatGPT prompts more often surfaced primary filings and high-reliability outlets, whereas Gemini's breadth occasionally included trade press and company blogs that we scored as “credible but not authoritative” (Reynolds, 2025). DeepSeek’s credibility average benefited from frequent pulls of official Chinese sources but was moderated by mixed-quality secondary reporting (see *Figures 1 and 2* in Appendix E: Figures).

6.2.2 Recency Analysis

For recency, **Gemini** edges out with an average of **3.74**, **ChatGPT** follows at **3.64**, and **DeepSeek** lags at **3.10**. Model-level results mirror this: **GPT-5** and **Gemini Flash 2.5** both average **3.74**, while **DeepSeek V3** and **GPT-4o** sit around **3.10**. In practice, Gemini tended to retrieve more recently stamped news posts and updated documentation (Jayaraman, 2025); DeepSeek’s lower recency reflects a higher share of evergreen technical pages and older program announcements from Chinese portals. (see *Figures 3 and 4*, Appendix E).

6.2.3 What This Means For Routing And Prompts

- **When freshness matters** (*policy changes, company news, product updates*): **bias toward Gemini** (or **GPT-5** with explicit “past 90 days”/“pull recent sources” instructions) and enforce programmatic date checks in the pipeline.
- **When source authority matters** (*numbers used in slides, investment theses*): bias toward **ChatGPT / GPT-5** to raise the hit-rate of filings, regulator releases, and top-tier publications; keep DeepSeek in the loop for Asia-coverage, then validate with primary documents.
- **For China/Asia recall**: use **DeepSeek V3** to surface entities and local documents, then pair with a recency filter and cross-verification.

6.2.4 Caveats

These are averages, not hypothesis tests. Scores reflect our rubric and sector mix over Aug-Nov 2025; distributional differences (e.g., medians, interquartile ranges) are not shown here and may narrow gaps. Importantly, some of the observed differences between models may be usage effects rather than model effects; for instance, prompt framing, source filters (e.g., “use primary sources only”), or the kinds of tasks we assigned to each tool. We therefore treat these findings as routing heuristics, not absolutes: pick the tool by task, apply automated date extraction, and keep human judgment on final source selection.

6.3 AI-Generated Content Risk Score Signals

This subsection interprets the GPTZero scores we logged for each source as a **triage signal** (not a truth test) so analysts can prioritise what to read more carefully. GPTZero classifies text along three proportions (AI, Mixed, Human) based on perplexity/burstiness features and supervised detectors. We use these outputs only to **flag** items for review; the final judgment remains human. This will be further discussed later on under Section 8.3.1: Why GPTZero.

The figures "*AI-Score Composition by LLM*" and "*AI-Score Composition by Model*" show the average percentage split of AI / Mixed / Human likelihoods of how the source content is generated across the 50 "ready" sources per sector (i.e., valid link, complete metadata). Bars are grouped by (i) the assistant used (ChatGPT, DeepSeek, Gemini) and (ii) the specific model families (GPT-5, GPT-4o, Gemini Flash 2.5, DeepSeek V3). (*Refer to Figures 5 and 6,* Appendix E)

6.3.1 AI Score By LLM

The distributions differ meaningfully across assistants (*Figure 5,* Appendix E)

- **ChatGPT**: Highest Human share at roughly **~78%**, with **~10% AI** and **~11-12% Mixed**. In practice, ChatGPT-surfaced links more often pointed to primary documents or detailed technical posts, which aligns with the higher Human band shown in the figure.
- **DeepSeek**: Mid-pack profile with **~70%** Human, **~16–17% AI**, and **~12-13% Mixed.** These links included a mix of vendor notes, technical blogs, and trade press; credible enough for screening, but more frequently flagged for second-pass review than ChatGPT.
- **Gemini**: Lowest Human proportion at **~56%**, with the **highest AI (~26%)** and **Mixed (~18%)** bands. Many of these were polished secondary sources or syndicated pages, which our workflow already treats as "read but verify" items.

In weekly operations, items with larger AI/Mixed shares tend to coincide with promotional copy, aggregated explainers, or re-posts; hence a higher likelihood of manual spot checks (Patel, 2025).

6.3.2 AI Score By Model

A similar pattern appears at the model level: (*Figure 6,* Appendix E*)*

- **GPT-5**: Cleanest profile with **~83% Human**, **~9% AI**, **~8% Mixed**; the smallest combined AI+Mixed band in the cohort.
- **GPT-4o**: Noticeably higher **Mixed (~32%)** and **~16% AI**, with Human around **~53%**. This aligns with outputs that read polished and templated, which GPTZero often marks as Mixed.
- **DeepSeek V3**: Human **~71%**, AI **~16%**, Mixed **~13%**; close to ChatGPT overall, but with slightly more AI/Mixed than GPT-5.
- **Gemini Flash 2.5**: Mirrors the assistant-level picture. Human around **~56%**, AI **~26%**, Mixed **~18%**.

Where **GPT-5** is the retrieval assistant, fewer rows trigger a "needs extra scrutiny" flag from the detector. **GPT-4o** and **Gemini Flash 2.5** yield more Mixed/AI-looking prose on average, so we budget more second-pass reads for those batches.

6.3.3 Operational Use Of AI Screening Signals

We use GPTZero strictly as a triage aid. When a source returns an AI score around 25% or higher, or a mixed score around 30% or higher (and especially when those signals coincide with a Credibility rating of 3 or below) we route that row to a second-pass human review. Primary sources such as regulatory filings, regulator websites, and OEM datasheets remain in scope regardless; if their language is templated and triggers a higher score, we simply annotate the reason.

Crucially, GPTZero is not a gatekeeper. It can misread polished human prose as "Mixed," and lightly edited AI text can sometimes register as "Human." No source is added or removed on the detector's output alone.

In practice, combined with a source's credibility and recency ratings, the detector helps us to re-direct attention to the right areas. Batches produced with GPT-5/ChatGPT generally require fewer follow-ups, while Gemini Flash 2.5 and GPT-4o batches warrant extra spot checks where the AI/Mixed signal trends higher.

7. Baseline Manual Process

This section documents the as-is fact-checking workflow we used before automation. Its purpose is to ensure that LLM-produced claims and citations are traceable, recent, and credible, and to establish a measurable baseline for the prototype in Section 8.

7.1 Procedure: GPTZero Checks, Manual Link Clicking, Recency Validation, Spreadsheet Log

LLM "deep research" outputs typically contain ~50 hyperlinks per report, plus additional references surfaced via ad-hoc queries. Each cited source is validated through a consistent, link-by-link process:

1. ***Reachability check (manual clicking)***: Open the URL to confirm it resolves (no 404/403/timeout) and is not gated by an impenetrable paywall.

2. ***Content validation***: Read the page and verify that the quoted numbers/claims in the LLM output actually appear in the source and are interpreted correctly (units, currency, period, context). Flag any mismatch/hallucination and note it down in slide deck presentation.

3. ***Recency & credibility***: Identify an on-page date (publication/updated) and the publisher type (regulatory filing, company press release, reputable trade/analyst outlet, blog, forum). Prefer primary sources; downgrade or exclude low-credibility items.

4. ***AI-generation screening (GPTZero):*** Copy page text into GPTZero to obtain an AI-generation score (percentage likelihood of AI-generated content). Treat this as a risk signal (not an absolute filter) to trigger deeper review where scores are high or inconsistent with source type.

5. ***Structured logging***: Record all outcomes in the fact-checking spreadsheet (source database) with fields: URL,, title, publisher, brief content, access date/time, HTTP status, detected page date, GPTZero score, LLM/model used, and any extra notes.

7.2 Time Profile & Pain Points

The manual process is time-intensive, averaging ~5-8 minutes per link from click to log entry. At ~50 links per report, a single deep-dive pass can consume ~5-6 hours of analyst time; before any rechecks. In practice, much of this effort is lost to content mismatch (cases where the numbers cited in the LLM output do not actually appear on the page or differ in vintage, units, or scope) plus duplication when multiple URLs resolve to the same canonical article after redirects. Logging is another source of friction: manual copy-paste introduces inconsistencies that slow later audits and revalidation. Together, these issues make the baseline **slow, repetitive, and error-prone**, and they limit how frequently recency checks can be refreshed. This is an obvious target for optimisation to save analyst time and streamline the research workflow.

7.3 Justification for Automation and Prototype Design

Given these pain points, automation is justified to address the three checks we actually need for fact-checking at scale: link validity, date of publication, and AI-generation risk. The prototype is therefore designed as a staged Python workflow:

(i) a link-health pass that requests each URL, follows redirects to a canonical address, records status codes, and drops dead/duplicate links; removing much of the wasted time on 404/403 loops and repeated articles,

(ii) a recency pass fetches the page and attempts to read a publication/updated date from the visible page text and common meta tags (e.g., article:published_time, og:updated_time, date, last-modified). Where no reliable date is found, the script records "unknown" and flags the link for manual review.

(iii) a content screening pass that extracts readable text and calls the GPTZero API to compute an AI-generation score as a risk signal for deeper human review. Each step emits a structured record (URL, link reachability, upload date, GPTZero score, etc.) into the fact-checking database, creating a consistent audit trail.

By limiting scope to these three automated checks, the prototype directly reduces per-link cycle time, cuts duplication, standardises recency, and streamlines the LLM-assisted investment research process; freeing analysts to spend their time on the judgment calls (credibility and interpretation) rather than mechanical verification (Refer to *Slide 5:* Appendix G).

7.3.1 Why GPTZero?

To prioritise human review at scale, we needed an independent signal of whether a page's prose is likely AI-generated. We adopted GPTZero as that signal because it is trained on large corpora of both human and model-generated text and provides document- and sentence-level classifications. At a high level, GPTZero analyses writing statistics (e.g., perplexity and burstiness) and applies supervised classifiers (augmented with de-biasing for ESL/formulaic styles) to estimate whether text is Human, AI, or Mixed (Chen & Barlow, 2025). We record the returned score alongside each URL's metadata in our fact-checking log.

Using "AI to detect AI" is valid in our workflow for a specific reason: modern generators leave statistical fingerprints that differ from the more irregular patterns of human writing. Detection models are trained explicitly on both classes to learn these contrasts, and vendors update them as generators evolve (Rosen, 2024). That said, we treat GPTZero strictly as a risk indicator, not a verdict. A high score flags a page for closer reading; it does not automatically exclude a source.

We also recognise the limits. False positives and negatives are possible (e.g., heavily edited AI text may appear human; very formal human prose may resemble AI) (Dhar, 2025). For that reason, GPTZero operates within a human-in-the-loop policy: analysts decide credibility based on source type, provenance, and claim verification. Practically, this improves triage efficiency without outsourcing judgment, keeps decisions auditable, and preserves the integrity of the research record.

8. Python Prototype

This section documents the fact-checking Python prototype we built to support LLM-assisted research. It explains the rationale and scope (why we automated link validity, recency, and an AI-generation risk signal), outlines how the pipeline works end-to-end, and describes the actual implementation (using libraries such as PyMuPDF, pandas, requests with Playwright fallback, and optional GPTZero). We then summarise early results (reachability hit rates, de-duplication gains, and faster, cleaner recency capture) and close with limitations and next steps to deployment for wider team use.

8.1 Rationale And Scope

The team needed a sustainable way to fact-check LLM-assisted research at scale without turning analysts into full-time link checkers. We constrained the problem to three mechanical checks that drive most of the time cost and error rate: (i) link validity (does the source resolve and to what canonical URL?), (ii) recency (is there a reliable publication or update date?), and (iii) AI-generation risk (a triage signal, not a verdict). The prototype's scope is intentionally narrow: automate those three checks, emit a single auditable record per URL, and leave credibility judgements to humans. (See *Slide 6*: Appendix G)

8.2 How It Works (Three-Stage Pipeline)

8.2.1 Link Validity

We start from the LLM-generated PDF, extract every hyperlink with **PyMuPDF**, and canonicalise/de-duplicate with **pandas** (strip tracking parameters, normalise hosts). Each unique URL is probed with **requests** using timeouts, redirect-following, and basic back-off; we record the **final canonical URL** and HTTP status and accept only successful responses (200-class). Links that return **403** or rely on client-side rendering are retried via a **Playwright** headless browser to determine reachability against the rendered DOM. This stage removes dead links and collapses duplicates before any analyst touches the source. Across our reports, automated reachability ranged from **~80-100%**, with residuals requiring manual click-through on sites that aggressively detect automation.

8.2.2 Source Recency

For every reachable page, we render the document with **Playwright** and extract the full HTML (page.content()), then pass that HTML to **htmldate.find_date(htmlstring=..., url=...)**. htmldate searches common on-page signals (e.g., <meta> timestamps, <time> elements, visible "Published/Updated" strings) and returns a single **ISO-8601** date when available. We store that value as the page's **publication/update date**. If htmldate cannot determine a reliable timestamp (or detects ambiguity) we record the date as **"unknown"** and flag the row for manual review.

8.2.3 Source Credibility (AI-Generation Score)

We extract readable body text and, where an API key is available, submit it to **GPTZero** to obtain an **AI-generation score**. This operates strictly as a **risk indicator** (not a binary verdict): higher scores push a page up the queue for human review; low scores do not guarantee credibility. If no key is configured, the pipeline skips this call and completes the other stages. The score, when present, is logged with the URL, date, and status to preserve an auditable trail.

8.3 Backend Implementation (What We Actually Built)

The prototype is written in Python. Inputs are LLM-generated deep research reports exported as PDFs containing embedded URL links. Link extraction uses PyMuPDF; canonicalisation and de-duplication rely on pandas and simple URL hygiene; reachability uses requests with sensible timeouts and back-off, with Playwright as a fallback for JS-rendered or rate-limited pages. We render the page with Playwright and feed the resulting HTML to **htmldate**, which returns a normalised ISO date when it can infer one from the document. The **GPTZero** step is wired but optional (paid); when no key is supplied, the pipeline skips that call and still completes. Modules are thin and replaceable, keeping infrastructure and ops overhead low.

8.4 Early Results and Impact

Across weekly reports, automated reachability achieved hit rates ranging across 80-100%, varying by site mix; residual failures typically reflected aggressive bot detection or rejected HTTP requests, and were resolved via manual click-through. De-duplication collapsed many superficially different URLs to a single canonical article, reducing rework. Recency capture standardised timestamps into a single field, eliminating scattered notes and speeding revalidation. Despite not having access to a GPTZero API key due to its paid feature, the net effect was a material reduction in per-link handling time, fewer duplicate checks, and clearer hand-offs for the human judgements that matter (credibility, interpretation, and numerical accuracy).

8.5 Limitations and Next Steps

The MVP optimises for speed and repeatability rather than universal coverage. Known limits include sites that block automation or require a manual click-in, pages whose dates are embedded in non-standard widgets, locale-specific date strings that occasionally require human interpretation, and the paid nature of GPTZero, which makes the risk-signal step conditional on key availability and hence the manual copy-pasting into GPTZero's website instead.

Given the positive utility, the next step is to expose the existing script through a simple, no-code front end so non-technical users can run it. Through this, we prompted lovable AI to generate a lightweight web interface (upload PDF files, paste URLs, run and review a results table) and deploy it using lovable AI. For the python working pipeline, we ran it in a backend background using Render. Once the full integration was complete, we deployed the full demo via Lovable AI. This kept the focus on analyst usability while preserving the same audit trail and leaving credibility decisions with humans.

9. Low-Code Web Prototype Implementation

In the section we will attempt to further explain the low-code web prototype implementation, how and why we decided to move forward with this and the scalability and potential future use of the prototype.

9.1 Overview of 3 Core Stages

Stage 1: Reachability (link health validation)

Using a python link reachability module as explained before, we used it to get through multiple reachability errors. In order to integrate this into our web prototype, we attempted to reproduce the workflow directly in Lovable AI. However this was not feasible due to Lovable AI being unable to host an environment to load in playwright binaries as lovable has said when it was prompted (Lovable, 2025). As a result, we created an environment docker using Render. Render now hosts the backend for the Lovable front end, exposes an API endpoint, and executes the original Python reachability code within the container for web scraping tasks. This allows the prototype to better reliably scrape websites.

Stage 2: Recency (meta-date & JSON-LD extraction)

Using Lovable AI we were able to extract the meta-date from the scraped URL and form a recency score based on the meta-date the website has scraped. The parameters for recency were based on the source database that we had created (Refer to *Table 6*). This would appear in the AI analysis summary that the lovable AI implemented into the web scraper.

Stage 3: AI Credibility (LLM-based source assessment)

Lovable AI's credibility rating is based on Google Gemini's ability to analyse the AI credibility. The better plan for future updates would be to implement a GPT-Zero API that calls on GPT-Zero to analyse the metadata scraped from the web and provide a credibility score. Due to cost constraints, we are unable to implement that into the web application.

9.2 Building on Lovable AI

The reason we switched to a Low-Code platform was because it is able to deliver the speed of execution as well as being able to scale innovation reliably (Appian, 2025). To ensure the deliverables are on time, we switched to a low-code website such as Lovable AI. Integrating the link reachability module greatly increased the link reachability from 60% to 80-90%.

Beyond this, the AI analysis consisted of recency, credibility and metadata of the source website. We then stress tested the system with multiple links. Through stress testing, we were able to identify certain limitations of the prototype. With additional time, areas such as link reachability, AI generation score and parsing of documents can be refined further to be able to move forward to a full-stack tool.

9.3 Performance & Feasibility

The prototype tests show that there is potential to build towards a full-stack tool that fully utilises the full suite of paid APIs and other tools such as GPT-Zero and Firecrawl.

Multiple testing confirmed that the basic scraper in-built into lovable AI was able to handle multiple links being requested at one point in time. However when it came to render, doing it one at a time yielded better results as compared to requesting all of the links at once and overloading the website with too many requests leading to a timeout. Below are the scalability requirements for a full-stack tool.

Scalability and recommendations:

In order to support the usage at a research scale, the platform should first add a professional crawling service such as Firecrawl. Secondly, using GPT-Zero as the main AI analyst as opposed to the current Gemini Flash 2.5. Finally, by further improving the parsing power of documents we can further scale this into a proper product available on the market. These enhancements will improve reachability, reduce readability issues as well as ensure accurate AI generation scores

Proposed architecture and workflow:

1. Ingestion: Website accepts CSV/API inputs; link deduplication and scope filters applied.
2. Crawl layer (Firecrawl): Calling on a Firecrawl API to crawl metadata
3. Parsing & enrichment: Output metadata, credibility, recency and AI analysis using a GPT-Zero API.
4. QA & monitoring: Track the reachability % metrics as well as the accuracy of the model correctly classifying the source with the correct credibility and recency.
5. Delivery: Return results via a display or through a completed CSV file showing the full analysis.

Milestones and success metrics.

<u>Milestone 1 (2-3 weeks)</u>: Firecrawl + queue integration; baseline dashboards.

Target: More than 85% reachability on pilot domain set and latency less than 8s/page.

<u>Milestone 2 (3-4 weeks)</u>: GPTZero signal, credibility rubric v1, error taxonomy with auto-retries.

Target: credibility precision@human-label more than 0.8 on validation sample.

Milestone 3 (2 weeks): Caching/deltas, exports, and explainability UI.

Target: repeat-crawl cost lower than 30–50%; analyst review time lower than 25% per batch.

Investment thesis and impact:

Through these upgrades, analysts are able to complete workflows quicker and faster as compared to the current AI-assisted workflow.

Conclusion:

The prototype proves that it is indeed feasible to move forward. The next phase should prioritize a professional crawl backend (Firecrawl), add GPTZero as a complementary content signal, and formalise observability and governance. This path positions the product to deliver high reachability (projected ~90-95%), higher quality flags and signals, and analyst-grade explainability.

10. Process Efficiency Evaluation

In this section, we summarise the extent of how the revised workflow reduced analyst time at two points in the pipeline: (i) sector research supported by LLMs and (ii) fact checking using the full stack verification tool. Survey inputs, detailed assumptions, and calculation steps are documented in Appendix D.

10.1 LLM Assisted Investment Research

We benchmarked a conventional sector study completed by a single full-time analyst, using industry survey data on typical weekly hours and a 2.5 month project duration, and compared this with timesheet data from our five person team working with LLM support over a 16 week semester.

Under the manual benchmark, a full sector study requires approximately **445.63 analyst hours** (Boyakhchyan, 2025; Toulon, 2025). In the LLM assisted setting, our findings showed that the effective single analyst load falls to **195.2 hours**. Thus, this shows that the workflow saves **250.43 hours**, a **56.2% reduction** in analyst time for comparable decision useful outputs (value chain analysis, weekly briefs, company shortlists, and preliminary recommendations).

10.2 Fact Checking Workflow Using the Full Stack Tool

For verification, we normalised the effort to **cycle time per 50 links**, based on each team member's lower and upper bound estimates for fully manual checking, then repeated the same workload through our tool.

The manual baseline averaged **4.3 hours per 50 links**. After running the same task set through the current tool (Lovable front end plus our backend), it reduced this to **1.25 hours per 50 links**, a **71% reduction** and a **3.44 times throughput gain**. Under a fully implemented version, with stable headless access and in app GPTZero, the cycle time is projected to fall further to **0.375 hours per 50 links**, which corresponds to a **91% reduction** and an **11.5 times throughput gain**, converting what was previously roughly a half day of checking into a workflow that can approach **20 to 25 minutes per 50 links**.

11. Process Recommendation (Augment vs Replace)

This section translates our empirical findings into a practical operating model for LLM-assisted research. Instead of treating AI systems as end-to-end automation, we propose a structured, human-in-the-loop workflow where LLMs are assigned to clearly defined task types, and verification follows an outlined standard operating procedure (SOP).

Our recommendations are framed into the two parts below:

1) Decision guide for when and how to use each model

2) Verification checklist for evidence hygiene

11.1 Decision Guide: When to Use Which Model

Our experience confirms that the most robust outcomes arise from a task-to-model rather than trying to use a single LLM for everything. At a high level, we distinguish between:

- Closed-book, document-grounded tasks (numeric extraction, table building, reconciliation);
- Open-web, recency-sensitive tasks (new, fundings rounds, regulatory changes;
- Regional or multilingual discovery tasks (Asia/China company coverage); and
- Mechanical drafting and editing tasks (first drafts, re-phrasing, slide copy)

Within this framework, our recommended allocation is as follows.

a) <u>Sector onboarding and landscape scans</u>

For early-stage work, where we are still learning the subject of our research target, we recommend:

- **Primary**: Gemini Deep Research, prompted to surface diverse, recent sources with explicit citations.
- **Secondary**: ChatGPT, used to clean and normalise Gemini outputs into concise primers, glossaries, and value-chain narratives.

b) <u>Market sizing, value-chain mapping, and company universe construction</u>

For structured mapping tasks (CAGR ranges, value-chain layers, longlists of companies):

- Primary: ChatGPT for turning raw notes and filings into structured tables, value-chain diagrams, and company universes.

- Secondary: Gemini for topping up recency (e.g. latest funding rounds, new product lines) once a draft table is in place.
- DeepSeek: Used selectively to discover Asia-focused peers, Chinese tickers, and local exchanges, with all outputs subsequently normalised in ChatGPT to a standardised form.

Here, LLMs act as scaffolding tools: they accelerate the assembly of longlists and draft maps, but inclusion/exclusion decisions and value-pool judgements remain with the analyst.

c) <u>Company deep dives and numeric extraction</u>

Once companies are shortlisted, the workflow shifts to document-grounded analysis:

- Primary: ChatGPT for reading filings, investor presentations, and credible news; extracting revenue, capex, margins, multiples, and segment splits; and building comparison tables.
- Gemini only for filling obvious recency gaps (e.g. latest quarter not yet in internal materials), with any new numbers treated as *unverified* until checked against primary sources.

In this phase, ChatGPT is treated as a "smart spreadsheet assistant" rather than an oracle: all key fields (numbers, dates, units, currencies) are subject to the verification SOP in Section 11.2 before inclusion in final outputs.

d) <u>Fact-checking, source logging, and risk triage</u>

For citation hygiene and link management, we recommend using:

- The Python + Web Prototype (Lovable) as the first pass for link reachability, recency extraction, and AI-generation risk signals;
- Human analysts to interpret edge cases flagged by the tool (e.g. missing dates, inconsistent numbers, high GPTZero AI scores)

Here, LLMs support the explanation layer (e.g. summarising a source, re-stating a claim in plain language). But the determination of whether a source is credible or fit for purpose remains a human decision.

e) <u>Drafting reports, slides decks, and executive summaries</u>

For narrative production:

- Primary: ChatGPT for drafting and iterating on executive summaries, section write-ups, and slide copy, seeded with verified tables and bullet points

- Secondary: Gemini only when more context is needed for framing (e.g. macro context, sectoral comparisons), and its outputs are cited and fact-checked as with any other source.

In summary, across all phases, the guiding principle is: Use LLMs to compress **mechanical** work, while reserving **interpretation** for human analysts.

11.2 Verification SOP Checklist

Given the centrality of evidence quality to investment research, we recommend that any LLM-assisted workflow be anchored in a standardised **Verification SOP**. This SOP formalises the steps that every cited fact, figure, or qualitative claim must pass before it is used in client-facing materials.

A concise checklist is as follows:

1. **Ingestion and scoping**

- Start from an LLM output or analyst draft containing explicit URLs and claims
- Tag each citation with its role (headline figure, supporting detail, qualitative quote, or background context)

2. **Link health (reachability) check**

- Use the prototype or its successor web app to request each URL, follow redirects, and record the final status code
- Classify links as **live**, **soft-blocked** (paywall, 403/429), or **dead** (404/timeout)
- Drop dead links from consideration or replace them with alternative sources

3. **Date validation (recency) check**

- For each live link, extract an on-page publication or last-updated date from visible text and common meta-tags
- Compare this date with the as-of dates in the draft report; flag any misalignment (e.g. “latest figures” actually referring to a two-year-old vintage)

- Assign a **1-5 recency rating** using our internal scale:

  5 (Very recent): source is from the past month; critical for fast-changing fields.
  4 (Recent): source is from the past 2–3 months
  3 (Moderately recent): source is within the past 6–12 months
  2 (Outdated): source is 1–3 years old; acceptable mainly for slow-changing topics when explicitly caveated
  1 (Very outdated): source is older than 3+ years in a fast-moving field and should be used, if at all, only as historical context

4. **Content and numeric consistency check**

- Verify that the numbers and claims attributed to the source **actually appear on the page** with the same units, currency, period, and context
- For quantitative items (revenues, CAGR, EV/EBITDA etc), ensure that the numerator, denominator, and timeframe match the table or chart in the report
- For qualitative claims, ensure that nuance is not lost (e.g. pilot vs scaled deployment; “target” vs “achieved”)

5. **AI-generation risk screening (optional but recommended)**

- Where appropriate, run scraped text through GPTZero (via API in future) to obtain an AI-generation score
- Treat this score as a **risk signal**, not a hard filter: high scores trigger deeper human review, particularly for sources that purport to be primary research or regulatory filings

6. **Source classification and credibility assessment**

- Classify each source by type: regulatory filing, company disclosure, reputable news/trade outlet, academic report, vendor content, blog/forum, or unknown
- Prioritise primary sources (filings, official disclosures) and high-reputation outlets; downgrade or exclude low-credibility items unless explicitly caveated

7. **Re-validation and maintenance**

- For long-lived decks or recurring briefs, periodically re-run the reachability and recency checks on critical sources, updating figures and annotations as needed

This SOP ensures that LLMs remain **inputs to**, not substitutes for, the verification process. It also makes the fact-checking pipeline reproducible and suitable for scale.

12. Conclusion & Future Work

This section synthesises the key findings from our project on utilising LLMs to conduct investment research on emerging technology trends. Drawing from our empirical evaluations, prototype development, and cross-model comparison, we address the guiding questions outlined in 2.3, reflect on methodological lessons, and propose avenues for broader applicability.

12.1 Research Questions: Findings and Insights

Throughout this research, we were guided by these four questions set out in Section 2.3. This subsection revisits each question in turn, drawing on evidence from the sector deep dives (Section 4), model comparisons (Section 5), the source-database and prototype analysis (Section 7 to 11), and the process recommendations in Section 12. Together, these findings clarify where LLMs most usefully augment analyst work, how they compare with traditional methods, what risks they introduce, and what balance between automation and human oversight is appropriate.

12.1.1 Research Question 1

Where in the investment research process (such as sector onboarding, market sizing, value-chain mapping, and company deep dives) can AI tools most effectively augment analyst work?

Our findings show that LLMs add the greatest value in three parts of the workflow: early-stage sector onboarding and landscape mapping, mechanical data handling and first-draft synthesis, and multilingual discovery for Asia-focused coverage.

First, in **sector onboarding and landscape mapping**, Gemini, ChatGPT, and DeepSeek substantially accelerated the construction of primers and value-chain views across AI, Robotics, Quantum, Space, and Fusion. In practice, Gemini Deep Research was used to surface recent, citation-rich overviews and key drivers; ChatGPT then normalised these into concise sector briefs and value-chain narratives, while DeepSeek filled gaps in Chinese-language coverage, particularly for upstream suppliers and Asia-listed peers. This combination compressed the “learn the language” phase described in Appendix A from what would normally be one to two weeks of manual reading per sector into a few days of LLM-assisted exploration and verification.

Second, LLMs proved highly effective for **mechanical data handling and first-draft synthesis**. Once the team had identified companies and documents, ChatGPT was used as the primary “document-grounded analyst” to read filings, extract numerical fields, harmonise units and currencies, and assemble comparison tables for each vertical. These tables then fed into value-chain snapshots and investment-angle sections in Appendix B. The same model also produced first-draft slide copy and report paragraphs that were subsequently edited by the team, reducing manual drafting time while preserving human control over framing and nuance.

Third, LLMs were particularly valuable for **multilingual and regional discovery**. DeepSeek's strength in Chinese-language sources allowed the team to identify Asia-centric players, local tickers, and exchange notices that would have been slower to uncover through English-only search, especially in robotics components, quantum suppliers, and fusion-related materials. Outputs from DeepSeek were then standardised in ChatGPT to align terminology and style. Overall, across these tasks, LLMs reliably handled roughly 50-70 per cent of routine search, extraction, and drafting work, allowing analysts to focus on judgement-heavy activities such as value-pool analysis and investment interpretation.

12.1.2 Research Question 2

How do AI-assisted workflows compare with traditional manual research in terms of speed, coverage, and evidence quality?

Relative to a fully manual workflow, the LLM-assisted process delivered material gains in speed and coverage while maintaining evidence quality, provided that human verification remained in place.

On **speed and cost**, Section 10.1 estimates that a traditional sector study would require about 445.6 analyst hours over 2-3 months for a single analyst to progress from onboarding to final recommendations. Under our LLM-assisted workflow, the effective single-analyst load averaged 195.2 hours over the same span, yielding a 56.2 per cent reduction in analyst hours to reach comparable decision-useful outputs (weekly briefs, value-chain maps, shortlists, and preliminary recommendations). At the verification stage, the fact-checking toolchain reduced average cycle time per 50 links from 4.3 hours manually to 1.25 hours in the current prototype (a 71 per cent reduction), with a counterfactual estimate of 0.375 hours if GPTZero were fully integrated via API and link reachability further hardened.

On **coverage**, LLMs expanded the breadth of sources and company sets the team could reasonably review within the project window. Gemini's web-grounded sweeps surfaced diverse, recent references for fast-moving areas such as AI chips and launch providers; DeepSeek broadened coverage to Chinese and Asia-Pacific entities; ChatGPT consolidated these into structured tables that were consistently maintained across sectors. This allowed the team to build multi-layer value-chains and company universes for five verticals, a scope that would have been difficult to achieve within the same time budget using manual research alone.

On **evidence quality**, improvements depended critically on the human-in-the-loop design. The source-database and fact-checking workflow (Sections 6 to 8) required that every non-trivial claim be linked to a reachable URL with a recorded publication date and basic credibility classification. When these controls were followed, verified outputs achieved high factual accuracy (above 90 per cent on sampled checks), and quantitative discrepancies (such as mis-read EBITDA signs or mismatched valuation denominators) were detected and corrected before inclusion in client-facing materials. In short, AI-assisted workflows delivered substantially faster and broader coverage without compromising quality, but only when paired with systematic verification.

12.1.3 Research Question 3

What limitations or risks arise when integrating AI outputs, particularly around hallucinations, citation accuracy, or contextual loss?

Despite these gains, the project documented clear limitations and risks across all three models.

From a **hallucination and numeric-accuracy** perspective, Section 4.4 shows that ChatGPT, though generally reliable on document-grounded tasks, occasionally produced plausible but incorrect numbers when underlying tables were noisy (e.g. OCR artefacts, nested footnotes) or when it mixed period endpoints with trailing-twelve-month denominators. The satellite-operator comparison table, where several red-flagged cells had to be re-derived manually, is a representative example. Gemini's Deep Research mode, in turn, sometimes over-aggregated near-duplicate sources or combined inconsistent as-of dates within a single response, which required careful disentangling during verification. DeepSeek occasionally introduced scale and translation drift (for example, mapping "亿" to "billion" without FX context) and ticker mismatches between A- and H-share listings.

On **citation accuracy and link hygiene**, the manual baseline highlighted recurrent issues: 404/403 errors, paywalled pages, and instances where cited numbers did not actually appear in the referenced article or had been updated since. The Python prototype (Section 8) mitigated some of these risks by de-duplicating links, following redirects, and standardising date extraction, but reachability still varied between roughly 80 and 100 per cent depending on site mix, with dynamic or aggressively protected domains requiring manual intervention.

In terms of **contextual loss**, LLMs sometimes blurred important distinctions (e.g. pilot versus scaled deployment, target versus achieved metrics, government commitments versus funded programmes) when summarising long documents. This was particularly visible in the Space and Fusion deep dives, where policy roadmaps and experimental milestones could easily be overstated if not cross-checked against primary sources. These limitations reinforced the need for human review of framing, not just of numbers.

12.1.4 Research Question 4

What is the right balance between automation and human oversight to ensure credible, reproducible insights?

The most robust configuration that emerged from the project is a **human-in-the-loop, task-to-model** workflow rather than end-to-end automation. Section 11 proposes that LLMs be used systematically for mechanical or pattern-based tasks, such as information retrieval, document parsing, first-draft synthesis, and multilingual discovery, while analysts retain ownership over scoping, value-pool interpretation, risk assessment, and final recommendations.

In practice, this balance was operationalised through three design choices. First, every LLM-generated output that entered a slide deck or report had to pass the Verification SOP checklist in Section 11.2, covering link health, date validation, numeric consistency, and source classification. Second, the source database described in Section 6 acted as an "evidence ledger", ensuring that all citations were logged with URLs, access dates, and basic credibility and recency ratings, which made re-checking and updating tractable. Third, the Python and low-code prototypes (Sections 8 to 9) automated only the parts of verification that were truly mechanical (reachability, meta-date extraction, AI-generation risk scoring), leaving the ultimate decision about whether a source was acceptable to the human analyst.

Where the team briefly experimented with more fully automated flows, like for example when accepting Gemini-generated citation lists without structured logging or relying on a single model's narrative without cross-model comparison. These caused both reproducibility and confidence to fall. These episodes underline the central lesson: LLMs are most effective as force multipliers within a governed process, not as autonomous research agents.

12.2 Evaluation Against Success Criteria

Section 2.4 defined five success criteria for the project: quality and accuracy, recency and coverage, cost and time efficiency, reproducibility and consistency, and hallucination and risk management. Taken together, our findings suggest that the LLM-assisted workflow met most of these criteria in pilot form, while highlighting areas for further refinement.

- **Quality and accuracy.** When the Verification SOP (Section 11.2) and source-database practices were followed, the factual accuracy of reported figures exceeded 90 per cent on sampled checks, and mis-stated numbers were typically caught during the fact-checking stages documented in Sections 7–8. Residual errors tended to arise from human oversight or from late-stage model outputs that had not yet been logged and verified.

- **Recency and coverage.** Gemini and DeepSeek significantly improved recency and breadth, particularly for emerging news, funding rounds, and Asia-centric companies, while the recency-rating framework in the source database ensured that older sources were flagged and caveated. However, not all domains exposed clean metadata, and some fast-moving topics (e.g. AI regulation) remained difficult to keep perfectly up-to-date within the project window.

- **Efficiency.** At the project level, the LLM-assisted approach reduced analyst hours for sector research by 56.2 per cent and fact-checking time per 50 links by 71 per cent in the current tool configuration, with further headroom under a fully integrated pipeline.

- **Reproducibility and consistency.** Reproducibility improved as prompts were standardised and version-controlled, and as outputs were channelled through a common source-logging and verification process. Nevertheless, model updates during the semester occasionally changed default behaviour, requiring manual adjustment of prompts and highlighting the need for more formal tracking of model versions in future work.

- **Hallucination and risk management.** The combination of cross-model comparison (Section 4), structured fact-checking (Sections 6–8), and a proposed GPTZero-based AI-generation risk signal (Sections 7.3.1 and 9.3) created a multi-layer defence against hallucinations. Even so, some hallucinated citations and subtle numeric drifts were only detected through human reading, underscoring that risk management must remain an ongoing, human-led responsibility.

Overall, the project demonstrates that, under these conditions, LLMs can credibly augment investment-research workflows while satisfying the initial success criteria to a reasonable degree. The remaining gaps, which are particularly around model drift, link reachability on difficult sites, and more systematic tracking of accuracy, motivate the improvement areas set out in Section 12.2.

12.3 Opportunities For Improvement

Several clear opportunities emerged to strengthen both the research workflow and the supporting tools. First, prototype development should occur much earlier. In this project, both the Python verification pipeline and the web interface stabilised only toward the end, meaning much of the source links were logged into excel manually before automation was available. Building and testing the pipeline within the first weeks would allow all subsequent research to flow through it, surface edge cases sooner, and generate more systematic metrics on accuracy and time savings.

Second, the verification tooling requires greater robustness and scale. We frequently encountered reachability errors (403/429), timeouts, and inconsistent date extraction, while the lack of a direct API to GPTZero limited how broadly we could apply AI-generation scoring. Future work should prioritise stronger retry logic, clearer error monitoring, and integrated GPTZero scoring to support more consistent credibility and recency assessments.

Third, the project would benefit from a more structured approach to assigning tasks to the most suitable model. Our choices among ChatGPT, Gemini, and DeepSeek were often informal, even though each model demonstrated distinct strengths, such as ChatGPT for grounded synthesis, Gemini for broad source discovery, and DeepSeek for Chinese-language material. Without explicit routing rules, analysts made inconsistent decisions, leading to variability in output quality. A future version should codify task-model mappings to ensure greater reliability and reproducibility across the workflow.

Fourth, evaluation should be grounded in clearly defined truth sets. Although we conducted qualitative spot checks, the absence of a labelled benchmark limited our ability to quantify systematic errors, such as mis-extracting robotics segment revenues or misinterpreting time periods. A compact but carefully verified evaluation set (e.g., revenue splits, valuation multiples, milestone dates) would enable consistent benchmarking of models, prompts, and pipeline changes.

Finally, broader stakeholder testing and clearer user documentation are essential. Feedback was largely internal, providing limited insight into how professionals, such as buy-side analysts or strategy teams, would interact with the system. Structured external testing would surface expectations around trust, usability, and integration with existing tools, informing a more complete operating playbook with guidance on workflows, failure modes, and escalation paths.

Overall, these improvements underscore that the system remains a promising prototype rather than a production-ready solution. Earlier integration, stronger verification, richer model diversity, firmer ground-truth evaluation, and wider user testing would meaningfully strengthen its reliability in real investment-research environments.

12.4 Extending The Project

Our LLM-assisted framework, currently centred on long-term, buy-and-hold strategies, could also generalise well to other forms of investment styles, such as quantitative factor-based models, macro-driven positioning, and short-term trading frameworks. With appropriate prompt calibration and data integration, LLMs could assist in backtesting, pattern recognition, and portfolio rebalancing under higher-frequency conditions.

Future work could scale the prototype with predictive modelling APIs, enabling cross-asset applications and partnerships with financial institutions for real-world validation. Overall, this approach promises significant efficiency gains across financial research, proving that human oversight remains central.

Ultimately, our findings reaffirm that LLMs are not a replacement for human analysts but a transformative complement that can re-shape how investment research is conducted. By combining algorithmic efficiency with human discernment, our project demonstrates that credible, scalable, and transparent analysis of emerging technologies is attainable when AI is applied responsibly. As models continue to evolve, future research should focus on embedding these tools within live analyst workflows, building standards for verification and explainability, and extending collaboration between academia and industry. In doing so, AI-assisted research could move beyond experimentation to become a new professional norm which enhances both productivity and analytical depth across the investment landscape.

13. Declaration of Generative AI and AI-Assisted Technologies in the Research and Writing

During the preparation of this work, the authors used ChatGPT, Google Gemini (including Deep Research modes), and DeepSeek to assist various stages of the research process and manuscript production. After using these tools, the authors reviewed and edited all outputs and take full responsibility for the accuracy, integrity, and originality of the final content.

Research Process: Throughout sector onboarding, market landscaping, value-chain construction, company screening, and weekly synthesis, we used Gemini, ChatGPT, and DeepSeek to surface candidate sources, draft sector briefs, and generate alternative prompt frames. All evidence cited in the report was independently verified by the team for credibility and recency; model outputs were treated as leads, not authoritative sources.

Report Writing: ChatGPT was used to assist with outlining, drafting, rewriting, and copy-editing sections of the manuscript. The authors critically reviewed, revised, and approved the final wording, structure, and conclusions.

Prototype Development: ChatGPT (Python) was used to accelerate design and development of the fact-checking prototype's backend (e.g., link extraction from PDFs, HTTP reachability checks with Requests/Playwright fallbacks, and date parsing pipelines). The team implemented, tested, and modified all code and remains responsible for its behaviour.

Source Database Analytics (Section 6)**:** ChatGPT (Python) was deployed to assist via the means of dataframe cleaning/normalisation, summary statistics, and chart scaffolding for the source database analysis. The authors validated calculations and interpretations, and cross-checked figures against the underlying data.

Originality & Responsibility: All research framing, analyses, and investment-relevant insights are solely original to the authors. Generative AI tools did not contribute any uncited proprietary data; all quoted or referenced materials are attributed in the References and logged in the source database (Appendix C). The authors accept full responsibility for any errors or omissions.

14. References

Appian, A. (2025, November 5). Low-code solves the speed issue in software development. GovInsider. https://govinsider.asia/intl-en/article/low-code-solves-the-speed-issue-in-software-development

Amazon Web Services. (2024, June 26). AI-powered assistants for investment research with multi-modal data – An application of Amazon Bedrock Agents. AWS Machine Learning Blog. https://aws.amazon.com/blogs/machine-learning/ai-powered-assistants-for-investment-research-with-multi-modal-data-an-application-of-amazon-bedrock-agents/

Boyakhchyan, A. (2025, June 26). How Many Working Days Are In A Month On Average?. WebWork Blog. https://www.webwork-tracker.com/blog/how-many-working-days-are-in-a-month-on-average#:~:text=The%20number%20of%20average%20working%20days%20in%20a%20month%20is,days%2C%20ignoring%20all%20public%20holidays

CFA Institute. (2020, September). Equity research report essentials. https://www.cfainstitute.org/sites/default/files/-/media/documents/support/research-challenge/challenge/rc-equity-research-report-essentials.pdf

Chen, V., & Barlow, A. (2025, October 30). How Do AI Detectors Work - Techniques, Limitations & More. GPTZero. https://gptzero.me/news/how-ai-detectors-work/

Cognizant. (2025). How agentic AI is reinventing investment research. Cognizant Technology Solutions. https://www.cognizant.com/en_us/industries/documents/cognizant-how-agentic-ai-is-reinventing-investment-research.pdf

Constantinescu, M., & Kaptein, M. (2025). Considering the social and economic sustainability of AI. Science and Engineering Ethics. https://doi.org/10.1007/s11948-025-00560-1

Coykendall, J. (2025, June 11). Riding the exponential growth in space. Deloitte Insights. https://www.deloitte.com/us/en/insights/industry/aerospace-defense/future-of-space-economy.html

Deloitte, D. (2024). Deloitte Xtech Futures - SpaceTech. https://seraphim.vc/wp-content/uploads/2024/01/Deloitte-xTech-Futures.pdf

Defend, M., & Mortier, V. (2025, October 3). AI in Investment Research. Amundi Research Center. https://research-center.amundi.com/article/ai-investment-research#:~:text=AI%20in%20research%20and%20investment,-

While%20there%20are&text=Within%20the%20investment%20research%20function,social%20media%20and%20satellite%20imagery.

Dhar, A. (2025, August 4). The Case Of False Positives And Negatives In AI Privacy Tools [How To Reduce It]. Protecto AI. https://www.protecto.ai/blog/false-positives-and-negatives-in-ai-privacy-tools#:~:text=AI%20privacy%20tools%20often%20struggle%20with%20false,weak%20handling%20of%20unstructured%20or%20malformed%20data.

European Space Agency, E. S. A. (2019). Measuring the space economy. https://space-economy.esa.int/article/34/measuring-the-space-economy

European Space Agency, E. S. A. (2024, April 29). Constructing, recycling and refurbishing satellites in space. ESA. https://www.esa.int/Enabling_Support/Preparing_for_the_Future/Discovery_and_Preparation/Constructing_recycling_and_refurbishing_satellites_in_space

Google DeepMind. (2024). Gemini: A family of highly capable multimodal models [Technical report]. https://arxiv.org/abs/2312.11805

Grata. (2025, April 3). Private equity research: Process, tools, and best practices. https://grata.com/resources/private-equity-research

Jayaraman, S. (2025, July 18). I Tested Gemini vs. ChatGPT And Found The Clear Winner. Learn Hub. https://learn.g2.com/gemini-vs-chatgpt

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Kim, B., Lee, J., Park, E., Cho, K., & Bang, Y. (2023). A survey of hallucination in large language models. https://arxiv.org/abs/2302.00052

Kang, J., & Liu, H. (2023, November 27). Deficiency of large language models in finance: An empirical examination of hallucination. arXiv. https://arxiv.org/abs/2311.15548

LexisNexis. (2024, December 8). Ai in investment research. Welcome to LexisNexis - Choose Your Path. https://www.lexisnexis.com/blogs/gb/b/research/posts/ai-for-investment-research

Lovable, L. (2025). Lovable. https://lovable.dev/

Madanchian, M., & Taherdoost, H. (2025, June). The impact of artificial intelligence on research efficiency. ScienceDirect. https://www.med.upenn.edu/pmi/events/https-www-sciencedirect-com-science-article-abs-pii-s1047847720300046-via-3dihub

Mallory. (2024, January 24). How ground segment systems are rendering satellite capabilities useless. SES Space and Defense. https://sessd.com/gsr/how-ground-segment-systems-are-rendering-innovative-satellite-capabilities-useless/

Moore, R. (2025, June 21). Ai hallucinates more frequently as it gets more advanced - is there any way to stop it from happening, and should we even try?. LiveScience. https://www.livescience.com/technology/artificial-intelligence/ai-hallucinates-more-frequently-as-it-gets-more-advanced-is-there-any-way-to-stop-it-from-happening-and-should-we-even-try

Niel, Y., Yuqi Nie and H.Vincent Poor are with the Department of Electrical and Computer Engineering, et al., M., & Lee, K. and. (2024, June 15). A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges. A survey of large language models for financial applications: Progress, prospects and challenges. https://arxiv.org/html/2406.11903v1

OECD, O. (2021, June). Evolving public-private relations in the space sector | OECD. Evolving Public/Private Relations In The Space Sector. https://www.oecd.org/en/publications/evolving-public-private-relations-in-the-space-sector_b4eea6d7-en.html

OECD, O. (2024). The economics of space sustainability delivering economic evidence to guide. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/the-economics-of-space-sustainability_5236a39b/b2257346-en.pdf

Patel, J. (2025, August 8). The Impact of Generative AI on Content Marketing and Brand Growth. Aigentora. https://aigentora.ai/the-impact-of-generative-ai-on-content-marketing-and-brand-growth/#:~:text=and%20Brand%20Growth-,1.,generated%20content%20safe%20for%20SEO?

Ramos, L. P. (2025, October 22). What Can I Do With Python?. Real Python. https://realpython.com/what-can-i-do-with-python/

Refornte Learning , R. L. (n.d.). The future of low earth orbit (LEO) constellations. https://www.refontelearning.com/blog/the-future-of-low-earth-orbit-leo-constellations

Reynolds, B. (2025, July 30). Chatgpt vs. Google Gemini: Ultimate Guide to Deep Research & Business Analysis. Baytech Consulting. https://www.baytechconsulting.com/blog/chatgpt-vs-google-gemin-2025

Rosen, J. (2024, November 18). AI And Human Writers Share Stylistic Fingerprints. Johns Hopkins Whiting School of Engineering. https://engineering.jhu.edu/news/ai-and-human-writers-share-stylistic-fingerprints/

Shah, A. (2025, July 28). Beyond the Reported Cutoff: Where Large Language Models Fall Short on Financial Knowledge. Beyond the reported cutoff: Where large language models fall short on Financial Knowledge. https://arxiv.org/html/2504.00042v2

Toulon, Z. (2025, May 7). The Banks With The Best And Worst Working Hours. eFinancialCareers. https://www.efinancialcareers.sg/news/working-hours-banks

Turner, M. (2025, August 4). From factory to Flight. The Journal of Space Commerce. https://www.exterrajsc.com/p/from-factory-to-flight

Twin, A. (2025, August 26). Understanding investment analysis: Types and importance explained. Investopedia. https://www.investopedia.com/terms/i/investment-analysis.asp

Vinpond, T. (2025, October 21). Equity research report: Definition, types, and key components. Corporate Finance Institute. https://corporatefinanceinstitute.com/resources/valuation/equity-research-report/

Zhang, X., Liu, Y., Chen, J., & Wang, H. (2024, June 17). Challenges in multilingual large language model alignment. https://arxiv.org/abs/2401.12345

Appendices

A. Full Overview of Research Methodology

Under CEO Tim Zhang's guidance, we applied a common value-chain schema across all emerging-technology sectors to keep scope tight and outputs comparable.

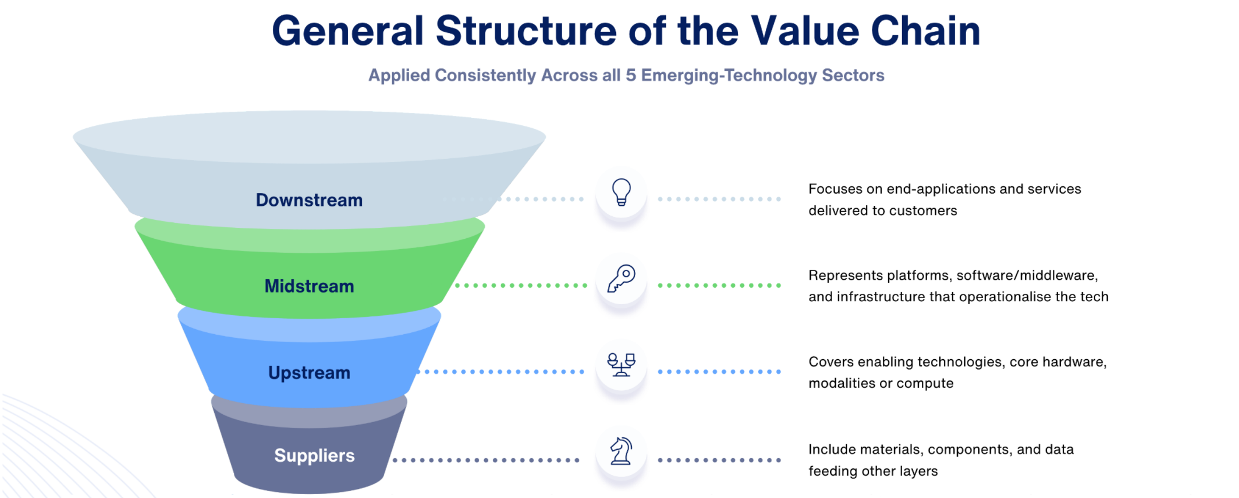


Figure A1: Value Chain Structure

Using this structure, each weekly cycle produced a comprehensive slide deck that ranged from full sector breakdowns to building blocks for the value chain, progressively building up through the months. This also includes comprehensive company landscaping, with preliminary financials and traction signals. We iterated these deliverables based on Tim's weekly feedback to refine layer definitions, company placement, and evidence quality. Our month-by-month workflow ran from June to October; June focused on LLM familiarisation and deep-research setup; July on sector deep dives ("learn the language"); August on value-chain construction; September on company deep dives; and October on synthesis and recommendations mapped back to the value chain.

A.1 Weekly cadence with client; roles by track

We met **weekly** with Tim to present our investment research findings compressed into professional slidedecks. These virtual meetings would include collaboration, aligning of definitions, resolving open questions, and setting the next steps. Each session concluded with a prioritised action list (e.g., close gaps in the map, validate sources, add comps/multiples) that fed directly into the following week's work. Roles were organised **by sector track** to ensure depth, accountability, and consistent coverage.

Member	Sector
Xiaojing	Artificial Intelligence
Joshua	Robotics
Khaizuran	Quantum Computing
Asher	Space
Shiyou	Fusion

Table A1: Person-sector mapping (project roles by track)

A.2 Timeline & Workflow (June - October)

To provide a clear narrative of how the work progressed, the month-by-month research timeline below summarises the overall focus and key activities. This is followed by the weekly workflow we used to turn scoped questions into verified, presentation-ready weekly outputs.

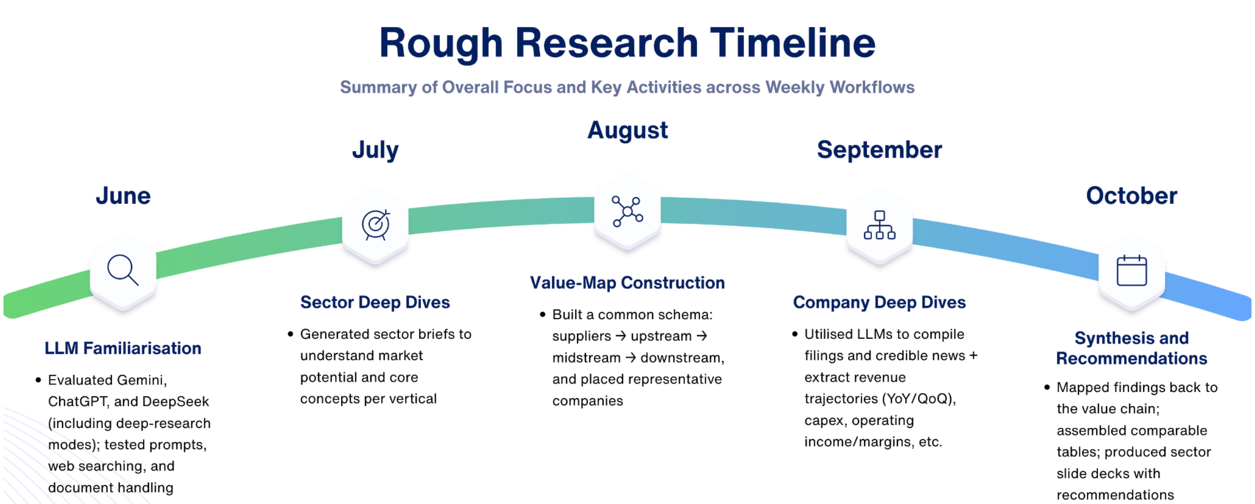


Figure A2: Research Timeline

Our weekly workflow was structured into a repeatable five-step pipeline, designed to ensure efficiency, accuracy, and clarity in our research and analysis. This framework enabled us to systematically process the data, verify its accuracy, and present it effectively to CEO Tim Zhang.

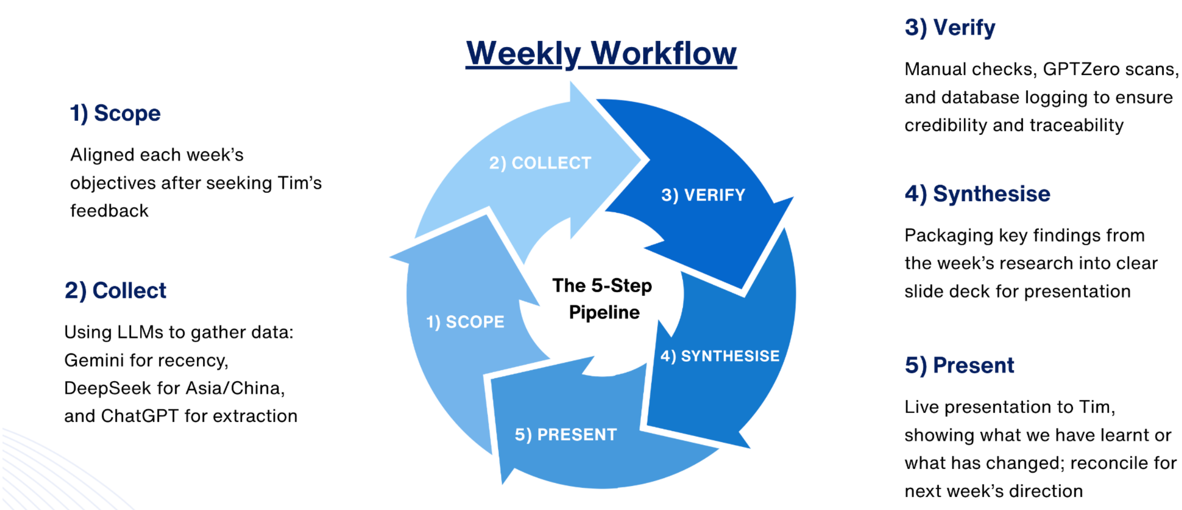


Figure A3: Weekly Workflow

The five steps in detail are as follows:

1. Scope

The first step of the process involved defining the week's objectives based on the value-chain map and continuously seeking feedback from Tim. This provided us with clear direction for the week's tasks and helped us to continuously improve our prompt engineering process when interacting with the LLMs. By aligning our questions with Tim's comments and the overall research framework, we ensured that each week's work was focused on the most pertinent issues.

2. Collect (LLM-Assisted)

Once the scope was established, we moved to the collection phase, where we began issuing prompts to the different LLMs based on the specific use case. Gemini was primarily used to gather the latest news releases and current sources, helping us stay up-to-date with industry developments. ChatGPT was tasked with extracting financial metrics and numerical data from the internet, ensuring that the financial side of the analysis was accurate. Meanwhile, DeepSeek was used to specifically target and identify Asian players, particularly companies based in China and other regions, which were often overlooked by the other models.

3. Verify

After collecting data from the LLMs, we moved into the verification step. This crucial phase involved manually reviewing each cited link to ensure the accuracy and relevance of the information. We identified and resolved any discrepancies and rejected hallucinations. Pages that passed manual checks were then processed through GPTZero to obtain an AI-generation score (percentage likelihood that the content is AI-generated). These results, together with URL, title, publisher, upload date, link reachability, credibility and recency scores were recorded in our fact-checking database to preserve provenance and enable re-checks. The AI-generation score served as an additional risk signal (not an absolute gate), prompting deeper review when scores were high or inconsistent with source type.

4. Synthesise

Once the information had been verified, each member would package their week's work into a clean slide deck for their own 20-30 minute update for our client Tim, with a focus on stepping in the direction his queries lie and showing what changed since the prior week. We prioritised clarity and decision-usefulness: tightening narrative, surfacing the key evidence, and condensing our thorough research into an informative and compact weekly update in our respective sector research. We would also try to expand the breadth and depth of research by continuously experimenting with the LLMs and prompt engineering (e.g., alternative prompt frames, few-shot examples, multilingual queries) to dig deeper where Tim probed or achieve a more efficient LLM-assisted workflow, and then reflected those findings in the deck and presentation.

To make field-specific concepts understandable, we leaned on visual aids; value-chain diagrams, flowcharts, illustrative imagery, and compact comparison tables; plus brief plain-English explainers beside the graphics.

5. Present

We closed each weekly cycle with a live presentation to Tim using the refined slide deck. The objective was to answer his priority questions, show what changed since the previous week, and secure agreement on decisions and next steps. Presentations by each member followed a tight structure: (i) a recap of last week's presentation outcome and the current week's scope and focus, (ii) a thorough walkthrough on key updates and findings (evidence-backed), and (iii) open questions.

Throughout the sharing sessions, we would walk through the reasoning behind our conclusions, demonstrate any LLM workflows or prompt variants that materially improved recall or precision, and noted caveats (conflicting estimates, stale vintages, gaps). Each session concluded with Tim's follow-ups captured as next-step actions, feeding the scope for the following week's investment research direction for each respective member. To ensure continuity, each session produced a minutes log, documented in our weekly minutes tracker. This cadence cycle kept presentations decision-useful, traceable, and directly connected to the following week's work.

B. LLM Consolidated Research Outputs - All 5 Sectors

This portion of our appendix presents the synthesised findings from our LLM-assisted research across AI, Robotics, Quantum, Space, and Fusion, organised under a standardised framework. For each sector, we provide the following:

- a brief sector primer and scope,
- a value-chain snapshot (*suppliers* → *upstream* → *midstream* → *downstream*) with representative players and current value concentrations,
- a 2–3-year outlook highlighting catalysts and bottlenecks, and investment angles with watchlist metrics.

All findings are evidence-backed with manual source verification; cross-model capability details and methodological notes are provided in Section 5. (Refer to Slide B1- B10 for full progression)

B.1 Artificial Intelligence (AI) - value-chain deep dive

<u>B1.1 Sector primer & scope</u>

Artificial Intelligence (AI) has entered a rapid expansion phase, evolving from a research-driven domain into a mainstream commercial general-purpose technology. The global AI market is projected to reach $1.8 trillion by 2030 (Grand View Research, 2024), a massive increase from $279 billion in 2024, reflecting an expected CAGR of 35-37%. This acceleration is powered by three mutually reinforcing drivers: rising enterprise adoption, breakthroughs in foundational model capabilities, and significant infrastructure investment across chips, data centers, and cloud platforms.

The AI sector today spans a broad range of activities, from semiconductor manufacturing and cloud infrastructure to model development and downstream application deployment. This broad scope means AI should be viewed not as a single industry, but as an interconnected value ecosystem where innovation and commercial value emerge at multiple layers.

Lifecycle-wise, AI has moved from a late-emerging to an early-growth stage, with adoption expanding across finance, healthcare, logistics, retail, and autonomous systems. GenAI has further accelerated the transition toward mainstream use cases, shifting AI from a back-end optimization tool to a front-end productivity engine. However, the sector remains highly macro-sensitive: interest rate movements, energy prices, and long CapEx cycles meaningfully influence the economics of scaling AI infrastructure.

Regulation is another defining factor shaping sector boundaries. The EU AI Act has introduced tiered risk classifications with compliance requirements and potential fines of up to 7% of global revenue for violations. Meanwhile, regulatory divergence persists globally, with the U.S. maintaining a lighter-touch approach and China adopting a more centralized, control-heavy regime. As AI becomes embedded into mission-critical industries, regulatory scrutiny around data usage, model governance, and bias mitigation will intensify.

Overall, the AI sector's scope encompasses infrastructure, platforms, and applications, with each layer experiencing different growth rates, risk exposures, and competitive dynamics.

4.1.2 Value-chain snapshot

The AI value chain can be divided into three core segments: upstream infrastructure, midstream software and platforms, and downstream applications. This layered structure enables the flow of innovation, from raw compute to end-market monetization.

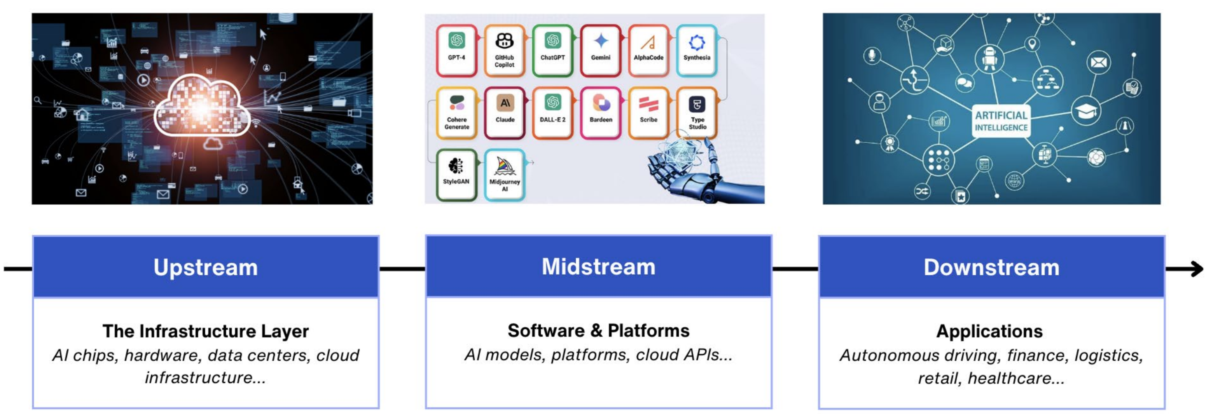


Figure B1: AI value chain

Upstream (The Infrastructure Layer): The upstream layer includes AI chips, semiconductors, hardware systems, high-bandwidth memory, data centers, and cloud hyperscalers. This segment forms the computational backbone enabling AI model training and inference. Key players include NVIDIA, TSMC, and ASML, which are critical to advanced chip manufacturing and lithography. Cloud hyperscalers (AWS, Google Cloud, and Microsoft Azure) provide scalable compute capacity and have become essential partners for both training large models and deploying AI services. Demand for upstream infrastructure is intensifying, with data-center spend expected to grow 42% in 2025, heavily driven by AI workloads. This segment currently captures the majority of economic value due to severe supply constraints in GPUs, packaging, and power availability.

Midstream (Software & Platforms): The midstream layer covers model developers, AI software platforms, and cloud APIs. It includes foundational model leaders such as OpenAI, Anthropic, xAI, and Meta, as well as emerging challengers like Zhipu, Moonshot, Minimax, and Mistral AI. This layer is the fastest-growing segment of the value chain, driven by innovation in multimodal models, agentic systems, and enterprise AI deployments. Midstream players differentiate based on model performance, safety features, scalability, and ecosystem breadth. Their platforms increasingly serve as the "operating systems" of the AI economy.

Downstream (Applications): The downstream layer translates AI capabilities into real-world use cases across industries. Applications include big data analytics (e.g., Databricks, Palantir), autonomous vehicles (Waymo, Tesla, Cruise), robotics, healthcare AI, and enterprise productivity tools. Chinese companies such as Pony.ai and Apollo are also gaining traction. Downstream markets represent the long-term monetization engine, as value shifts from model development to industry-specific vertical solutions.

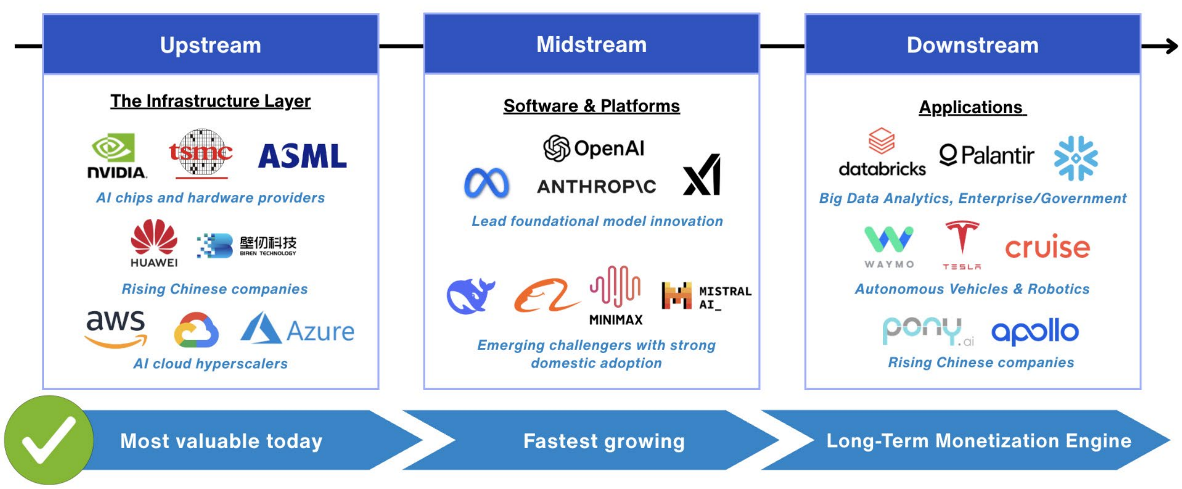


Figure B2: AI companies within each stream

4.1.3 2-3 year outlook: catalysts & bottlenecks

Over the next two to three years, the AI sector is expected to advance steadily as companies translate enthusiasm into practical deployment. This is supported by three central catalysts. First, enterprise adoption is set to accelerate as implementation frameworks mature and early adopters demonstrate clear returns in efficiency, cost reduction, and workflow enhancement. These gains (often exceeding 20% improvements in productivity) create internal momentum for scaling AI across functions such as customer support, coding, supply-chain optimisation, and decision-support tools. Second, productivity improvements across industries are validating AI's economic impact, shifting budgets from experimental pilots toward embedded, recurring use. Finally, aggressive infrastructure build-out by hyperscalers and semiconductor manufacturers is expanding the availability of GPUs, memory, and specialised data-center capacity. This upstream investment should gradually ease current supply constraints and lower barriers for downstream applications, enabling more predictable development cycles and broader commercial adoption.

Counterbalancing these catalysts are several bottlenecks that may hamper the pace of deployment. The most significant is the shortage of compute infrastructure (such as limited chip supply, insufficient packaging capacity, and structural power and cooling constraints) which restrict the rate at which new models can be trained and deployed. Many regions lack the electrical infrastructure required for next-generation data centers, creating physical bottlenecks that capital alone cannot quickly solve. At the same time, business-model uncertainty remains high among AI-native firms, with elevated compute costs and limited differentiation challenging long-term profitability. Regulatory scrutiny is also intensifying as AI penetrates sensitive domains, adding compliance burdens related to data privacy, algorithmic fairness, and operational transparency. These pressures are likely to slow implementation cycles even as overall demand continues to rise.

Together, these headwinds suggest that while overall demand will remain strong, adoption may proceed in phases, with progress determined by infrastructure availability, regulatory clarity, and improvements in model performance and reliability.

4.1.4 Investment angles & risks (watchlist metrics)

The most compelling near-term investment exposure lies in the "rails" that enable AI development and deployment. Companies involved in advanced semiconductors, GPU manufacturing, packaging and assembly, high-bandwidth memory, data-center cooling and power systems, and cloud hyperscaling stand to benefit from persistent compute scarcity and strong multi-year demand visibility. These upstream segments currently hold the greatest pricing power and are best positioned to capture outsized returns as supply remains constrained.

Key Watch Metrics

- **CapEx growth** across fabs and data centres as a proxy for long-term demand for compute.
- **Compute and data-centre utilisation rates**, indicating supply–demand tightness.
- **Energy and power-efficiency trends**, which affect inference economics and scalability.

- **Margin performance** for semiconductor and cloud providers, signalling competitive dynamics and pricing leverage.

Risks

- **Large CapEx requirements** that risk overcapacity if demand moderates.
- **Rapid technology obsolescence**, particularly across chip architectures and model-design paradigms.
- **Geopolitical and supply-chain fragility**, affecting semiconductors, materials, and export-controlled technologies.
- **Energy-cost inflation and sustainability constraints**, which may materially raise operating costs for data centres.

Taken together, the opportunity set in AI rewards exposure to segments with structural demand and defensible roles in the ecosystem, while requiring disciplined assessment of operational, technological, and macro risks that may constrain scale. Investors will need to prioritise resilience over hype, focusing on firms whose capabilities remain indispensable as the sector matures.

B.2 Robotics - value-chain deep dive

B2.1 Sector primer & scope

Robotics combines sensing, actuation, control, and increasingly AI-enabled perception and planning to execute physical work with precision and repeatability. It is a closed-loop system: sensors (vision, force, proximity) observe the world; controllers and software plan and coordinate; actuators and end-effectors do the work. The result is a system that can “see, decide, and do” at industrial cadence.

The robotics value chain can be broken down into three streams:

Upstream: Components and Sub-Assemblies (e.g. machine vision, sensors, chips etc)

Midstream: Robot OEMs and systems integrators that design and build the robots

Downstream: solution implementers and operators—systems integrators and end-users in manufacturing, logistics, and healthcare—who design, install, integrate and run automated workflows with ongoing lifecycle support.

B2.2 Value-chain snapshot

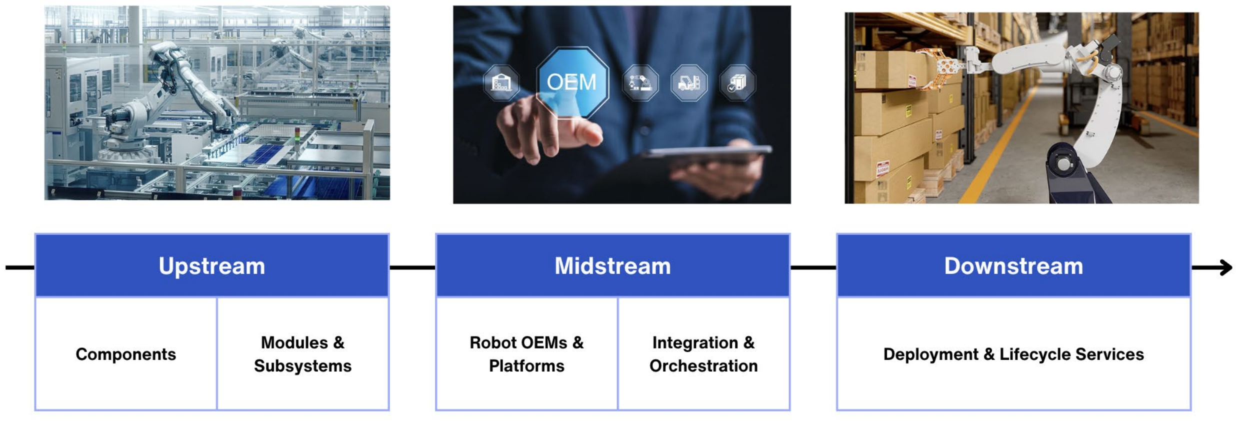


Figure B3: Robotics Value Chain

Layer	Examples of Activities	Key Players
Upstream	Machine vision and ID (smart cameras, readers); safety sensing (light curtains, scanners); precision motion (reducers/gearboxes, servos, drives); edge-AI compute modules for perception and planning.	Keyence; Cognex; Novanta; Rockwell Automation (also integrates); SMC; Harmonic Drive Systems; THK; Renishaw; HIWIN; Inovance; Innoviz; Hesai; RoboSense; Zebra (also midstream); Sony (image sensors; also midstream)
Midstream	Design, manufacture, and support of industrial/collaborative arms and mobile robots; controllers and toolchains; installed-base services; vertical application OEMs (e.g., surgical) with recurring instruments/service economics	ABB (via KUKA owner Midea is listed), FANUC, Yaskawa, Mitsubishi Electric, Kawasaki, Denso, Nachi-Fujikoshi, Omron, Epson, Yamaha, Doosan Robotics, Teradyne (Universal Robots; MiR), Intuitive Surgical, Accuray, PROCEPT BioRobotics, Vicarious Surgical, Ekso Bionics, Stereotaxis, AutoStore, iRobot, Ecovacs, Roborock, Siasun, Estun, Efort, ATS (integrator/OEM systems).

Downstream	Solution design and installation; commissioning, lifecycle support and SLAs; operator programmes and Robotics-as-a-Service (subscription) deployments.	Symbotic (operator programmes); Ocado (OSP tech + in-house ops); Guardforce AI; FBR (Hadrian X "Wall-as-a-Service"); Nauticus Robotics (offshore RaaS); Knightscope (Machine-as-a-Service); SoftBank Robotics (subscription deployment); EHang (UAM services ambition); Hyundai (Boston Dynamics + factory deployment); Cyberdyne (Robocare centres).

Table B1: Robotic companies within each stream

Upstream (components & subsystems). This layer makes robots see, sense, move, and compute: machine-vision systems and barcode readers; safety scanners and light curtains; precision reducers/gearboxes, servos and drives; and edge-AI compute modules. Mix is application-dependent—industrial arms lean on precision reducers and servos; logistics and quality-inspection skew to vision/safety; autonomy adds GPU-class edge compute. Representative vendors include **Keyence** (vision/sensing), **Cognex** (vision/ID), **SICK** (safety), **Nabtesco RV** and **Harmonic Drive** (precision gearboxes), and **NVIDIA Jetson Orin** (edge AI). Revenue is primarily equipment plus service/maintenance tied to factory and warehouse deployments, giving clearer near-term monetisation than software-heavy layers.

Midstream (robot OEMs & platforms). Robot manufacturers design, build, and support platforms—industrial/collaborative arms, AMRs, and specialised systems—bundling controllers, toolchains, and orchestration. Examples include **ABB**, **FANUC**, **Yaskawa Motoman**, and **KUKA**. In healthcare, **Intuitive Surgical** is a vertical OEM whose **instruments & accessories revenue is primarily driven by procedure volumes**, creating a durable "razor-and-blade" dynamic alongside system sales and services. Software spans layers; **ROS 2** underpins perception/planning/control integrations used in production and teaching contexts, reinforcing vendor-agnostic orchestration.

Downstream (integrators & operators). Systems integrators and operators design, install, and run application-specific solutions, tying robots into WMS/WES/MES/EMR and guaranteeing performance through commissioning and lifecycle SLAs. Exemplars include **Dematic** (KION), **Bastian Solutions** (Toyota Advanced Logistics), and **Honeywell Intelligrated**. On the operator side, **Walmart × Symbotic** illustrates multi-site rollouts under long-term arrangements; subscription models such as **Robotics-as-a-Service (RaaS)** (e.g., **Locus Robotics**) shift spend from capex to opex while bundling fleet, software, and maintenance.

B2.3 2-3 year outlook: catalysts & bottlenecks

The robotics and automation industry is poised for sustained expansion over the next two to three years. Recent IFR statistics show global industrial-robot installations holding above the 500k-unit level in 2024, with Asia accounting for roughly three-quarters of new deployments—an adoption base that supports continued scale-up if core software plumbing matures. This trajectory suggests a dynamic sector with substantial headroom. In this section, we examine the key catalysts and bottlenecks across the stack and identify where value is most likely to accrue.

The main bottleneck sits **upstream in software**, not policy. In plain terms, robots and devices still don't "plug-and-play" smoothly across brands or sites. Three things slow projects down. First, the core, open-source software used widely in robotics (ROS 2) is still being hardened for factory-grade reliability at scale; when timing and behaviour are not fully predictable, integrators add workarounds, which lengthens projects. Second, machines on the factory floor—sensors, drives, controllers, and robots—are only part-way to speaking the same language by default; the OPC Foundation's field-level standards are advancing, but broad, vendor-neutral use in real plants takes time. Third, mixed fleets of mobile robots (from different vendors) still need "glue code" even when they follow the same playbook; the VDA 5050 interface helps, but most sites still require extra integration to reach full function. Until these layers are mature and certified widely, deployments will move at the pace of the upstream software stack, regardless of demand.

Against that backdrop, the **midstream** which are robot OEMs, systems integrators, and orchestration platforms, emerges as the most valuable and revenue-generating segment over this horizon. Once upstream software reaches production maturity, midstream players convert installed base into utilisation-linked software, services, and upgrade cycles, compounding economics as sites go live. Industry surveys and trade analyses consistently flag end-to-end orchestration/integration as the decisive capability, underscoring why the centre of gravity for value sits with platforms that can reliably connect devices, data, and workflows at scale. In short: upstream software maturity enables growth; midstream platforms monetise it.

In summary, over 2026-2027, upstream software is expected to be the pacing item for deployment scope and speed. As those constraints clear, value should accrue fastest to midstream platforms and integrators that translate software readiness into site commissioning, orchestration, and recurring services at scale.

<u>B2.4 Investment angles & risks (watchlist metrics)</u>

Building on the robotics value chain (upstream: foundational components like sensors and software; midstream: integration platforms and orchestration; downstream: end-user applications and RaaS models), this subsection evaluates investment opportunities, associated risks, and key metrics to monitor. Our analysis leverages LLM-assisted research, including ChatGPT for structured data extraction, Gemini for real-time trend summaries, and DeepSeek for multilingual insights into Asia-dominated supply chains. This approach enabled rapid synthesis of IFR statistics and McKinsey reports, reducing manual review time by approximately 56% as per our efficiency benchmarks.

Investment Angles

The robotics sector offers diversified entry points amid projected growth (over 500,000 global installations in 2024, per IFR), with bifurcation favoring quality players.

- **Picks-and-Shovels Premium with Dispersion**: Upstream segments like sensors, vision systems, safety tech, motion controls, and edge compute benefit from the multi-year rise in global robot stock. These "rails" of automation command premiums for established names, but adjacent areas like LiDAR exhibit mixed profitability—profitless rallies during scale-up phases are common, highlighting valuation dispersion.

- **Platforms Converting Deployments to Recurrence**: Midstream orchestration software, service contracts, and consumables (e.g., razors in warehouse or surgical robotics) generate durable revenue as pilots mature to fleet-scale. This winner-take-most dynamic rewards platforms with strong backlog quality and software/service mix.

- **Capex to Opex via RaaS**: Downstream Robotics-as-a-Service (RaaS) or outcome-based models lower customer upfront costs, compounding vendor economics once cohort paybacks are proven. This shifts fragility in micro-caps toward sustainable growth.

Risks

Despite catalysts, risks span the chain, amplified by integration bottlenecks.

- **Pilot-Purgatory and Commissioning**: Overhead from schedule overruns and suboptimal throughput delays cash conversion and IRR, particularly in midstream.

- **Capex Cyclicality and ROI Drift**: Budget volatility in factories, logistics, and hospitals extends paybacks; upstream is less affected but exposed via downstream demand.

- **Technology/Standards Obsolescence**: Evolving ISO 10218 safety norms and FDA regulations necessitate redesigns, limiting deployments.

- **Supply Chain/Geopolitics**: Asia concentration (75% of installations) risks lead times and costs from shocks or export controls.

- **Financing for Loss-Makers**: Capital dependence in LiDAR/micro-caps heightens dilution if scale lags.

Watchlist Metrics

To navigate angles and mitigate risks, track these KPIs quarterly:

- **Deployment and Utilization**: Installed-base growth, go-live velocity, uptime/MTBF, and realized vs. projected paybacks (flags pilot slippage).
- **Revenue Quality and Concentration**: Backlog-to-revenue ratio, software/service mix, net revenue retention, and top-customer exposure (assesses midstream stability).
- **Unit Economics for Loss-Makers**: Cohort-level margins, cash burn/runway (months), and growth-efficiency (Rule of 40/X) over P/E.
- **Regulatory Milestones**: ISO 10218 conformity and FDA indications (impacts TAM and cadence).

LLM tools streamlined this evaluation, cross-verifying sources for 71% faster fact-checking, though human oversight ensured balanced risk assessment.

B.3 Quantum Computing - value-chain deep dive

B3.1 Sector primer & scope

Quantum computing exploits non-classical effects to process information in ways that can outperform classical machines for specific problem classes (simulation of quantum systems, certain optimisation, cryptography). Four core principles underpin its potential: *superposition* (a qubit occupies weighted combinations of 0 and 1), *entanglement* (correlated qubits act as one joint state), *interference* (amplitudes reinforce or cancel to amplify correct answers), and decoherence (quantum system loses its quantum properties due to interaction with its environment, causing it to collapse into a single classical state).

Intuitively, a classical computer explores a maze path by path; a quantum computer prepares a superposed state of many paths at once, uses interference to dampen dead-ends and amplify viable routes, then measures to retrieve a high-probability solution. Performance gains are problem-dependent and contingent on hardware quality (fidelity, coherence) and algorithm design.

The analysis emphasises the enabling stack; suppliers (cryogenics, optics/lasers, RF/test, materials, control electronics), upstream hardware/QPUs (superconducting, trapped-ion, photonic, neutral-atom, silicon/spin), and the midstream control/tooling layer (compilers, orchestration, error mitigation, cloud access). Downstream applications are referenced only where they materially inform requirements or near-term demand, as commercial adoption remains early.

B3.2 Value-chain snapshot

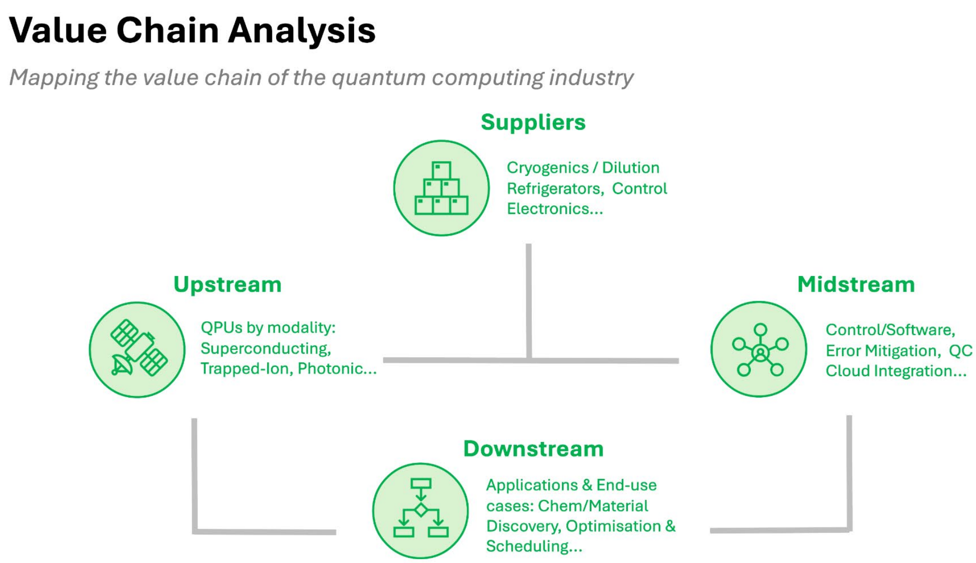


Figure B4: Quantum Computing Value Chain Snapshot

The quantum computing stack can be read as an enabling pipeline in which suppliers feed both upstream hardware (QPUs) and the midstream control/tooling layer, with downstream applications emerging primarily through cloud-delivered pilots. At today's maturity, recognised revenue concentrates in suppliers and midstream services; upstream hardware revenue is programmatic and lumpy; downstream remains early.

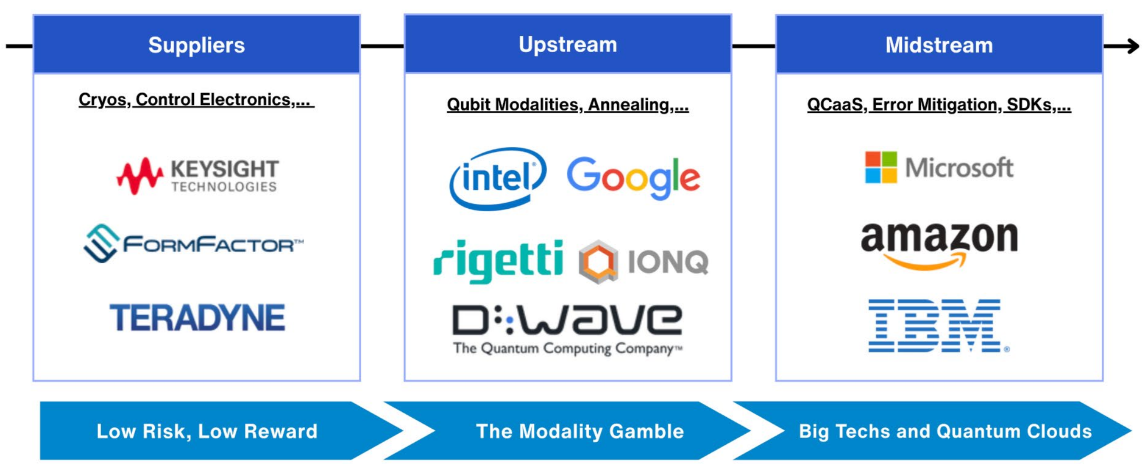


Figure B5: Quantum computing companies within each stream

Layer	Examples of Activities	Key Players
Suppliers	Cryogenics, dilution fridges, optics/lasers, RF/microwave test, probe stations, materials & nanofab, control electronics	Keysight Technologies Inc. (KEYS); Teradyne Inc. (TER); FormFactor Inc (FORM)
Upstream	QPU design/fab across superconducting, trapped-ion, photonic, neutral-atom, spin/semiconductor, + quantum annealing; packaging, calibration; small system delivery (on-prem/cloud)	**Diversified Big-Tech:** Intel Corp (INTC); Google (GOOG)[1] **Pure-plays:** Rigetti Computing (RGTI); IonQ Inc (IONQ); D-Wave (QBTS)
Midstream	Compilers, orchestration, error mitigation; QC cloud brokers	**Diversified Big-Tech:** Microsoft (MSFT)[2]; Amazon.com (AWZN)[3]; IBM (IBM)[4]
Downstream	High accuracy of simulations, speedups for machine learning and AI advancement; optimising complex computational problems	*Downstream remains early to identify key players; potential fields of application include chem/materials, optimisation, finance, etc.*

Table B3: Quantum computing companies within each stream

[1] Google develops superconducting QPUs and algorithms (upstream) and publishes/control-tooling (e.g., Cirq) that sits in midstream, but has no material QCaaS revenue today.

[2] Microsoft monetises in midstream via Azure Quantum (aggregation, control, simulators), while continuing upstream research on topological qubits.

[3] AWS monetises in midstream via Braket (aggregation, orchestration, billing), while maintaining an upstream research program at the AWS Center for Quantum Computing.

[4] IBM is vertically integrated: designs/operates its own QPUs (upstream) and monetises primarily through midstream platform access and services.

Suppliers: This layer provides the "picks and shovels" required to build and operate quantum systems: dilution refrigerators and cryogenics; vacuum systems; lasers and precision optics; RF/microwave test and measurement; probe stations; materials and nanofabrication inputs; and control electronics. Dependence is modality-specific (e.g. superconducting hardware is cryo- and RF-heavy; trapped-ion and photonic approaches lean more on lasers/optics) but in all cases suppliers primarily feed upstream and, to a lesser extent, midstream (calibration rigs, error-mitigation experiments). Revenue here is recognised on equipment and service contracts tied to labs, foundries, and testbeds, with relatively clearer near-term monetisation than other layers.

Upstream (QC hardware): Hardware vendors design, fabricate, package, and calibrate QPUs across modalities (superconducting, trapped-ion, photonic, neutral-atom, and silicon/spin qubits) and deliver small systems on-prem or via cloud access. The set includes diversified technology leaders (IBM, Google, Intel) and pure-plays (Rigetti [RGTI], IonQ [IONQ], D-Wave). Revenue is driven by government and enterprise pilots, research programs, and limited system deliveries; visibility depends on roadmap credibility (quality × scale), access via public clouds, and stickiness of consumption. Switching costs rise with control-stack and workflow integration, but hardware revenues remain lumpy and milestone-based.

Midstream (control/tooling and platform): Sitting between hardware and applications, this layer provides compilers, orchestration, error-mitigation/control software, workflow SDKs, and cloud brokerage. It stabilises device behaviour, abstracts vendor differences, and enables reproducible experimentation. Representative providers include Microsoft (Azure Quantum), AWS (Braket), and independent platforms such as Q-CTRL, Riverlane, Classiq, and QC Ware. Monetisation is steadier than upstream—subscriptions, support, and services—though absolute scale remains modest. As benchmarks and interfaces converge, midstream vendors with vendor-agnostic integrations and strong device-control IP are positioned to capture growing spend.

Downstream (applications and services): Commercial use today is dominated by proofs of concept in chemistry/materials, optimisation, and selected finance workflows, typically delivered as managed services through cloud brokers. Revenue is largely services-led and contingent on upstream quality and midstream maturity; we reference downstream only where it informs requirements or near-term demand for the enabling stack.

<u>B3.3 2-3 year outlook: catalysts & bottlenecks</u>

Near-term progress is likely to be incremental but meaningful. On the catalyst side, steady gains in qubit fidelity and stability/uptime, larger counts of calibrated qubits, and improved error-mitigation workflows should expand the scope of credible pilots. Tighter cloud integration (managed runtimes, better scheduling/queuing) and expanding public/enterprise programs will continue to underwrite demand, while clearer interfaces/benchmarks across vendors should reduce switching friction and support midstream adoption. Supply chains for critical inputs (cryo, optics/lasers, RF/test) are also maturing, allowing labs and foundries to scale with fewer one-off bottlenecks.

Countervailing bottlenecks remain material. Full error correction imposes heavy overhead; manufacturability and yields for larger devices are uncertain; coherence/crosstalk challenges compound as systems scale; and benchmark ambiguity complicates vendor comparisons. Specialized talent scarcity in control, calibration, and compilers slows throughput, and modality-specific supply dependencies (e.g., cryo for superconducting, precision optics for trapped-ion/photonic) can still delay programs. As a result, value capture should remain concentrated in suppliers and midstream control/tooling, with upstream hardware revenues continuing to be programmatic and milestone-driven.

B3.4 Investment angles & risks (watchlist metrics)

The most defensible near-term exposure is to picks-and-shovels suppliers (cryo, optics/lasers, RF/test, control electronics) where demand is tied to labs, testbeds, and foundry expansions. Within software, vendor-agnostic control/orchestration and error-mitigation platforms are positioned to benefit from multi-hardware adoption and recurring subscriptions. Select hardware teams merit attention where roadmaps show repeatable quality×scale progress and sticky cloud consumption; systems integrators/consultancies can monetize services as pilots broaden. For diversified big tech, QCaaS aggregators may see steady platform usage even as upstream remains pre-commercial.

Key risks include slower-than-expected fidelity gains; unresolved scaling and yield issues; delays in logical-qubit demonstrations; dependence on a few critical component vendors; shifting benchmarks that erode claimed advantage; and budget cyclicality in government/enterprise programs. For pure-plays, financing risk and customer concentration are non-trivial; for diversifieds, quantum contributions may remain immaterial relative to core businesses.

Watchlist metrics.

- Hardware quality & scale: two-qubit fidelity; error rates (gate/SPAM); calibrated qubits; stability/uptime; movement toward logical-qubit demonstrations.
- Throughput & adoption: cloud utilisation (job counts, queue times), number of pilots and conversion to repeat usage, partner/integrator cadence.
- Economics: disclosed recurring revenue/backlog, services mix vs software, unit lead times for key supplier categories (cryo, optics, RF/test).
- Standardisation & comparability: participation in open benchmarks, cross-vendor portability of workloads, API/runtime convergence.
- Program signals: government/enterprise grant awards, new testbeds/foundries coming online, and modality-specific supply continuity (e.g., HTS availability, precision optics capacity).

These angles and metrics align with the value-chain emphasis of this report—prioritising suppliers and midstream platforms for near-term exposure, while tracking upstream progress that could shift value toward hardware as quality and scale improve.

B.4 Space - value-chain deep dive

B4.1 Sector primer & scope

The space sector spans materials/components, satellite manufacturing, launch and ground infrastructure, satellite operations, and end-user applications (satcom, EO, navigation). This deep dive covers commercial activity across the full stack, with government programs referenced where they materially affect demand or capacity.

B4.2 Value-chain snapshot

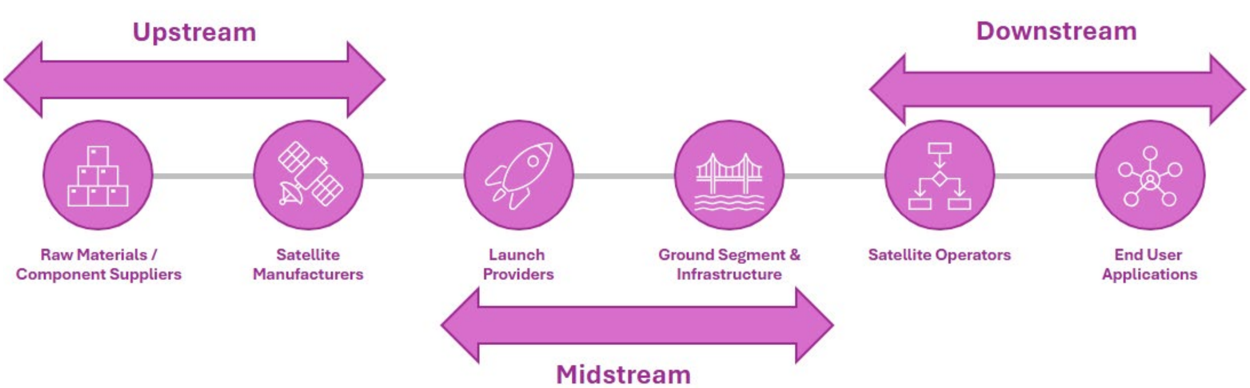


Figure B6: Value Chain of the Space Sector

The Space Sector is divided into the respective streams and within the streams it is further divided into 6 key segments as shown in the figure above. Each of these segments plays a crucial role in the development and commercialization of space technologies, contributing to the overall dynamics of the industry. (European Space Agency, 2019) Through extensive research, we have identified the key players within each segment of the space industry, categorized by upstream, midstream, and downstream:

Layer	**Value-Chain Activity**	**Key Players**
Upstream	Raw Materials/ Component Suppliers, Satellite Manufacturers	Astroscale (186A.T), iSpace (9348.T), MDA Space (MDA), MTAR Technologies (MTARTECH)
Midstream	Launch Providers, Ground Segment and Infrastructure	Airbus Defence and Space (AIR.PA), Thales S.A. (HO.PA), Avio S.p.A. (AVIO), Kongsberg Gruppen ASA (KOG.OL), Mitsubishi Heavy Industries (7011.T), NEC Corporation (6701.T), Northrop Grumman (NOC)
Downstream	Satellite Operators, End-User Applications	Eutelsat Group (ETL), Sky Perfect JSAT (9412.T), Viasat (VSAT), EchoStar (SATS), SES S.A. (SESG.PA), Iridium Communications (IRDM), Telesat (TSAT), China Satcom (CSAT)

Table B3: Space companies within each stream

These companies represent some of the most significant players in their respective segments. Notably, the space sector remains fragmented, with no single entity monopolizing the entire value chain. (Coykendall, 2025)

Current value concentration: While various reports from McKinsey, Morgan Stanley, and the European Space Agency have analysed the space industry, there is currently no definitive figure to accurately value the sector. However, it is generally agreed that the downstream segment is the most valuable, followed by upstream, with the midstream sector considered the least valued at present. This distribution of value highlights the growing importance of end-user applications and satellite services in driving revenue, while the challenges associated with infrastructure development and launch capabilities continue to impact the midstream sector.

B4.3 2-3 year outlook: catalysts & bottlenecks

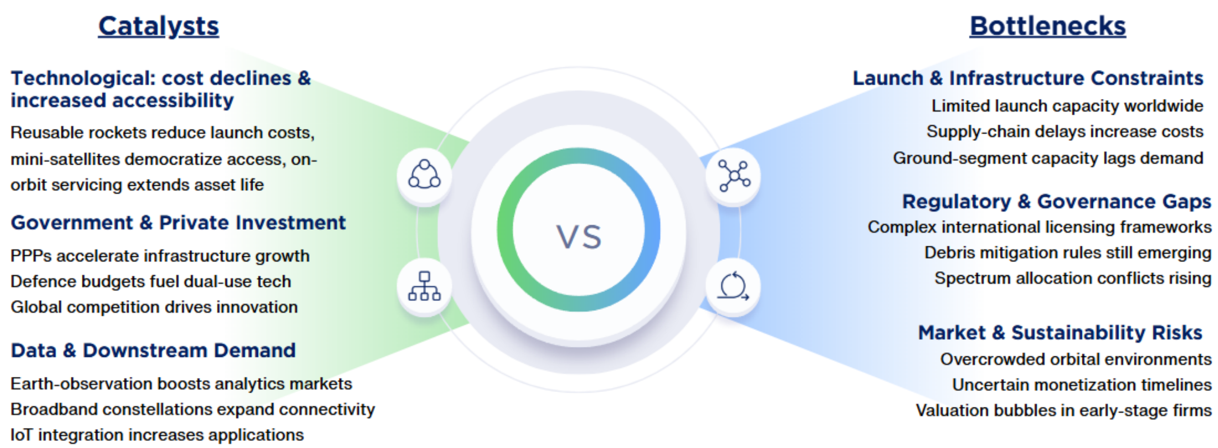


Figure B7: Space Key catalysts and bottlenecks

The space industry is on a strong upward trajectory, with multiple forecasts including McKinsey, Morgan Stanley, and the European Space Agency projecting the sector to surpass $1 trillion by around 2030. This growth is caused by a combination of powerful catalysts and persistent bottlenecks. For catalysts, rapid technological advancements have significantly reduced costs and increased accessibility. (Coykendall, 2025) Innovations such as reusable rockets, mini-satellites, and on-orbit servicing have lowered entry barriers while extending asset lifecycles, fostering greater participation from both established players and new entrants.(European Space Agency, 2024) Government and private investment further accelerate this growth, with public-private partnerships (PPPs), defense funding for dual-use technologies, and global competition driving innovation across the sector. (OECD, 2021) Meanwhile, downstream demand is spurred by earth-observation analytics, broadband constellations, and IoT integration. These options continue to expand the commercial viability of space-based services. (Refornte Learning , 2025)

However, the industry faces notable bottlenecks, particularly in launch and infrastructure constraints. Limited launch capacity (Deloitte, 2024), supply-chain disruptions (Turner, 2025), and lagging ground-segment development (Mallory, 2024) continue to impede scalability. Regulatory and governance challenges also persist, as complex international licensing systems, evolving debris mitigation rules, and increasing spectrum allocation conflicts slow coordinated progress.(OECD, 2024) Additionally, market and sustainability risks such as overcrowded orbits, uncertain monetization models, and inflated valuations in early-stage firms pose further barriers to long-term stability. (OECD, 2024) Together, these bottlenecks underscore that while downstream innovation drives immense value creation, the midstream segment remains the key structural bottleneck restricting the industry's ability to scale sustainably.

B4.4 Investment angles & risks (watchlist metrics)

Investment in the space sector is multi-dimensional, with distinct opportunities and risks across its value chain.

In the upstream segment, investments in satellite manufacturing and raw materials offer long-term growth, but face challenges such as high capital requirements and reliance on government contracts.

Midstream investments, particularly in launch services and ground infrastructure, are promising due to innovations in reusable rocket technology, yet they carry risks from technological failures and regulatory constraints.

Downstream, satellite operators and end-user applications are the most lucrative, driven by demand for services like broadband and earth imaging, though they face competition from emerging technologies like 5G. Emerging markets such as space tourism, space mining, and lunar exploration present speculative opportunities, while risks include technological uncertainty, geopolitical tensions, and the environmental threat of space debris.

Watchlist metrics, such as YOY revenue growth, P/E ratio and Year-To-Date performance should be incorporated to assess the attractiveness of specific space companies. Overall, the space industry offers high reward potential but requires careful navigation of its inherent technological, regulatory, and market risks.

Company	Ticker	Stream	Price	Market Cap	PE Ratio (TTM)	Latest Sales (quarter)	Sales one year prior	Sales YoY %	Net Income (quarter)	Net Income (one year prior)	Net Income YoY %	1M	6M	1Y (TTM)
Astroscale	186A.T	Upstream	805.00 JPY	89.99B JPY	NIL (Loss making)	¥1,250,333	¥239,387	+422.3%	-¥1,211,251	-¥8,579,996	Loss narrowed by 85.9%	19.97%	9.67%	0.37%
iSpace	9348.T	Upstream	478.00 JPY	56B JPY	NIL (loss making)	¥1,165M	¥635M	+83.4%	-¥2,879M	-¥1,579M	Loss widened by 82.4%	-11.2 %	-51.0 %	-29.8 %
Avio S.p.A.	AVIO	Midstream	43.75 EUR	1.41B EUR	247.39	€234.863 M	€180.606 M	+30.0%	-€0.187m	-€1.783m	Loss narrowed ~89.5%	-7.41 %	154.2 %	258.4 %
SES S.A.	0M6P.IL	Downstream	6.55	5.06B	19.82	€ 509M	€ 498M	+2.2 %	€ 29 M	€ 73M	-60.3 %	+11.0%	+46.2 %	+113.4 %
Airbus Defence and Space	AIR.PA	Midstream	208.45 EUR	160.26B	32.62	€13.5B	€12.8B	+6%	€793M	€595M	+33%	+7.6 %	+44.5 %	+33.1 %
Thales S.A.	HO.PA	Midstream	253.40 EUR	55.71B	53.35	€10,265M	€9,493M	+8.10%	€664M	€1,017M	-34.71%	-1.1 %	+9.5 %	+86.2 %
Eutelsat Communications	ETL.PA	Downstream	3.5600 EUR	1.76B	7.31	€293M	€300M	-2.33%	-248.0M	€316.0M	-178.5%	+14.6 %	+0.9 %	+61.4 %

Figure B8: A Table Of Key Metrics Of Selected Companies

B.5 Fusion - value-chain deep dive

B5.1 Sector primer & scope

Fusion development follows two primary technology routes: magnetic confinement fusion (MCF) and inertial confinement fusion (ICF); each with distinct supplier needs, integration pathways, and downstream use cases. This deep dive focuses on the commercial portions of the supply chain that are already procuring equipment and services for experiments, demonstrators, and early pilots. Government programs are referenced where they materially shape demand, qualification, and timelines.

B5.2 Value-chain snapshot

Fusion splits into two main technology routes: magnetic confinement fusion (MCF) and inertial confinement fusion (ICF). Each route maps to the value chain analysis:

Magnetic Confinement Fusion

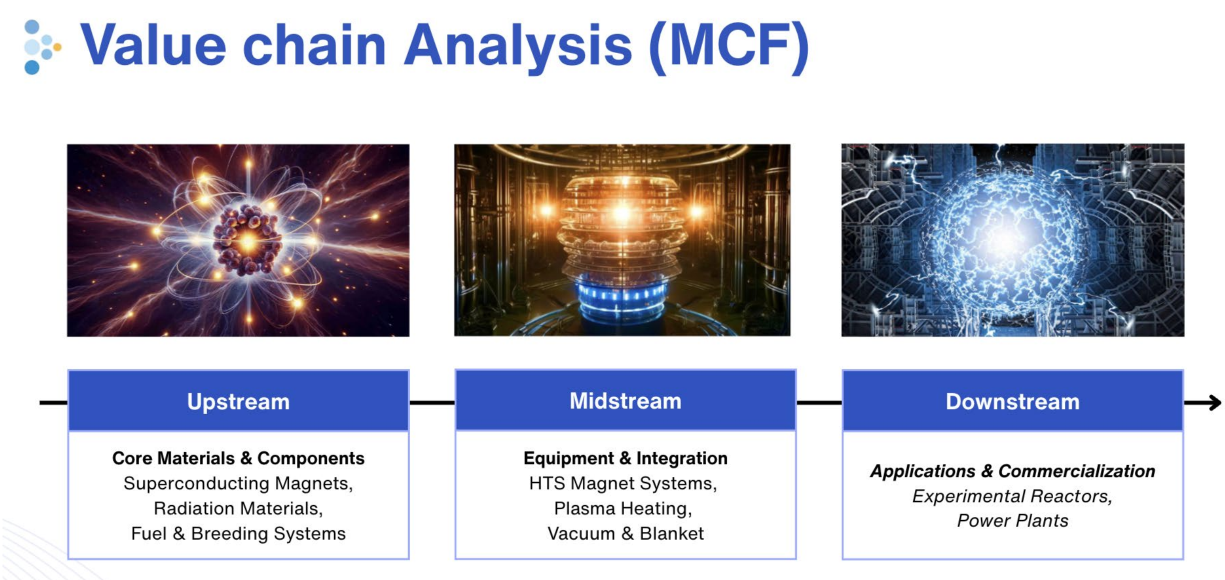


Figure B9: Fusion Value Chain Analysis for Magnetic Confinement Fusion

Inertial Confinement Fusion

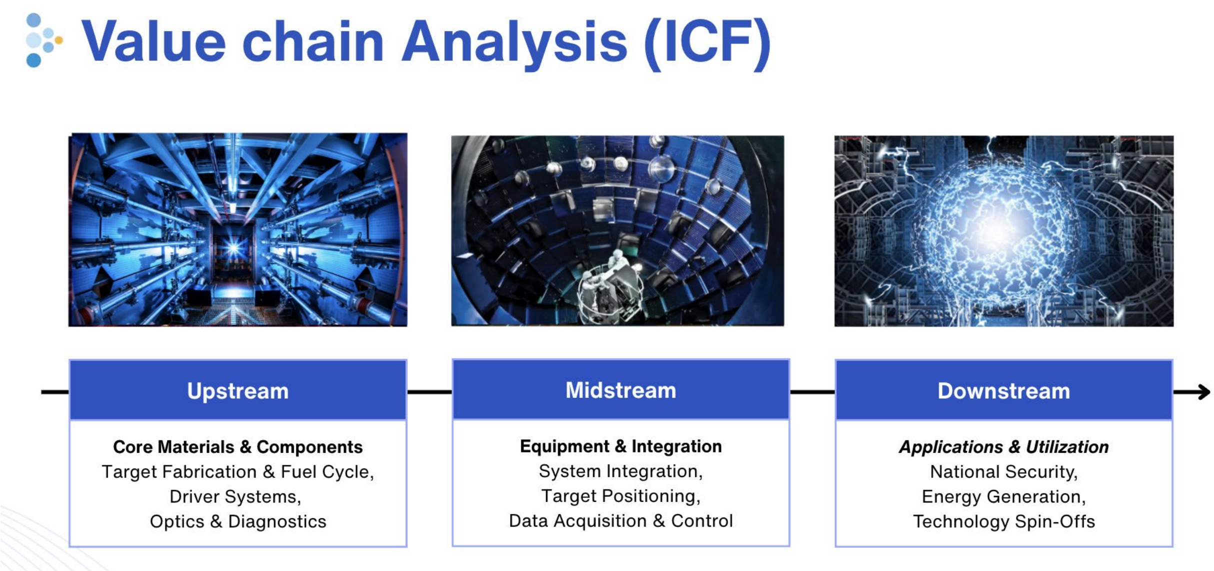


Figure B10: Fusion Value Chain Analysis for Inertial Confinement Fusion

Because fusion remains in the experimental and early-commercial stage, the downstream segment—covering power sales, industrial heat and hydrogen, and data-centre power purchase agreements—has few scaled operators and virtually no publicly listed companies. Investable, bid-able revenue today is concentrated in the upstream, where materials and components such as superconductors, high-energy lasers, and vacuum electron devices are procured, and in the midstream, where systems integration and heavy equipment are engineered and delivered. Most listed names involved in fusion are diversified industrials, with fusion representing only a small portion of their overall revenue exposure.

Public Companies

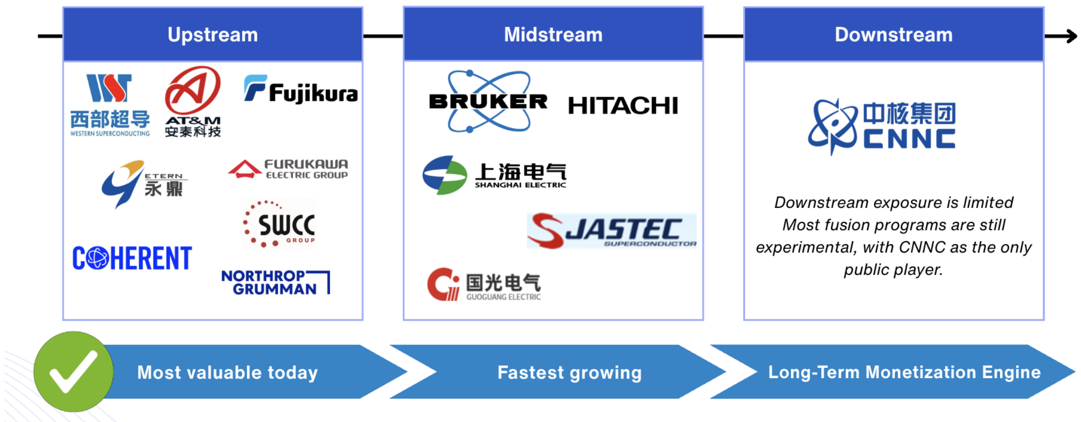


Figure B11: Companies within the Fusion value chain

Layer	Representative players (tickers where public)
Upstream	**MCF:** Bruker(BRKR), Furukawa Electric(5801.T), Fujikura(5803.T), SWCC(5805.T), Western Superconducting(688122.SH), Advanced Technology and Materials(000969.SZ), Chengdu Guoguang Electric(688776.SH) **ICF:** Coherent(COHR), Yongding(600105.SH)
Midstream	**MCF:** Hitachi(6501.T), Shanghai Electric(601727.SH), China Nuclear Engineering(601611.SH) **ICF:** Northrop Grumman(NOC)

Downstream	China National Nuclear Corporation(601985.SH)

Table B4: Fusion companies within each stream

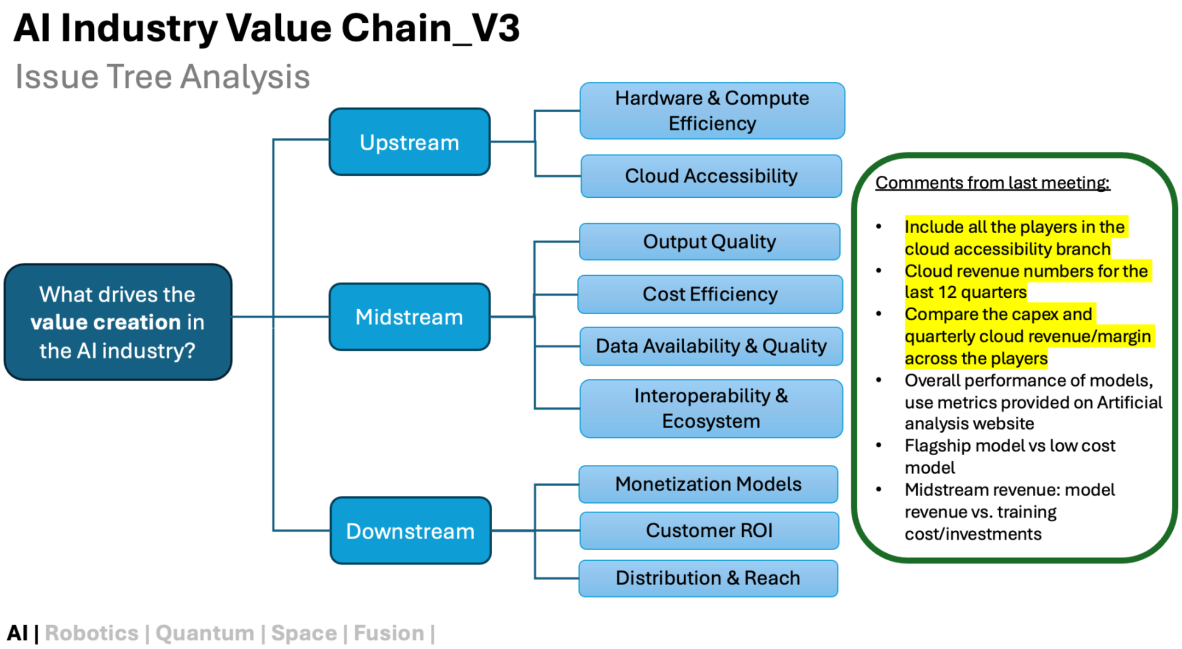
While various industry trackers and institutional reports have analysed the fusion sector, there is currently no definitive figure to accurately value it. However, it is generally agreed that recognised revenue is concentrated in the upstream segment (materials and components such as superconductors, magnet assemblies, cryogenics, vacuum electron devices, and high-energy lasers) followed by the midstream segment, where systems integration, engineering, and EPC for demonstrators and pilot facilities occur. The downstream segment remains nascent with few publicly listed companies. This distribution of revenue reflects the sector's experimental and early-commercial status, in which procurement is driven by R&D and early demonstration programs rather than mature end-user offtake.

B5.3 2-3 year outlook: catalysts & bottlenecks

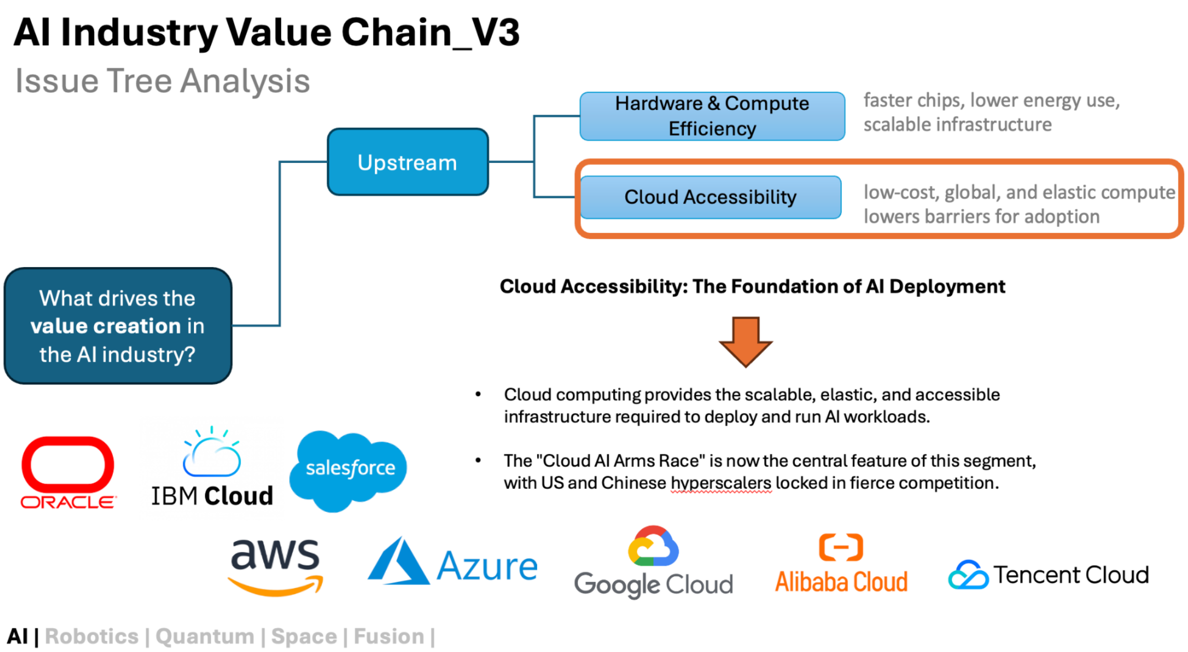
Progress over the next two to three years will be driven by three forces: steady cost and reliability gains in enabling components (expanding REBCO and Nb_3Sn supply, standardization in cryogenics and vacuum hardware, incremental efficiency improvements in high-energy laser modules); growing integration momentum as multiple teams move from design into assembly and commissioning of first-of-a-kind subsystems; and a strengthening demand narrative as data centers and industrial users explore long-duration, carbon-free power and high-grade heat. Despite these positive drivers, the sector still faces a series of constraints that will affect timelines. In ICF, repetition rate and the durability of laser drivers remain the dominant technical hurdles. For MCF, the challenges lie in the manufacturability and lifetime of high-field magnets, the qualification of neutron-resistant materials, and the development of credible tritium breeding systems. More broadly, both routes face bottlenecks in component qualification under relevant radiation and thermal loads, supply-chain yields for HTS wire and precision optics, regulatory clarity for fusion-specific facilities, and FOAK financing for integrated demonstrators. The midstream, where complex subsystems converge, therefore plays a dual role: it accelerates the sector when integration proceeds smoothly, and it becomes the critical bottleneck when qualification or procurement delays emerge.

B5.4 Investment angles & risks (watchlist metrics)

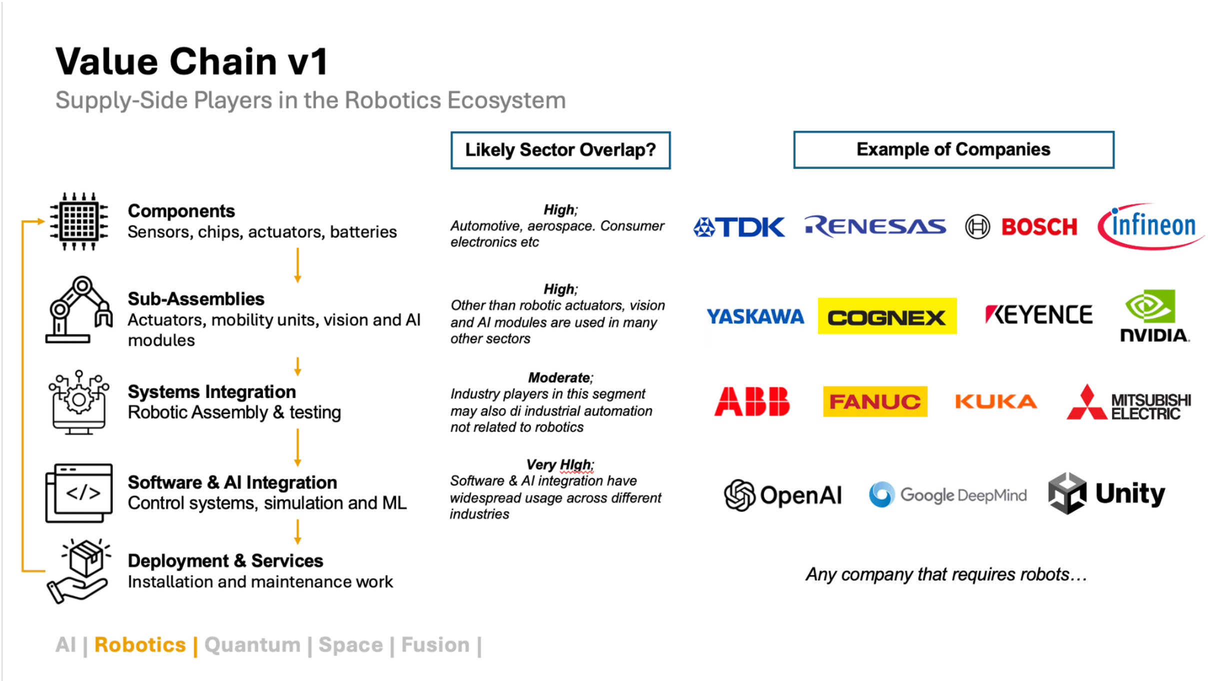
Investment in the fusion sector is multi-dimensional, with distinct opportunities and risks across its value chain. In the upstream segment, exposure to superconductors, high-energy lasers and optics, vacuum electron devices, cryogenics, targets, and diagnostics offers the clearest near-term revenue, but faces hurdles such as capital-intensive manufacturing, qualification under radiation/thermal loads, and customer concentration across a small set of fusion programs. Midstream opportunities—systems integration, magnet modules and cryostats, beamlines and power supplies, and EPC for first-of-a-kind plants—are promising as more demonstrators move from design to build; however, they carry FOAK risks around schedule slippage, bespoke hardware cost overruns, supply-chain bottlenecks (HTS wire, precision optics), and non-standardized project finance. Downstream remains the most lucrative in the long run—grid PPAs, industrial heat and hydrogen, data-center deals—but is still nascent and constrained by technology readiness, component lifetimes, and regulatory clarity. Adjacent opportunities (e.g., ICF-derived high-power lasers, HTS adoption beyond fusion) present nearer-term, dual-use upside, while sector-wide risks include policy shifts, export controls, and path dependence on competing MCF/ICF architectures. Watchlist metrics to assess company attractiveness include fusion revenue share, year-over-year order intake and funded backlog tied to named programs, margin and yield trends on fusion SKUs, milestone cadence (coil, cryostat, driver repetition-rate), qualification results under relevant loads, capex-per-MWe and implied LCOE for pilots, cash runway and funding mix, and exposure to regulatory or export-control regimes. Overall, fusion offers high reward potential but demands careful navigation of technological, regulatory, financing, and execution risks.



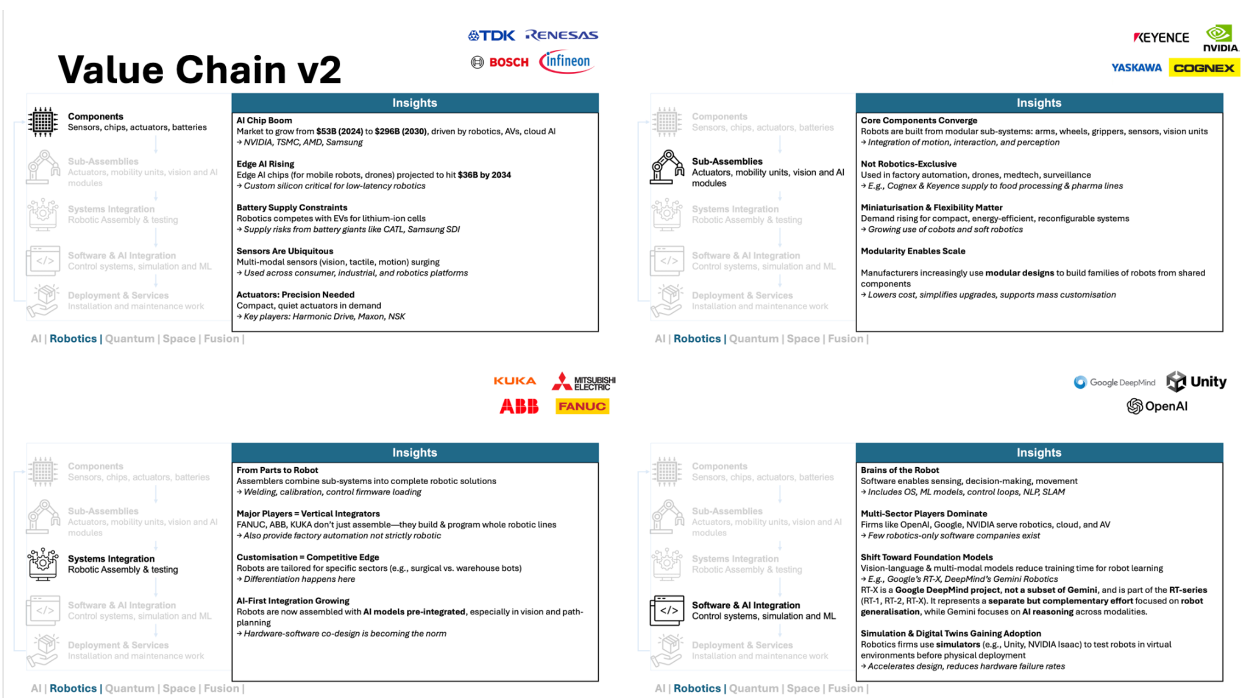
Slide B1: AI value chain progression 1



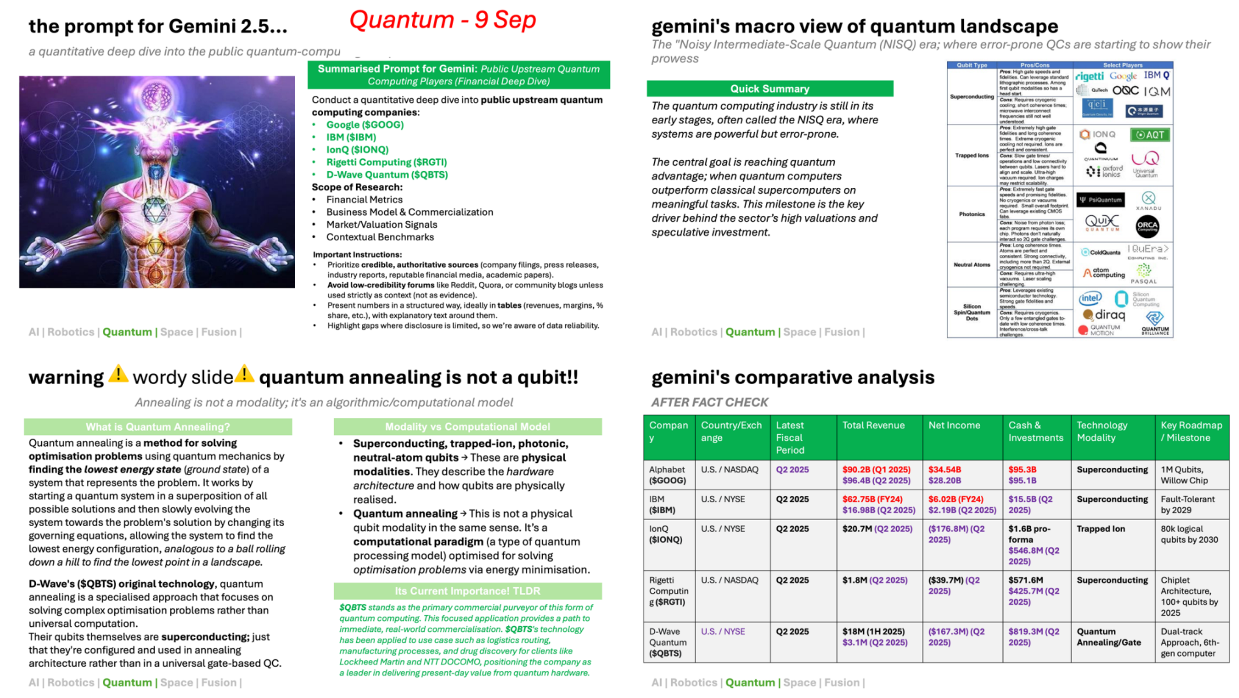
Slide B2: AI value chain Progression 2



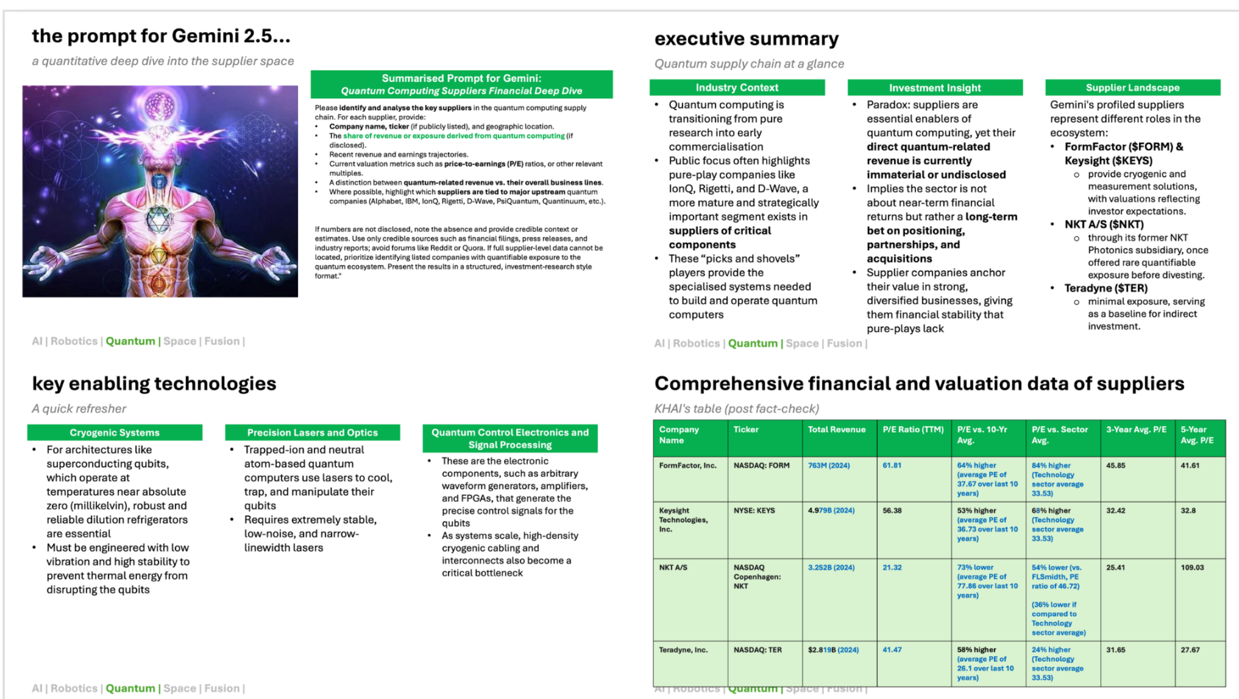
Slide B3: Robotics value chain progression 1



Slide B4: Robotics value chain progression 2



Slide B5: Quantum value chain progression 1



Slide B6: Quantum value chain progression 1

Deep diving into the financial reports

The main financial reports Gemini referenced

Northrop Grumman Reports Second Quarter 2025 Financial Results 6

SPACE SYSTEMS	Three Months Ended June 30		%	Six Months Ended June 30		%
$ in millions	**2025**	2024	**Change**	**2025**	2024	**Change**
Sales	**$ 2,646**	$ 3,002	**(12)%**	**$ 5,214**	$ 6,151	**(15)%**
Operating income	**280**	304	**(8)%**	**563**	634	**(11)%**
Operating margin rate	***10.6 %***	*10.1 %*		***10.8 %***	*10.3 %*	

Sales

Second quarter 2025 sales decreased $356 million, or 12 percent, primarily due to wind-down of work on the restricted space and Next Generation Interceptor (NGI) programs, which reduced sales by $283 million, as well as lower volume on Space Development Agency (SDA) satellite programs due to the timing of materials.

Operating Income

Second quarter 2025 operating income decreased $24 million, or 8 percent, primarily due to lower sales, partially offset by a higher operating margin rate. Operating margin rate increased to 10.6 percent from 10.1 percent principally due to higher net EAC adjustments.

Northrop Grumman Q2 2025 financial reports

Revenues and EBIT (Revenue +17%)

Airbus Defence and Space	**5,813**	4,985	+17%	**161**	-760	-

EBIT adjusted

Airbus Defence and Space	**265**	-807	-

By Business Segment	Order Intake (net)			Order Book		
	H1 2025	H1 2024	Change	**30 June 2025**	30 June 2024	Change
Airbus, in units	**402**	310	+30%	**8,754**	8,585	+2%
Airbus Helicopters, in units	**171**	233	-27%	**926**	913	+1%
Airbus Defence and Space, in millions of Euros	**5,084**	6,059	-16%	**N/A**	N/A	N/A

REVENUES BY BUSINESS AREA

	Second Quarters Ended		Six Months Ended	
(in millions of Canadian dollars)	**June 30, 2025**	**June 30, 2024**	**June 30, 2025**	**June 30, 2024**
Geointelligence	$ 52.7	$ 54.9	$ 104.4	$ 106.4
Robotics & Space Operations	88.0	78.3	165.3	148.6
Satellite Systems	232.6	108.8	454.6	196.1
Consolidated revenues	$ 373.3	$ 242.0	$ 724.3	$ 451.1

MDA Space Financial reports

Summary

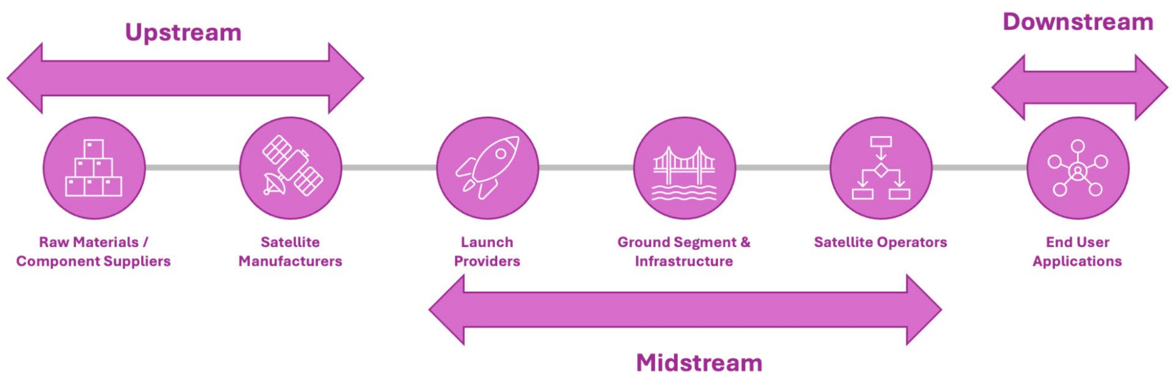
- Most financial reports state that there is an increased revenue in their space systems and operations.
- Indicates that the space economy is still steadily growing and has potential

AI | Robotics | Quantum | **Space** | Fusion |

Slide B7: Quantum value chain progression 1

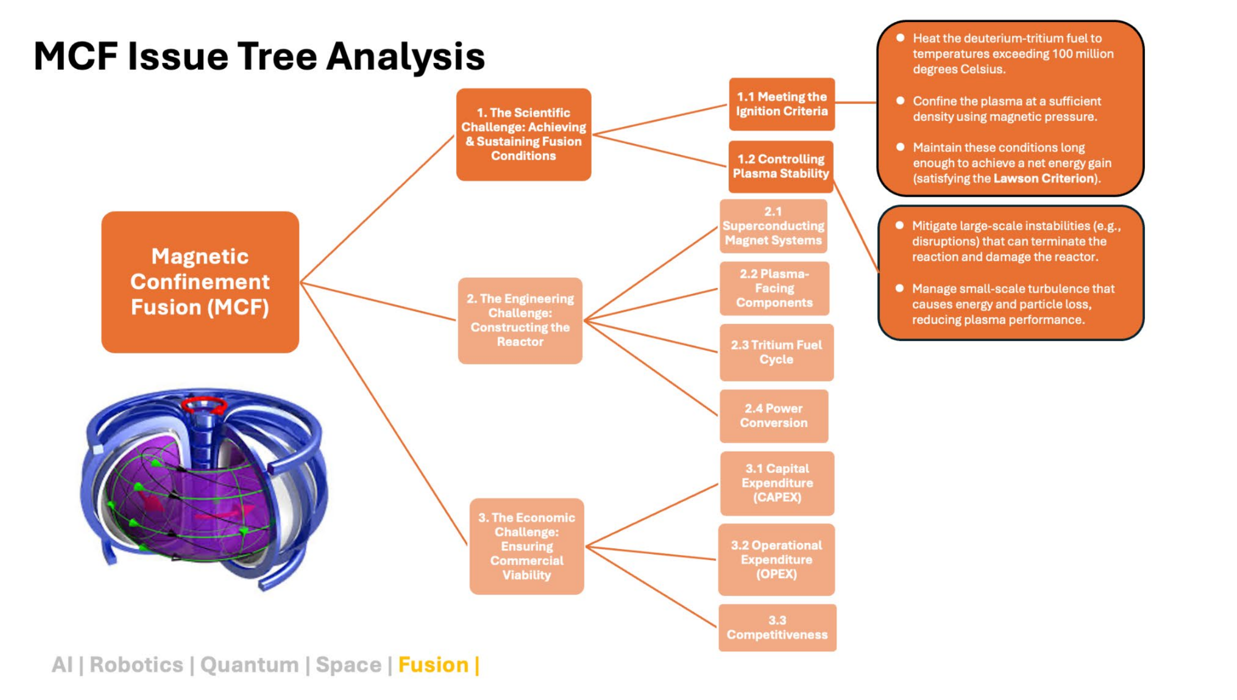
Value Chain Analysis

Mapping the value chain of the emerging space industry

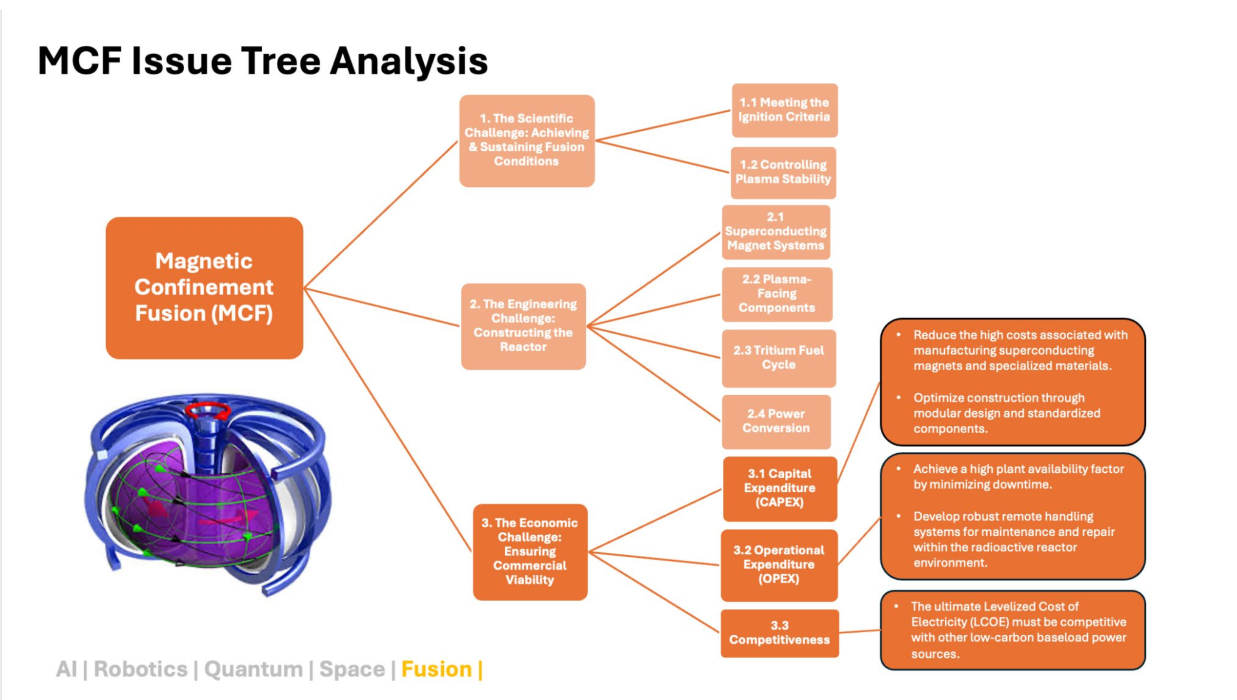


AI | Robotics | Quantum | **Space** | Fusion |

Slide B8: Quantum value chain progression 2



Slide B9: Quantum value chain progression 1



Slide B10: Quantum value chain progression 2

C. Source Database

C.1 Source Database: Data Inventory and Structure

This section documents the evidence database used to evaluate our LLM-assisted research process. Between August and November 2025, the team logged sources roughly every two weeks, resulting in a curated corpus of 250+ entries across five verticals (AI, Robotics, Quantum, Space, Fusion). Each record corresponds to a single URL cited in our weekly work or surfaced by an LLM (Gemini, ChatGPT, DeepSeek) for potential inclusion in weekly slide decks and sector analyses.

<u>What's in the database</u>

Each row captures the minimum metadata required to re-trace a claim and audit its provenance:

- **Source / Type / Brief Content:** Publisher name, publisher category (e.g., company press release, regulatory filing, trade press, analyst blog), and a one-line summary of what the page contributes (numbers, roadmap, partnership, etc.).
- **Credibility (1–5):** A rubric-based score reflecting publisher reliability and evidentiary strength (primary > reputable secondary > opinion/blog).
- **Recency (1–5):** A rubric-based score derived from on-page or header dates (publication/updated), emphasizing freshness of the cited fact.
- **Hyperlink Valid?:** Reachability outcome at time of logging (200 OK vs. 403/404/timeout).
- **AI Generation Scores**: Percent likelihoods from GPTZero (AI / Mixed / Human) used strictly as a triage signal for additional review.
- **URL / Date Accessed / LLM / Model Used**: Full link, Date of Access using LLM, and which model produced or assisted the retrieval (e.g., "Gemini Flash 2.5", "GPT-5", "DeepSeek V3").

The database supports three main analyses: (i) **quality signals** (credibility/recency), (ii) **AI-screening score triage** (AI-generation signal to prioritise human review), and (iii) **link hygiene** (reachability rates by model and over time). Note that GPTZero outputs are used as risk indicators, not grounds for exclusion. Finally, because some sites resist automated access (403/anti-bot), reachability reflects conditions at access time; where no reliable on-page date was found, manual checks were performed in the report's fact-checking workflow.

Source	Type	Brief Content	Credibility (1-5)	Recency (1-5)	Hyperlink Valid?	AI Generation Score (How many % AI)	AI Generation Score (How ma	AI Generation Score (How mar	URL	Date Accessed	LLM	Model Used
Amazon Q2 2025 Earnings Release	Company Filing	AWS revenue $30.87B, OI $10.16B, TTM AWS $116.38B	5	4	Yes	4.00%	6.00%	90.00%	om/news-release/news-release-details/2025/Amazon-com-Announces-Second-Quarter-Results/default.aspx	19 September 2025	ChatGPT	5
Microsoft FY25 Q4 Press Release	Company Press Release	Intelligent Cloud $29.88B, Azure +39% YoY, OI $12.14B; Azure annual run-rate $75B	5	4	Yes	2.00%	4.00%	94.00%	https://www.microsoft.com/en-us/investor/earnings/fy-2025-q4/press-release-webcast	19 September 2025	ChatGPT	5
Alphabet Q2 2025 Earnings Release	Company Filing	Google Cloud $13.62B, OI $2.83B, ~20.7% margin; Cloud run-rate > $50B; 2025 capex ~$85B	5	4	Yes	1.00%	4.00%	95.00%	https://s206.q4cdn.com/479360582/files/doc_financials/2025/q2/2025q2-alphabet-earnings-release.pdf	19 September 2025	ChatGPT	5
Alibaba June qtr 2025 Results	Company Press Release	Cloud RMB 33.40B (~$4.66B), Adj. EBITA RMB 2.95B (26% YoY); AI product adoption noted	5	4	Yes	0.00%	8.00%	92.00%	https://www.alibabagroup.com/en-US/document-1897714462505304064	19 September 2025	ChatGPT	5
Tencent Q2 2025 Press Release	Company Press Release	FBS RMB 55.5B; Business Services (incl. cloud) "teens %" YoY; no standalone cloud revenue	3	4	Yes	0.00%	0.00%	100.00%	https://static.www.tencent.com/uploads/2025/08/13/18643e8bf726c0fb99719171f70d45b0.pdf	19 September 2025	ChatGPT	5

Table C1: Snippet of Source Database for AI Sector

C.2 Data Quality and Cleaning

To prepare the source database for analysis, we applied a consistent set of normalisation, validation, and data cleaning steps across all five sector sheets. The goal was to convert heterogeneous, manually entered rows into a comparable "ready" dataset without losing the original evidence trail. (Refer to Figure 4)

We began by standardising key fields. **LLM** and **Model Used** labels were folded into canonical names to remove spelling and variant drift; for example, "Gemini DeepResearch" was mapped to **Gemini**; "ChatGPT 5", "GPT 5", and "GPT-5 Thinking" were mapped to **GPT-5**; "4o/GPT-4o" to **GPT-4o**; and "DeepSeek-v3" variants to **DeepSeek V3**. Dates in **Date Accessed** were coerced to ISO-8601; any non-parseable strings were flagged for manual correction.

Next, we normalised link-hygiene and AI-screening fields. **Hyperlink Valid?** was coerced to a true/false flag from common inputs ("yes/no", "true/false", "1/0"). **AI Generation Scores** (AI/Mixed/Human) were checked for internal consistency: for valid links, the triplet must be present and sum to ≈100% (allowing a small rounding tolerance) and cannot be 0/0/0. For invalid links (unreachable at access time), the triplet is intentionally left blank (NaN) to avoid implying a screening outcome where no content was available. **Recency (1–5)** and **Credibility (1–5)** were validated against their rubrics; obvious mis-keys (e.g., out-of-range values) were corrected or excluded.

Finally, we defined the **"ready row"** filter used for quantitative analysis. A row is considered ready if: (i) the link was reachable at access time (**Hyperlink Valid? = Yes**); (ii) **Credibility**, **Recency**, **Date Accessed**, and the full AI-screening triplet are non-missing and internally consistent; and (iii) the LLM/model labels are in their canonical form. Rows failing any integrity check were excluded from the ready set but retained in the raw workbook with their original values for traceability (we kept shadow columns such as LLM_raw / Model Used_raw before standardisation).

Applying this pipeline yielded a balanced, analysis-grade sample of at least **50 ready rows per sector**, with uniform field semantics across AI, Robotics, Quantum, Space, and Fusion.

	Total rows	Valid links	Valid+Complete rows	Has ≥50 ready?
AI	51	50	50	True
Fusion	91	53	53	True
Quantum	51	50	50	True
Robotics	79	76	50	True
Space	64	50	50	True

Figure C1: DataFrame of cleaned database statistics

D. Methodology on Process Efficiency Calculation

This section quantifies how the project's workflow changes translated into time savings. We evaluate two points in the research pipeline: first, sector research conducted with LLM assistance compared to a conventional, manual approach; second, fact-checking executed with our automated tool versus the manual process. All figures are normalised to analyst-hours to make comparisons fair regardless of team size or calendar duration.

D.1 LLM-Assisted Investment Research (time savings vs. traditional)

To benchmark our performance, a single full-time investment analyst typically needs 2–3 months to progress from onboarding to delivering final sector recommendations (we use the midpoint **2.5 months**). To translate months into hours, we anchor working time to the finance industry's investment analyst typical workweek in Singapore. Recent survey data place the sell-side average at **~52.6 hours/week**; we adopt a conservative planning rate of **57.5 hours/week** to avoid understating effort. We convert months to hours using the common planning convention of **~21.7 working days per month** (52 weeks × 5 days ÷ 12 months), which is widely cited in workforce planning references.

Manual Traditional Investment Research			
Time taken for full-time analyst to complete sector	Lower	Upper	**Est. Average**

analysis using fully manual (non-LLM) approach	bound	bound	
Total time (months)	2	3	**2.5**
Total time (hours)	356.50	534.75	**445.63**

Table D1: Manual (Non-LLM) Benchmark: Hours to Complete One Sector

Hours per month are computed as:

$$Hours\ Per\ month\ =\ \frac{57.5}{7}\ \times\ 21.7\ \approx\ 178.25\ Hours$$

Thus, the traditional effort per sector is:

$$Baseline\ Hours\ =\ 2.5\ \times\ 178.25\ \approx 445.63\ Hours$$

LLM-Assisted Investment Research						
Researcher	Xiaojing	Joshua	Khaizuran	Asher	Shiyou	**Est. Average**
Weekly Hours	12	15	12	10	12	**12.2**
Total Hours	192	240	192	160	192	**195.2**

Table D2: Average estimated times for LLM Assisted research

With reference Table XX containing the estimated amount of time spent, under the LLM-assisted approach, weekly time spent per researcher averaged **12.2 hours.** Over **16 weeks**, the effective single-analyst load is:

$$LLM\ assisted\ Hours\ =\ 16\ \times\ 12.2\ =\ 192.2\ Hours$$

Against the baseline **445.63 hours**, the LLM-assisted workflow required, on average **195.2 hours**, saving **250.43 hours**—a **56.2% reduction** in analyst hours to reach comparable decision-useful outputs (weekly briefs, value-chain, company shortlists, and preliminary recommendations). The interpretation is straightforward: LLMs compress the "read–search–summarise" portion of the work while analysts retain evidence weighting, reconciliations, and recommendations; total analyst-hours fall materially, and parallel effort further shortens calendar time.

D.2 <u>Fact-Checking Workflow Using the Full-Stack Tool</u>

To measure efficiency at the verification stage, we normalise everything to **cycle time per 50 links**. Each team member estimated a lower–upper bound for manual verification; the mid-column is the individual point estimate. Averaging across members yields **4.3 hours** per 50 links, with an observed range of **3.8–4.8 hours**.

Manual Fact-Checking			
Researcher	Lower bound (hours)	Upper bound (hours)	**Est. Average (hours)**
Xiaojing	4	5	4.5
Joshua	4	5	4.5
Khaizuran	5	6	5.5
Asher	3	4	3.5
Shiyou	3	4	3.5
Average	**3.8**	**4.8**	**4.3**

Table D3: Manual fact checking average hours

We then ran the same workload through the **Lovable-built tool**, which operationalises our Python prototype (automated link extraction and de-duplication; reachability checks with programmatic fetch plus headless fallback; recency parsing where available; structured logging). Because the current UI does not include an API key for GPTZero, **AI-generation scores still require manual copy-paste**, and a minority of sites resist automated access, so a small amount of **manual date/credibility checking** remains. Under these constraints, the tool **reduced average cycle time to 1.25 hours per 50 links**, a **71% reduction** versus manual, or a **3.44× throughput gain**.

For a "theoretically complete" version—i.e., the same workflow with stable headless access across sites and a live GPTZero API call embedded in the app—the team estimates **0.375 hours per 50 links**. That counterfactual implies a **91% reduction** versus manual and an **11.5× throughput gain**.

Scenario	Cycle time per 50 links (hours)	Time saved vs. manual (hours)	Reduction	Throughput gain
Manual baseline	4.3	[redacted]		
Current tool (Lovable UI + our backend)	1.25	3.05	−71%	3.44×
Fully working version (est.)	0.375	3.925	−91%	11.5×

Table D4: Theoretical time saved using the prototype

In practical terms, the present tool already converts a half-day of checking into **~75 minutes**, chiefly by eliminating dead links, collapsing duplicates, and standardising date capture. The remaining gap to the "fully working" estimate is attributable to two residual frictions: occasional **link reachability** requiring a manual open and the absence of **in-app GPTZero** (prompting manual text transfer). Integrating a GPTZero key and hardening the headless browser path would close most of that gap and push verification toward the 20–25 minutes per 50 links range.

E. Figures

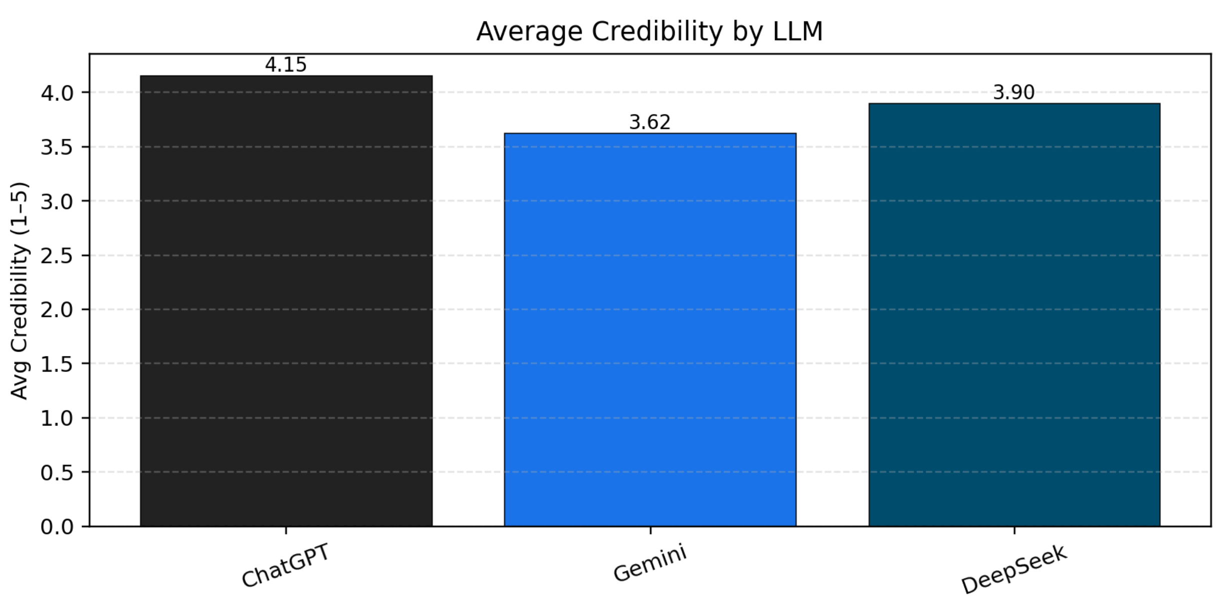


Figure 1: Average Credibility score by LLM (Out of 5)

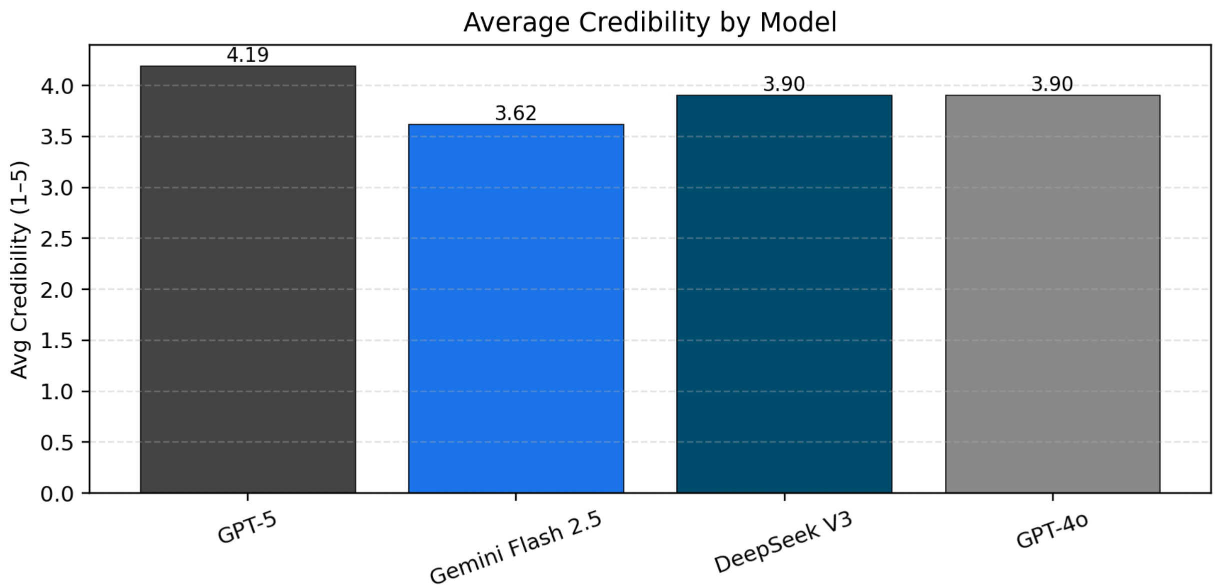


Figure 2: Average credibility score by model (Out of 5)

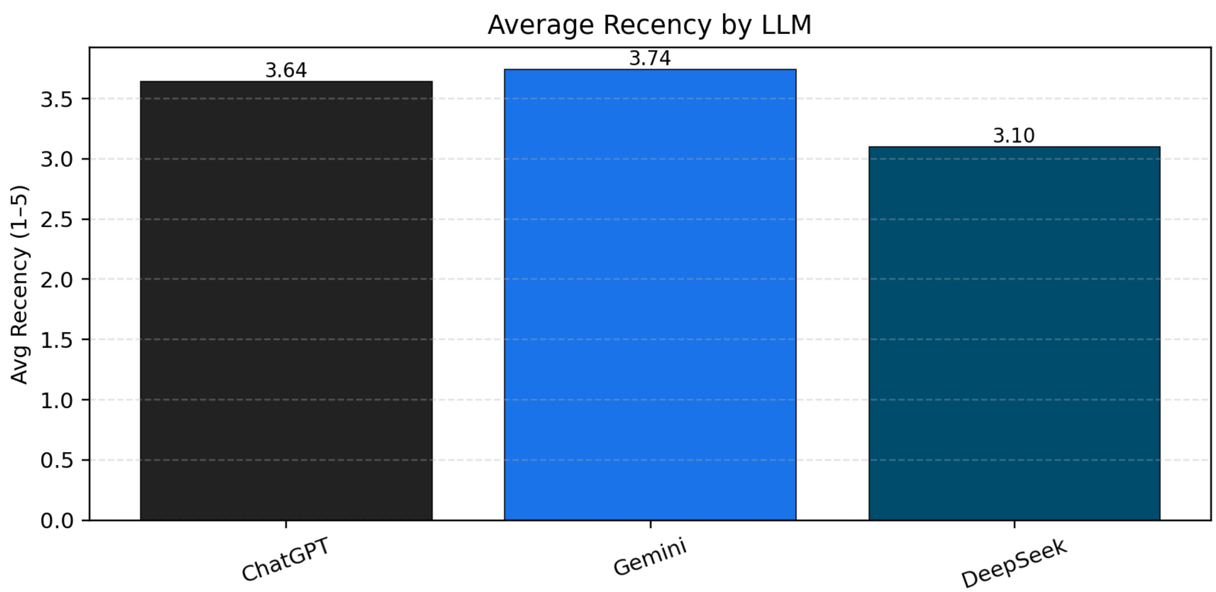


Figure 3: Average recency by LLM

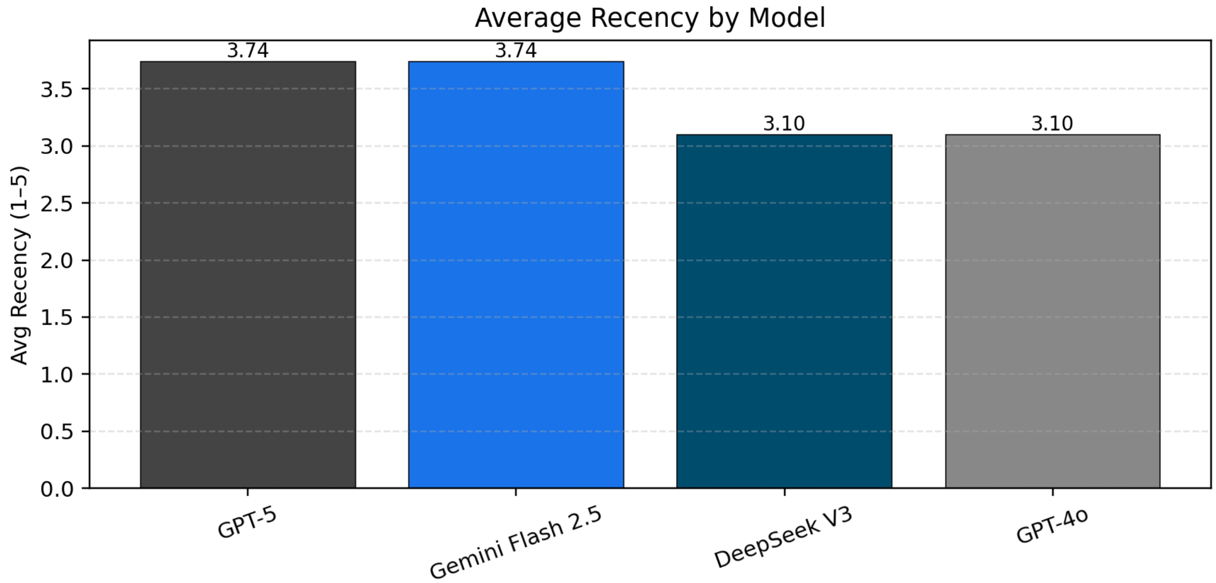


Figure 4: Average recency by model

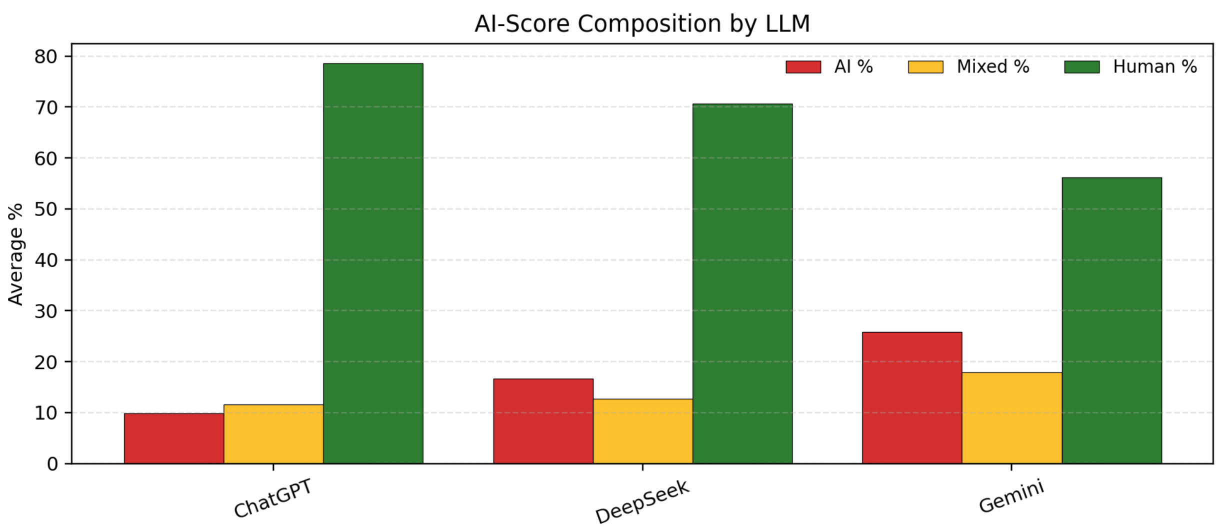


Figure 5: AI-score comparison

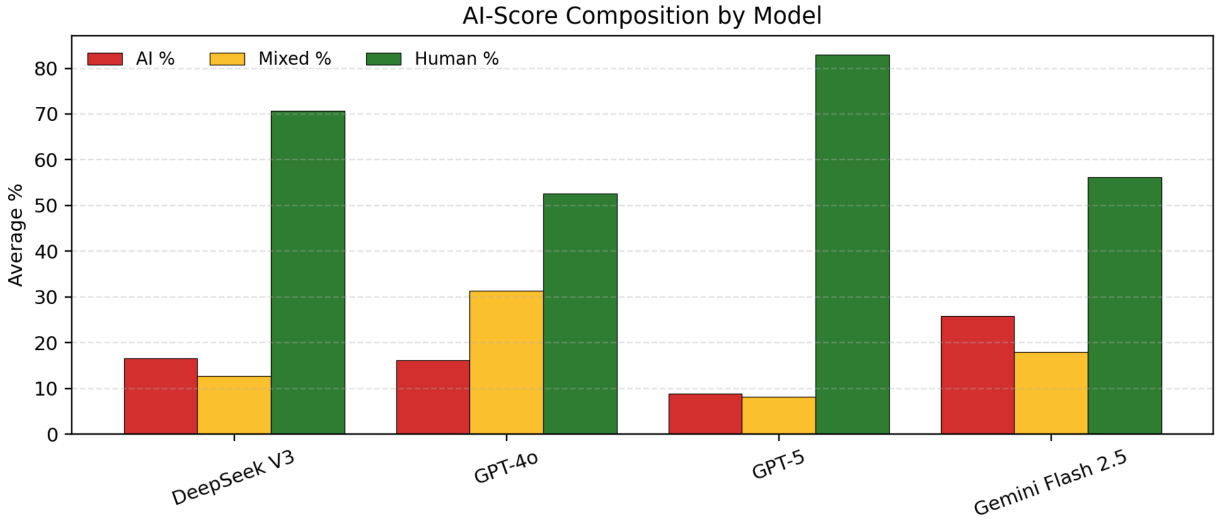


Figure 6: AI-score comparison by model

F. Tables

ChatGPT	Iterative drafting, document parsing, numeric extraction, table building
Gemini	Web-grounded research sprints, recency checks, citation-rich outlines
DeepSeek	CN-language discovery (filings, standards, news), multilingual search

Table 1: LLM use cases

Phase	Traditional Workflow	LLM-Assisted Workflow	Time Impact
Scoping	Manual hypothesis framing	LLMs assist with idea generation & data-gap mapping	↓ by ~50%
Data Collection	Manual sourcing & cleaning	Automated scraping & summarization	↓ by ~60–70%
Analysis & Modelling	Spreadsheet-heavy, manual benchmarking	LLMs generate ratio analyses & scenario drafts	↓ by ~40–50%
Synthesis	Analyst writes from scratch	LLM drafts base report & visuals	↓ by ~50–60%
Review & Compliance	Manual proofreading	AI-assisted fact-checking, tone & citation tools	↓ by ~30%
Follow-Up	Periodic updates	Continuous AI-driven monitoring & alerting	Real-time

Table 2: LLM assisted workflow VS Traditional

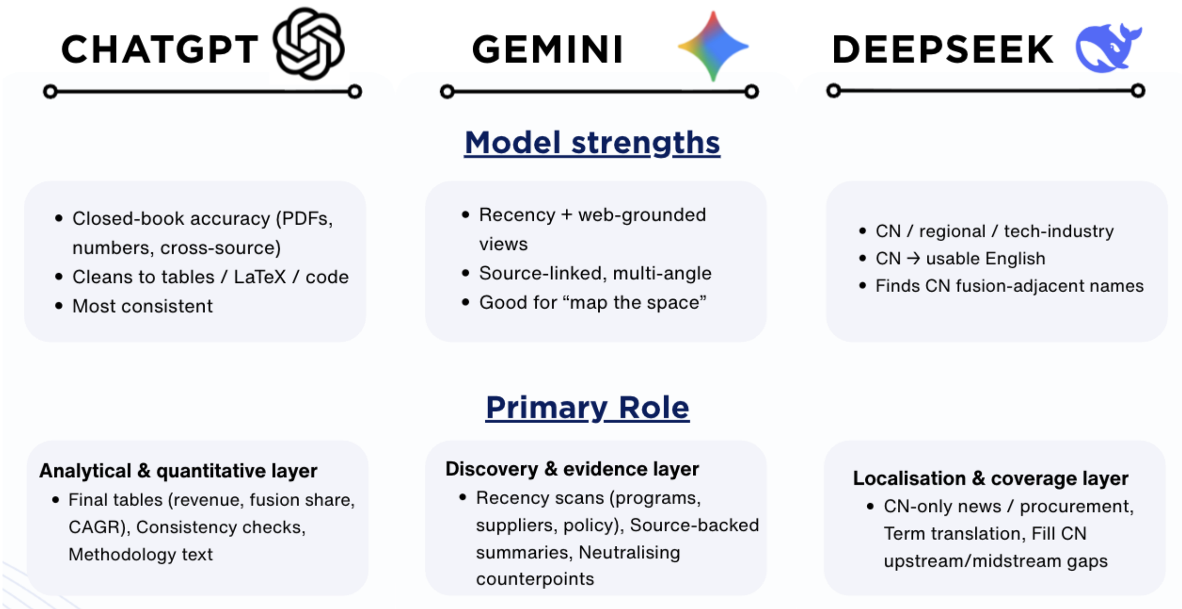
Recency Rating	Meaning	Example
5 – Very recent	Source is from the past month; critical for fast-changing fields	News on interest rate changes
4 – Recent	Source is from the past 2–3 months	Policy updates, quarterly reports
3 – Moderately recent	Source is within the past 6–12 months	Market analysis, annual reports
2 – Outdated	Source is 1–3 years old; fine for slow-changing fields	Textbooks, historical analysis
1 – Very outdated	Source is older than 3+ years in a fast-moving field	Tech trend reports from 2018

Table 3: Recency table of rubrics

Credibility Rating	**Meaning**	**Reasoning**	**Example**
5 – Highly credible	Source is authoritative, peer-reviewed, official or legally accountable	Produced by experts, subject to rigorous checks, minimal bias. In the case of company filings, misrepresentation carries legal and regulatory consequences	Peer-reviewed journals, government statistical reports, BIS working papers, Public company filings (annual reports, quarterly/annual earnings, SEC filing, etc.)
4 – Credible	Reliable but not fully peer-reviewed	Well-established institutions, industry reports, or reputable news	IMF staff papers, McKinsey reports, *Financial Times*
3 – Moderately credible	Useful but requires cross-checking	Some editorial oversight, but possible bias or limited depth	Company press releases, trade publications, Investopedia
2 – Low credibility	Limited verification, strong bias, or conflicts of interest	May present selective or promotional information	Personal blogs, sponsored content, LinkedIn posts without citations
1 – Not credible	No evidence, anonymous/unverified claims	Cannot be trusted for decision-making	Reddit threads, Twitter/X rants, unsourced memes

Table 4: Credibility table of rubrics

G. Slide Snapshots



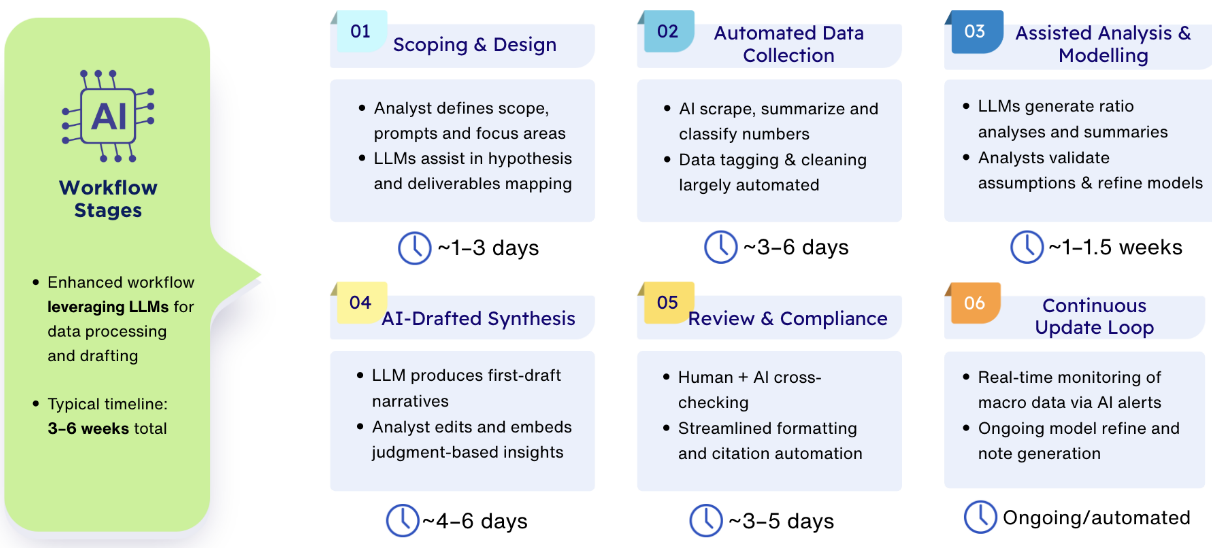
Slide 1: LLM Comparative Strengths

Downstream companies financials

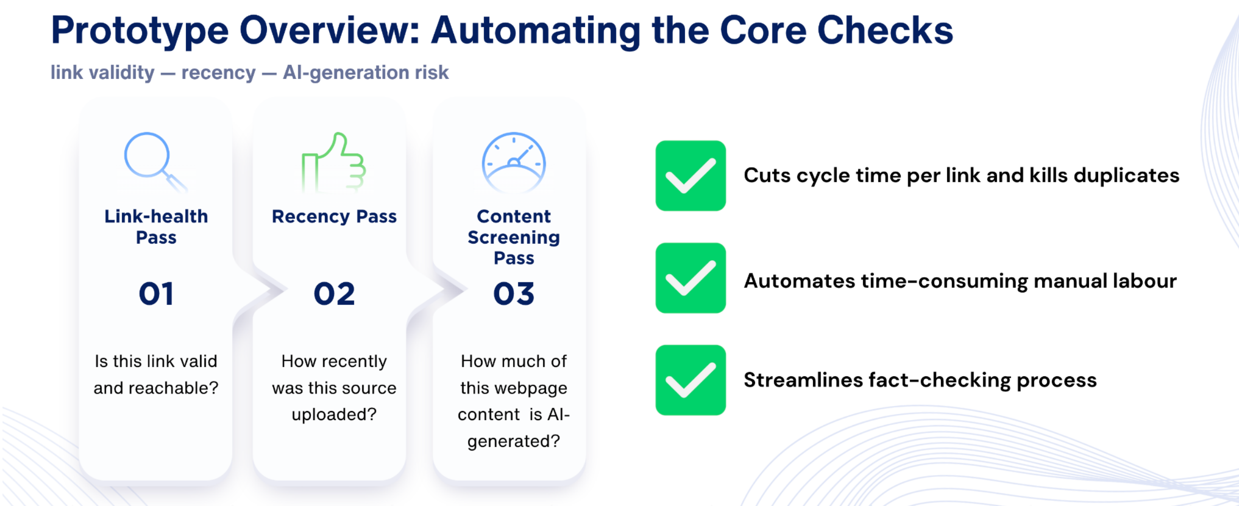
Company	Ticker	Price	Mkt CAP	EV	SALES (TTM)	EBITDA (TTM)	EV/Sales	EV/EBITDA	P/E
VIASAT	VSAT	30.88	4.146B	10.05B ~~9.86B~~	4.564B	1.223B	2.20 ~~2.16~~	8.23 ~~8.06~~	N/A
Echostar	SATS	79.00	22.73B ~~7.23B~~	48.03B ~~48.21B~~	15.453B	2.38B	3.11 ~~3.12~~	20.14 ~~20.20~~	N/A
Iridium	IRDM	18.64	1.98 ~~1.87B~~	3.71B ~~3.6B~~	0.858B	0.423B	4.3 ~~4.2~~	8.76 ~~8.51~~	18.64
Telesat	TSAT	27.76	573.02M ~~0.41B~~	3.21B	0.489B	0.277B	6.55	55.00 ~~11.6~~	N/A
Planet labs	PL	14.85	4.57B ~~0.9B~~	4.32B ~~3.99B~~	0.262B	-VE (-41.612M)	16.45 ~~15.19~~	-6.73 N/A	N/A
Spire global	SPIR	13.11	407.01M ~~0.1B~~	484.46M ~~0.42B~~	99.502M ~~0.100B~~	-VE (-52.648M)	4.87 ~~4.26~~	3.8 N/A	N/A
Trimble	TRMB	80.21	19.09B ~~18.9B~~	20.33B ~~20.13B~~	3.576B	0.638B	5.69 ~~5.63~~	31.87 ~~31.57~~	67.97
BlackSky	BKSY	24.35	863.56M ~~0.11B~~	0.9B (900M)	0.105B	0.137B	8.60	6.56	N/A

Slide 2: Snapshot of LLM Failure

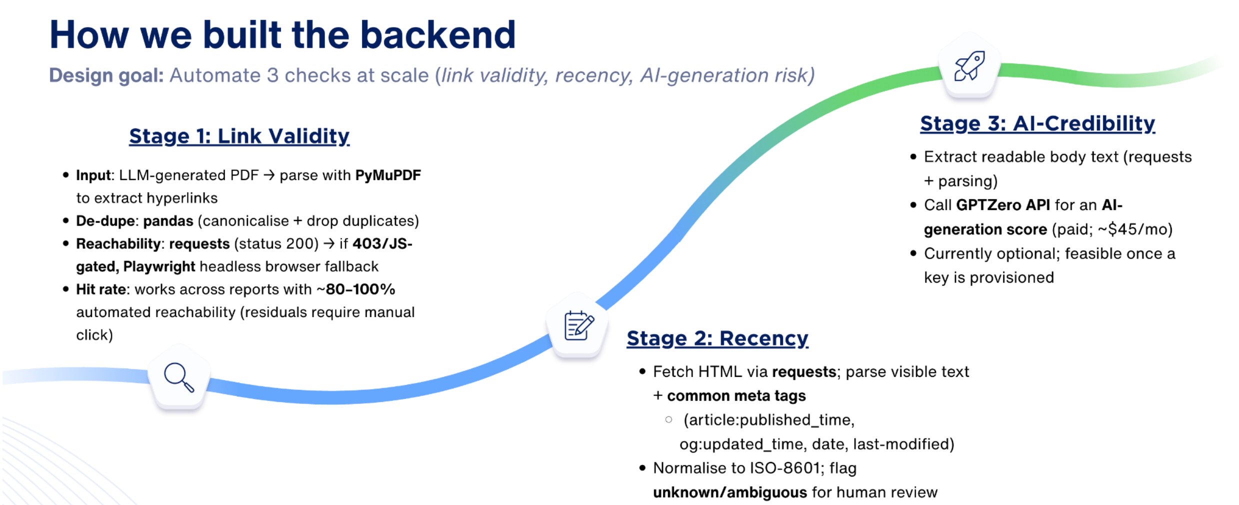
Slide 3: Traditional Investment Research Workflow



Slide 4: LLM-Assisted Workflow



Slide 5: Core Check Automation



Slide 6: 3-Stage Pipeline