A lighthouse stands atop a hill composed of numerous blue, translucent cubes. The lighthouse is illuminated from within, casting a warm glow. The background is a dramatic sky with soft clouds and a hint of sunset or sunrise. The overall scene is surreal and futuristic.

# SCBX AI Outlook 2025

Beaconing the Future of Artificial Intelligence.

SCB<sup>x</sup>





## SCBX AI Outlook 2025: Beaconing the Future of Artificial Intelligence

At the end of 2022, OpenAI's launch of ChatGPT marked a defining moment—the world was introduced to Generative AI (Gen AI), and everything changed.

What once felt like a distant frontier suddenly became a present-day force. AI rapidly embedded itself into both personal routines and enterprise strategies, emerging as a powerful game-changer—streamlining lives, unlocking new efficiencies, and transforming how organizations operate. In pursuit of competitive advantage, businesses across sectors are now racing to understand, adopt, and innovate with AI.

SCBX AI Outlook 2025: Beaconing the Future of Artificial Intelligence is designed to be a lighthouse amid this accelerating wave—offering clarity and direction as leaders navigate the rising tides of technological disruption. This report explores the defining AI trends shaping the year ahead and provides a strategic lens for turning uncertainty into opportunity.

The report is divided into four acts, each highlighting a major force reshaping the AI landscape:

### **ACT I: Two Philosophies, One Future.**

The Battle Between Open-Source and Closed-Source AI Intensifies

### **ACT II: Tiny Titans - Small, but Mighty.**

More Versatile, Smaller, and Smarter: 3 Trends of the Next AI Evolution

### **ACT III: AI at Your Fingertips.**

Agentic AI: Rise of the Agents

### **ACT IV: Not Quite Human, But Almost There.**

Artificial General Intelligence (AGI) and the Unresolved Path to Human-Level AI

The report concludes with an EPILOGUE: The AI Storm – Infinite Impact. Case Studies from Inside the Eye of the Typhoon—featuring real-world case studies from SCBX's use of the "Typhoon" the group's AI Engine across various business units.

As the AI currensurges forward, this report aims to serve not just as a forecast—but as a strategic beacon for those ready to ride the wave and lead.



# Contents



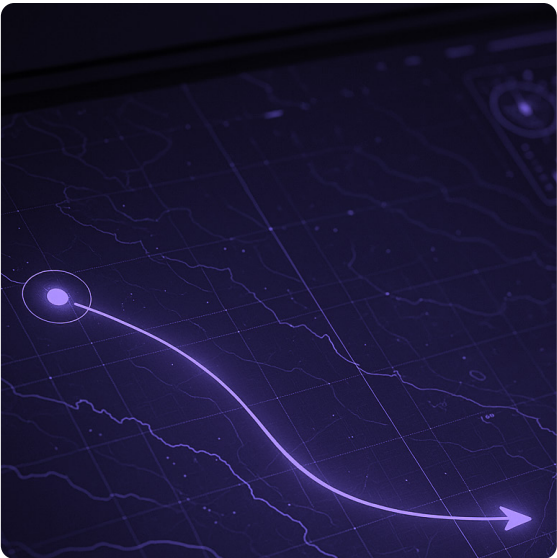
**ACT I:**  
**Two Philosophies, One Future.**  
The Battle Between Open-Source and Closed-Source AI Intensifies

Page **4-9**



**ACT II:**  
**Tiny Titans - Small, but Mighty.**  
More Versatile, Smaller, and Smarter: 3 Trends of the Next AI Evolution

Page **10-16**



**ACT III:**  
**AI at Your Fingertips.**  
Agentic AI: Rise of the Agents

Page **17-21**



**ACT IV:**  
**Not Quite Human, But Almost There.**  
Artificial General Intelligence (AGI) and the Unresolved Path to Human-Level AI

Page **22-26**

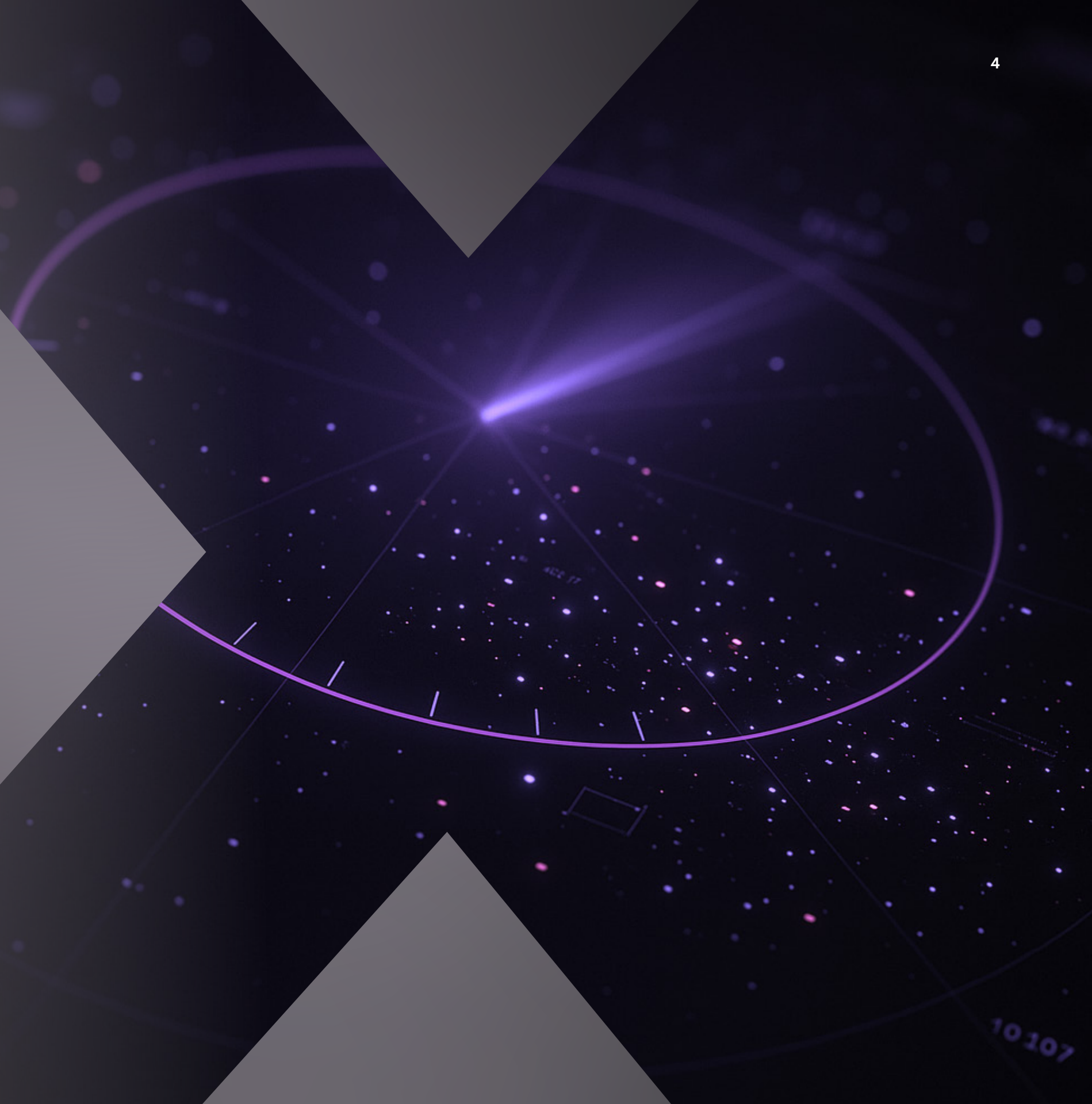


**EPILOGUE:**  
**The AI storm - Infinite impact.**  
Case Studies from Inside the Eye of the Typhoon  
กรณีศึกษาของ SCBX

Page **27**



# ACT I: Two Philosophies, One Future.





# The Battle Between Open-Source and Closed-Source AI Intensifies

The development of advanced artificial intelligence (AI) models is currently dominated by two dominant philosophies open-source and closed-source.

Open-source AI models are transparent by design. Their underlying code and architectural insights are made freely available to the public, empowering developers worldwide to use, modify, and build upon them without restrictions. Closed-source models, by contrast, are developed and tightly controlled by specific organizations. Their internal mechanics, data, and decision-making logic remain proprietary—shielded from external scrutiny.

Well-known closed-source models such as OpenAI’s ChatGPT and Anthropic’s Claude typically exhibit high performance but are often considered “black boxes,” with internal workings that remain obscure to users. On the other hand, notable open-source models such as DeepSeek R1 from DeepSeek and Llama from Meta allow the developer community to participate, leading to rapid and collaborative innovation.

The debate over “Open vs. Closed-Source AI Models” has become a critical topic shaping the direction of the AI industry. Each approach offers distinct advantages and drawbacks. As the capabilities of open-source models increasingly rival those of their closed counterparts, their impact on industries and economies grows—making it equally essential for businesses to adapt and integrate AI solutions appropriately.<sup>[1]</sup>

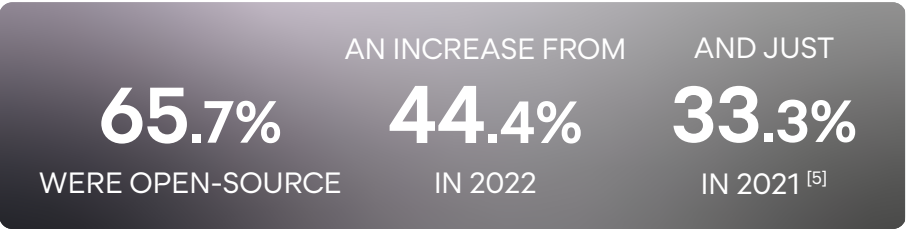
## Impacts and Transformations from Open-Source and Closed-Source Models

Closed-source models developed by major corporations such as OpenAI, Google, and Anthropic initially propelled AI’s advancement—starting around 2018 with the release of GPT-1—and sparked widespread public interest. OpenAI’s ChatGPT, powered by proprietary models like GPT-3.5 and GPT-4, showcased the remarkable potential of AI in conversation and content generation, prompting many organizations to adopt AI technologies in earnest.<sup>[2]</sup>

However, the nature of these closed systems as “black boxes” has raised concerns about transparency and data security. Many organizations hesitate to submit sensitive data into systems they cannot fully control, highlighting the risks associated with depending on closed models that may lack robust internal safeguards.

Conversely, open-source models have significantly altered the landscape of AI by granting public access and encouraging communal development. This collaborative power has led to the rapid emergence of innovative AI models.<sup>[3][4]</sup>

This trend is increasingly evident in findings from the Stanford Institute for Human-Centered Artificial Intelligence (HAI), which reported a record-breaking 149 foundation models launched in 2023 alone. Of these, over



Such statistics point to a swift industry shift toward openness, moving away from domination by a few tech giants. Today, a growing number of organizations and research institutions are releasing models in open-source formats. This has led to a broader dissemination of knowledge and technology, enabling more experimentation and innovation beyond the closed doors of large corporations.

The escalating competition between open- and closed-source models has become a focal point, with many businesses now opting for open-source models to power their chatbots and AI applications—directly challenging paid services based on proprietary models.

Furthermore, open-source models fuel innovation by allowing developers and everyday users to customize tools for unique needs or niche problems without the burden of building models from the ground up.



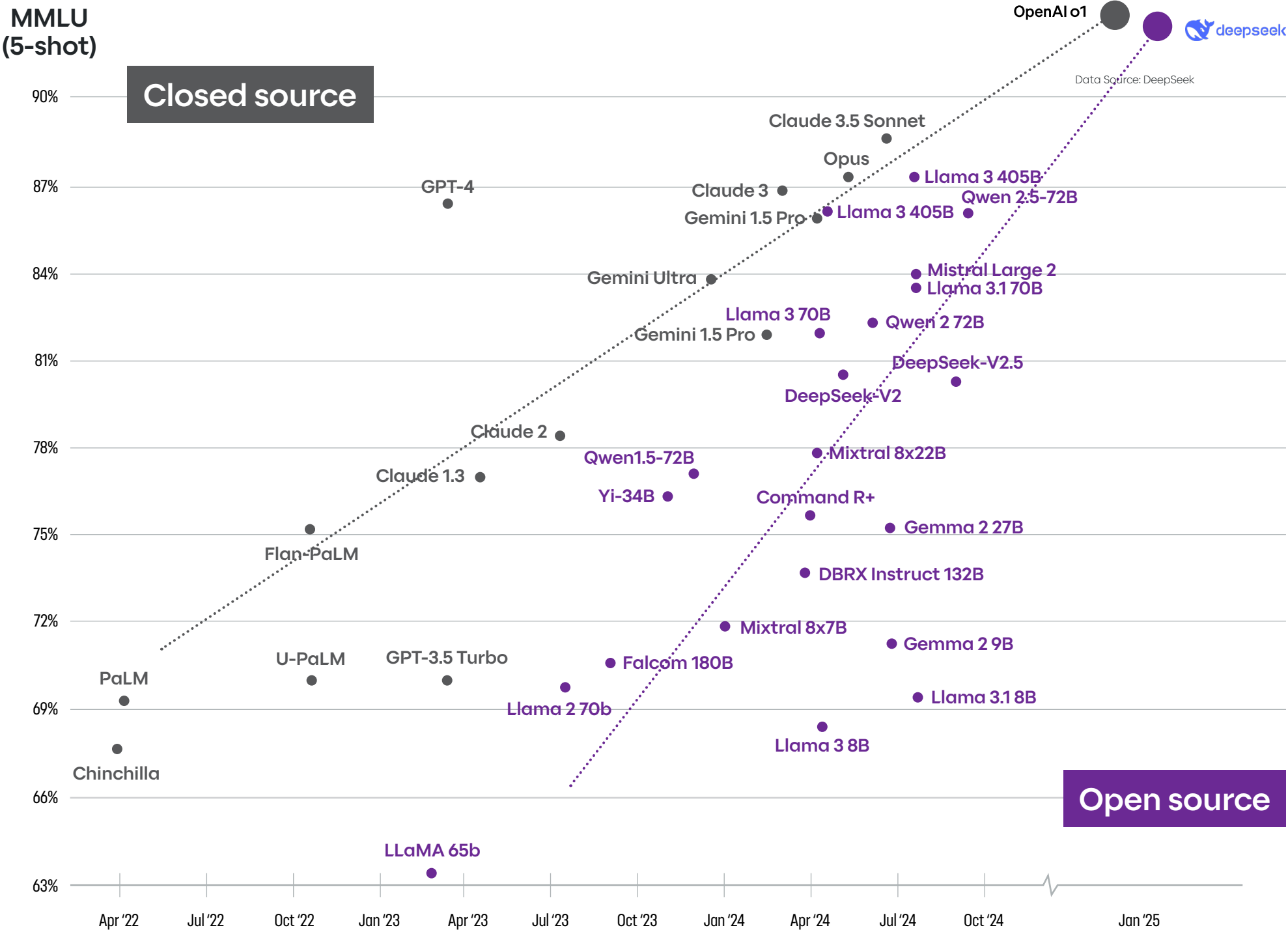
# DeepSeek and the Narrowing Gap Between Open and Closed Models

Although leading closed-source models have historically held an advantage in performance, the gap between closed- and open-source models is narrowing rapidly.<sup>[6]</sup>

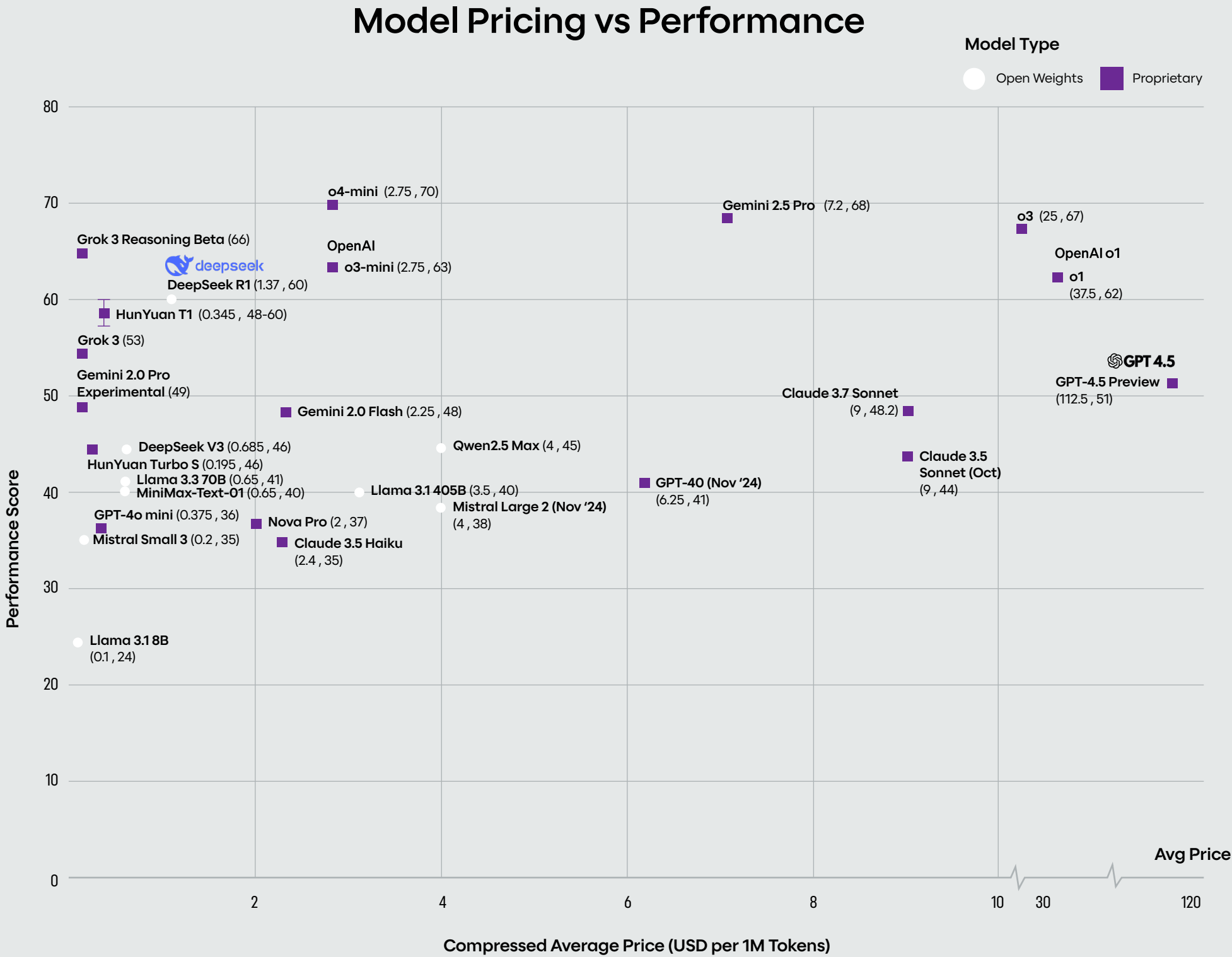
DeepSeek is a prime example of a fast-growing open-source model. Within just two years, it has released multiple versions, including DeepSeek-Coder, DeepSeek-V2, and most recently, DeepSeek R1—demonstrating a consistent trajectory of innovation. A key highlight of DeepSeek R1 is its ability to close the performance gap with closed models—models that typically require significantly higher resources and development costs—more quickly than expected. This is particularly evident in its performance on the MMLU (Massive Multitask Language Understanding) benchmark <sup>[7]</sup>, a widely recognized standard for evaluating AI models on language understanding, general knowledge, and advanced problem-solving capabilities.

DeepSeek R1 scored comparably to OpenAI’s “o1,” one of the leading closed-source models, suggesting that the performance of open-source models is now catching up at an accelerated pace. Notably, DeepSeek was reportedly developed with a budget of just USD 5.6 million (approximately THB 190 million). However, many analysts believe this figure may underestimate the true cost, as it likely excludes various additional expenses such as multiple training cycles, data collection, infrastructure, and operational costs. Taking these into account, the total investment may have been significantly higher than reported.<sup>[8]</sup>

The impact of DeepSeek has already rippled across technology stock markets following its launch on January 27, 2025. Companies in the datacenter energy sector—such as Vertiv and GE Vernova Inc. (GEV) saw stock prices fall by as much as 30%. Semiconductor stocks, including AI-related growth stocks like Nvidia and TSMC, dropped by over 20%. In contrast, cloud service providers such as Google, Microsoft, Amazon, and Meta were not significantly affected. Ultimately, the stocks that gained value were those associated with the application layer—such as Adobe, Salesforce, and SAP.







Between January 28–30, 2025, several of the stocks that had sharply declined began to rebound; however, they had yet to recover to their pre-DeepSeek launch levels. Cloud service providers remained relatively stable, with certain companies, such as Google and Meta, even experiencing a slight increase.

The fact that DeepSeek was able to develop a model with performance comparable to closed-source models within just two years reflects a growing trend—open-source models may soon possess capabilities equivalent to their closed counterparts. This development could lead to accelerated AI advancement, cost reduction, and broader accessibility of AI technologies across various industries. The ongoing competition between the two model types is likely to reshape the AI landscape rapidly—in terms of cost, adoption, and continuous development.



For instance, as illustrated, OpenAI’s GPT-4.5 demonstrates strong control over hallucination issues (the generation of inaccurate information). However, the processing cost per 1 million tokens is approximately USD 112.5, with a performance score of 51 (performance being assessed based on reasoning, general knowledge, mathematics, and programming capabilities). In contrast, DeepSeek R1 processes 1 million tokens at a cost of only USD 1.37, while achieving a performance score of 60.<sup>[9]</sup>



Beyond performance and cost, another emerging trend now being actively pursued by both sides is the push for speed—specifically, the acceleration of reasoning-based responses. In earlier model generations, generating well-reasoned answers required significant computational power during inference. As a result, the speed of logical, coherent responses is becoming another critical metric by which future AI models will be evaluated.

To say that “open-source AI models are closing in fast”—with the potential to match or even surpass closed-source models in certain areas—would not be an overstatement.

Nonetheless, leading closed-source models still retain clear advantages in specific tasks, such as advanced language understanding or managing extended contexts—areas in which models like GPT-4 excel. These advantages stem from the vast resources and proprietary, large-scale datasets used to train such models, which are available exclusively to the organizations that own them.

However, given the exponential rate at which open-source models are currently progressing—doubling in public releases year over year—and the fact that they are continually “refined” by developers around the world, the performance gap is steadily narrowing.

Moreover, open-source models offer distinct advantages in flexibility and customization, which, in some cases, make them better suited to specialized applications than general-purpose closed models. For example, a company seeking an AI assistant for analyzing domain-specific documents may choose to fine-tune an open-source model on its proprietary data, resulting in an AI system that deeply understands its specific context—something a generic closed-source model may be unable to deliver with the same level of precision.

# Conclusion: Open-Source AI Models Are Transforming the Economics of AI

The competition between open-source and closed-source AI is not only reshaping the technological landscape, but also redefining the economics of AI through the principles of Economies of Scale and Economies of Speed.



## 1. Economies of Scale:

Open-source models significantly reduce the cost of AI development by distributing workloads across a global community of developers. Unlike closed-source models, which require substantial investment from a single entity, open-source models allow businesses to adopt and fine-tune existing models without starting from scratch—lowering costs and accelerating the commercial deployment of AI.



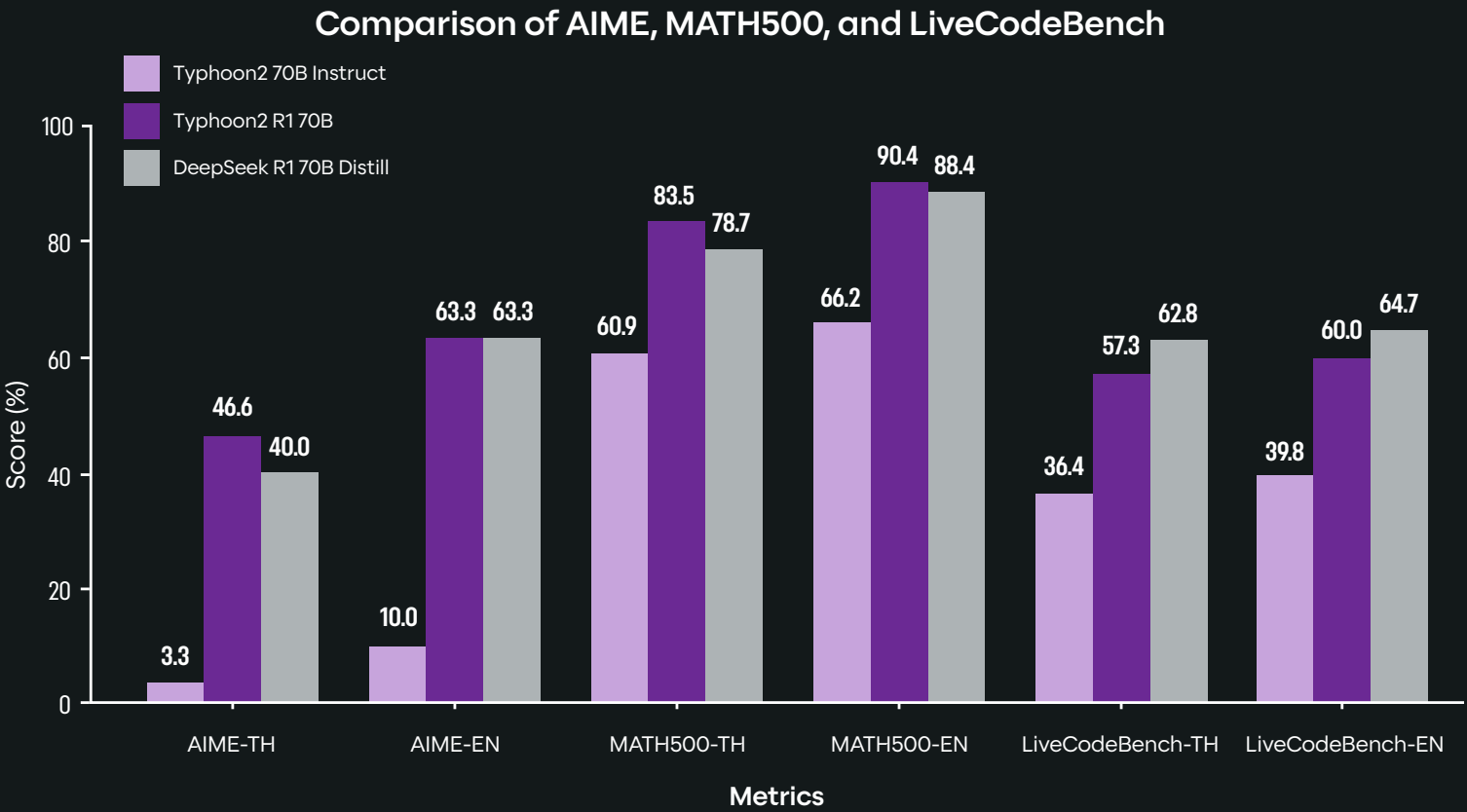
## 2. Economies of Speed:

By making source code publicly available, the development cycle of AI is dramatically accelerated. The ability for a broad developer base to collectively refine models leads to faster problem-solving and innovation compared to the limited internal processes of closed-source development. This global adaptability and shared learning enable open-source models to reach performance parity with closed-source models in significantly shorter timeframes.

A striking example is Typhoon2 R1 70B, an open-source model by SCB10X, developed with a budget of just USD 120 (approximately THB 4,000) and within a remarkably short timeframe of 15 hours. This was achieved by merging two models—Typhoon2 70B, known for its Thai language proficiency, and DeepSeek R1 70B Distill, recognized for its strong reasoning capabilities. The resulting model exhibited significant improvements in both reasoning and Thai language understanding, surpassing the performance of DeepSeek R1 in key areas.<sup>[10] [11]</sup>



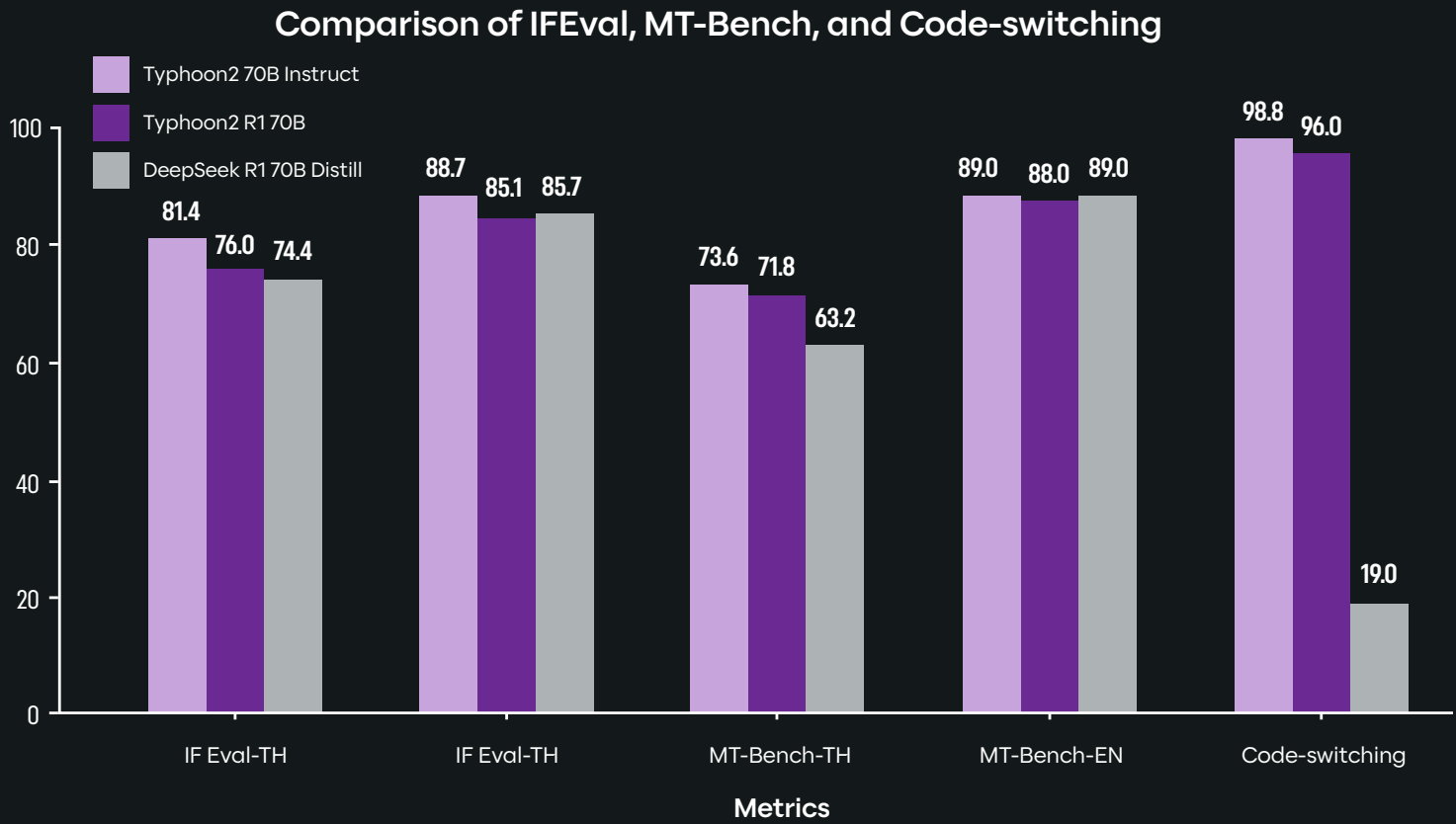
## Reasoning Benchmarks



Closed-source models offer top-tier performance and reliable support, but this comes at the cost of limited transparency and a concentration of power in the hands of a few developers.

On the other hand, open-source models promise collaboration and equitable access, and they are increasingly demonstrating capabilities that rival those of closed-source models in many areas. Striking a balanced integration between these two approaches will be essential for the continued growth of the AI industry and for maximizing its benefits to the economy and society at large.

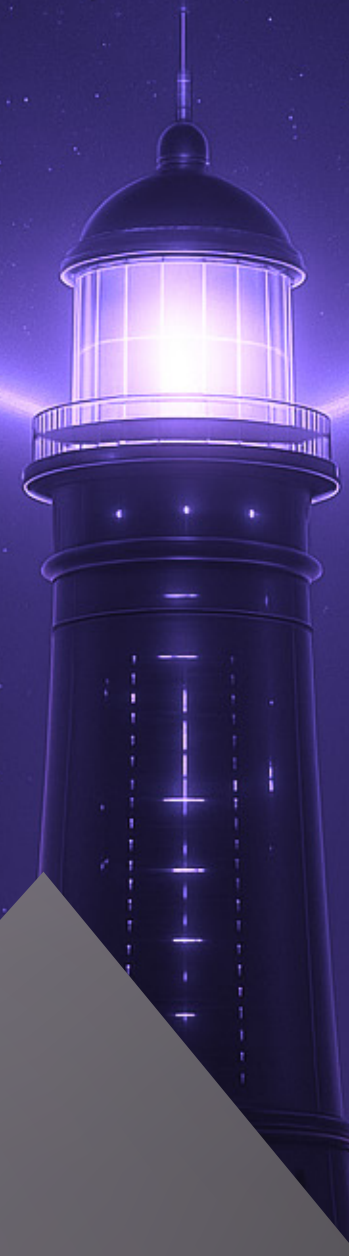
## Instruction Following/Code-Switching Benchmarks



Ultimately, whether it is open-source or closed-source AI models, the shared objective is to harness artificial intelligence for the greatest possible benefit to humanity. The coexistence of both approaches will enable this goal to be achieved more rapidly and comprehensively—while maintaining a healthy equilibrium between innovation, transparency, and accountability in the future.



# ACT II: Tiny Titans – Small, but Mighty.





# More Versatile, Smaller, and Smarter: 3 Trends of the Next AI Evolution

Artificial Intelligence (AI) continues to evolve at a remarkable pace, with three emerging trends redefining the technological frontier:

1. Versatile Multimodal models capable of interpreting and responding to diverse types of data.
2. Smaller, task-specific models that are more efficient and adaptable.
3. Scaling Intelligence for Improved Reasoning, shifting focus toward more efficient approaches—pursuing deeper reasoning capabilities without ever-larger model sizes.

Each of these trends holds distinct implications—reshaping how businesses operate, how industries innovate, and how AI integrates into everyday life. Collectively, they signal a shift toward more practical, scalable, and human-aligned AI—pushing the boundaries of what intelligent systems can accomplish in the real world.

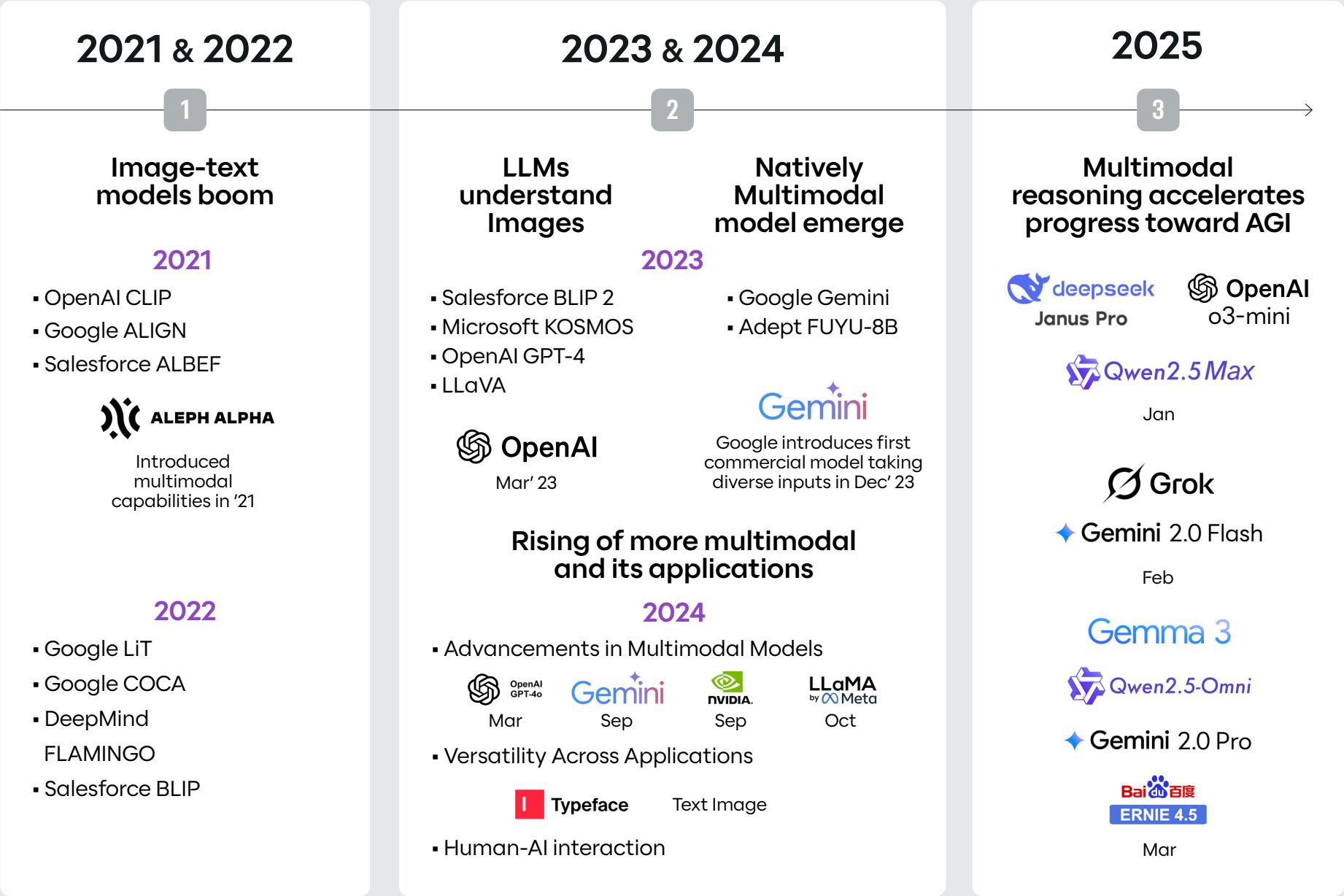
## 1 Versatile Multimodal Models

Multimodal AI models are capable of processing multiple types of data simultaneously, such as text, images, audio, and video. This enables the system to comprehend context more deeply and perform a broader range of complex tasks. Multimodal large language models (Multimodal LLMs) are gaining strong momentum, as they offer richer contextual understanding and greater flexibility for a wide variety of real-world applications.

This trend has led to the development of new models that can interpret diverse data formats and has paved the way for their integration with AI agent systems, enabling them to operate autonomously and intelligently. Demis Hassabis, CEO of Google DeepMind, stated that “the combination of multimodal capabilities and AI agents is one of the key steps toward AGI (Artificial General Intelligence).”<sup>[1]</sup>







As illustrated in, this trend has been gaining momentum since 2021–2022, which marked the beginning of the growth era for models capable of handling multiple data modalities. Since then, it has continued to expand and is expected to maintain its upward trajectory into 2025 and beyond. The increasing development and widespread application of multimodal models—along with advancements in multimodal reasoning, or the ability to draw inferences from diverse data types—are seen as accelerating progress toward achieving AGI (Artificial General Intelligence).<sup>[2] [3] [4]</sup>

## Impact on the Economy and Industry

Multimodal models possess significant potential for generating economic value due to their ability to address problems across multiple industries using a single model. For instance, in the training and education sector, they can be used to create intelligent teaching assistants capable of understanding both text and visual content—thereby enhancing the explanation of complex topics for learners.

One example is SCB X Public Company Limited, which has implemented this technology in its Smart RM Training system. This system simulates customer interactions to train sales staff by using a voice-to-voice, real-time model, representing an advanced application of multimodal LLM. The result is that employees develop skills more efficiently and comprehensively, while the company also benefits from reduced training costs.

In the medical field, these models can assist doctors in diagnosis by analyzing X-rays or MRI results alongside textual medical reports. In customer service, intelligent chatbots capable of understanding both speech and images from customers allow for more accurate and context-aware responses. This trend enhances operational efficiency across sectors and elevates user experiences, as AI becomes capable of understanding context derived from multiple data sources.

## Future Outlook

This trend is expected to become increasingly significant. Future models may be capable of receiving input from various sensory modalities—such as audio, visual, text, and IoT sensor data—and processing them collectively to produce coherent and deeply reasoned outputs. For example, factory assistant robots may be able to visually observe a situation while simultaneously responding to verbal instructions, or data analysts may naturally explain numerical data in conjunction with graphs and images.

In addition, the integration of multimodal models with AI agents—specifically through Full-Duplex in Agentic AI, or the ability for AI agents to engage in simultaneous two-way communication—not only enhances the efficiency and naturalness of human-AI interactions, but also allows the AI to rapidly analyze and process data. This capability leads to more accurate and timely decision-making, which is especially valuable in applications requiring immediate response, such as customer service, technical support, or emergency assistance.<sup>[5]</sup>

This approach increases the usability of AI without requiring human intervention at every step, fundamentally transforming business operations across various sectors. However, future challenges include managing the computational and energy costs associated with these large models, as well as efficiently handling diverse types of data.

2

## Small Models for Specific Tasks

The second trend centers on the development of small AI models or Small Language Models (SLMs) that are fine-tuned for high efficiency in specialized tasks. These models have significantly fewer parameters compared to large models such as GPT-3 or GPT-4, but they are trained for expertise in specific domains. As a result, they offer faster processing, lower resource consumption, and easier, more cost-effective fine-tuning.

These advantages are driving the growing popularity of small models. They can also be deployed directly on edge devices, such as smartphones or IoT equipment, which enhances accessibility while preserving data privacy, as processing takes place on the user's device rather than through centralized servers.



## Impact on the Economy and Industry

Specialized small models provide opportunities for small and medium-sized enterprises (SMEs) to adopt AI more broadly, without requiring large-scale infrastructure investments. This helps reduce the technological gap between large corporations and smaller players.

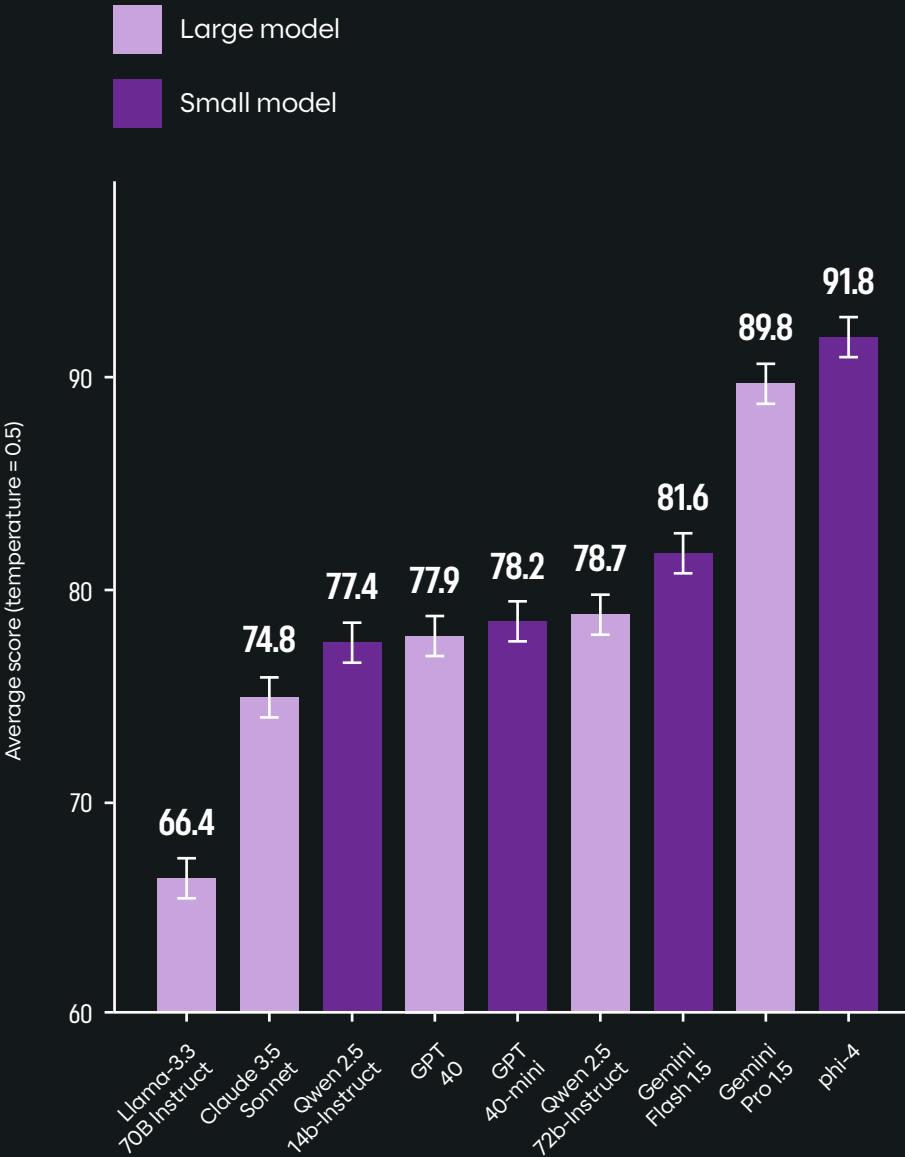
For example, in the smartphone industry, intelligent voice assistants on newer mobile devices now process voice commands locally, eliminating the need to transmit data to centralized servers. This results in faster response times and prevents users' voice data from being exposed externally. In healthcare, dermatology diagnostic apps running on tablets or smartphones in remote areas can screen patients without internet connectivity. Additionally, task-specific small models often offer higher transparency and security, as their smaller codebase and parameter count make them easier to audit and control.

Despite their smaller size, many specialized small models demonstrate impressive performance, rivaling or even surpassing larger models in certain applications. A recent study by Microsoft introduced Phi-4, a model with only 14 billion parameters, which exhibited strong capabilities in complex reasoning and outperformed larger models like Gemini Pro 1.5 in mathematical tasks, as shown in the referenced image. Phi-4 was also designed with a focus on Responsible AI from the outset.

Another notable example is the startup Mistral AI, which released Mistral 3B and 8B — containing 3 billion and 8 billion parameters, respectively. These small models have emerged as top performers on standard benchmarks, outperforming other models in similar size categories.<sup>[6] [7]</sup>

Innovations like these indicate that “small AI models can compete with large ones in specialized tasks, save resources, and run on client-side devices”—a concept that is transforming industry perspectives by challenging the assumption that only large models are powerful. Moreover, MIT Technology Review listed small language models among its “Top 10 Breakthrough Technologies of 2025,” highlighting that after the rapid progress brought by large language models in previous years, it is now time to focus on smaller, smarter models that can enable wider real-world deployment of AI.

## Average performance on November 2024 AMC-10/12 tests







## Future Outlook

We will witness an increasing number of small, specialized models tailored to a wide variety of tasks—such as domain-specific coding assistants, medical data analysis models focused on particular diseases, or product recommendation engines designed for niche customer segments.

The competition among technology companies will shift toward determining who can develop small models that perform as well as large models in specialized tasks, while consuming significantly fewer resources. Achieving this would provide a substantial competitive business advantage.

Arvind Krishna, CEO of IBM, emphasized that if we can train models to achieve similar levels of performance at just one-hundredth of the cost and run them on a smaller infrastructure, every organization will be compelled to adapt in order to remain competitive.<sup>[8]</sup>

This transformative trend will further accelerate the growth of open-source AI, as small models are easier to build upon. Developer communities can collaboratively enhance and share these models more efficiently. As small models continue to become more intelligent, everyday users may soon gain access to high-quality AI without relying on large corporations or expensive hardware. This will help democratize AI, expanding its opportunities and benefits across society.

3

# Scaling Intelligence for Improved Reasoning

The third trend challenges the prevailing notion of scaling up—the idea that increasing model size leads to better AI performance—which has dominated the field over the past decade.<sup>[9]</sup>

New generations of large language models (LLMs), such as GPT-3 and GPT-4, contain massive numbers of parameters and are trained on vast datasets. This scaling approach has been shown to enhance language understanding and reasoning capabilities, aligning with Scaling Laws, which suggest that increasing model size, dataset volume, and compute power typically leads to exponential improvements in performance.

However, as models grow larger, signs of the law of diminishing returns have begun to emerge. That is, the rate of improvement in AI capabilities slows down, even with continued increases in parameters and training data. This raises concerns over the rising costs and diminishing benefits of further scaling.

In addition, data availability and training costs are becoming major barriers. Researchers predict that, within a few years, the world may face a shortage of high-quality training data. Estimates suggest that the supply of high-quality textual data could be exhausted by 2026. If this trend continues, it could significantly hinder AI advancement unless new strategies are adopted.

Thus, the conventional path of scaling by simply increasing model size may be reaching its limits. The AI community must seek alternative approaches to building more intelligent systems without relying solely on size expansion.



## Impact on the Economy and Industry

This scaling challenge has two major implications. On one hand, developing state-of-the-art models with advanced reasoning capabilities (such as GPT-4) requires massive investment—in terms of computational resources (e.g., thousands of GPUs and substantial electricity consumption) and enormous datasets. Only a few large corporations possess the resources to pursue such efforts, potentially leading to technological monopolies and widening inequality, as smaller companies and developing nations may struggle to keep pace.

On the other hand, as the limitations of scaling up become more evident, both industry and researchers are turning to alternative, more efficient development methods. If more effective strategies are found, they could drastically reduce the cost of AI development and open the field to new players, yielding long-term economic benefits.



## New Approaches to AI Development

Beyond increasing model size, one promising approach is to allocate more computational power during inference rather than during training. This strategy involves allowing the model to “think longer” when faced with complex problems, rather than responding instantly as current models typically do. This concept is often referred to as Test-time Compute, or “on-the-fly computation” during real-time use.

For example, instead of generating a single answer immediately upon receiving a query, the model could explore multiple potential solutions internally—through various steps or reasoning paths—evaluate them, and select the most appropriate answer to present to the user.

This method mimics the way humans take time to think through complex problems and could enable models to solve multi-step reasoning tasks more effectively, without any increase in model size. Several research institutions have begun experimenting with this approach. OpenAI, for instance, has developed a model series under the codename “o1,” designed for multi-step reasoning during response generation and trained with expert-curated data to enhance deep reasoning capabilities.





The result is a model that is not particularly large but is capable of solving complex problems more effectively. This approach is seen as one that could fundamentally reshape the competitive dynamics of the AI industry, as it may no longer be necessary to rely on massive computational chips or extensive energy consumption. If models can become more intelligent through alternative means, companies may begin to compete based on the quality of algorithms and training techniques rather than solely on the scale of their resources.

Beyond the use of inference time, other approaches are also emerging, such as the use of synthetic data to supplement training (to address the shortage of real-world data), or improvements in model architecture to enhance learning efficiency without increasing parameter count, such as models that simulate step-by-step reasoning.



## Examples and Future Outlook

Ilya Sutskever, co-founder of OpenAI and Safe SuperIntelligence and an early proponent of the “scale equals success” paradigm, recently acknowledged in an interview with Reuters that the traditional approach of simply increasing model size is reaching saturation. He emphasized the need to return to an “age of wonder and discovery” in search of new ideas to build more intelligent AI. His earlier statement that “if you have a sufficiently large dataset and you train a very large neural network, success is guaranteed” may no longer hold true in this decade.<sup>[10]</sup>

This shift implies that the future of AI will not be driven by size alone but will increasingly focus on qualitative efficiency. In the near future, we are likely to see a convergence of multiple approaches—for example, mid-sized models trained in multi-step reasoning, combining real and synthetic data, and utilizing a mixture of specialized smaller models rather than relying on a single massive one (mixture of experts).<sup>[11]</sup>

These strategies could significantly reduce the burden of computational and energy resources. If successful, they would yield benefits for both the economy and the environment while opening opportunities for new players to develop intelligent models without being hindered by the prohibitive costs that have previously created barriers to entry.

These are the three major trends shaping the next era of AI. They reflect a direction in which the AI industry is striving to expand the technological frontier toward models that can receive and process more diverse data, become smaller, smarter, and more accessible. Multimodal models aim to equip AI with a holistic understanding of the world through various data types, thereby enhancing its capabilities in handling complex real-world scenarios. Small, specialized models aim to broaden AI adoption by reducing resource constraints and enabling all sectors to access AI that is well-suited to their specific needs. Meanwhile, scaling model computation during inference for improved reasoning represents an effort to push the boundaries of AI performance further without being limited by traditional methods.

These three trends do not develop in isolation but are mutually reinforcing in the creation of future AI that is versatile, resource-efficient, and capable of solving complex problems through reasoning. By staying informed of and adapting to these trends, businesses and the economy as a whole can fully leverage the benefits of AI, while simultaneously preparing to address the new challenges that will emerge, including:

1. The use of AI in governance and the development of Ethical AI
2. The advancement of AI under constraints on data and development resources
3. The socioeconomic impacts of AI



# ACT III: AI at Your Fingertips.





# Agentic AI: Rise of the Agents

Over the past two years, artificial intelligence has made remarkable strides—particularly through the rise of Large Language Models (LLMs), which have revolutionized how we interact with machines. Yet most AI in use today still falls under the category of Generative AI: systems that respond to prompts by generating content but remain fundamentally reactive and instruction-bound.

That's now changing. Enter Agentic AI—a more advanced form of AI capable of perception, reasoning, planning, and autonomous execution. Rather than merely following human instructions, these systems understand goals and take initiative using tools and contextual knowledge. It's the difference between a task-oriented assistant and a strategic partner with decision-making ability.

Global organizations like the World Economic Forum and leading institutions such as Stanford HAI<sup>[1]</sup>, MIT Sloan<sup>[2]</sup>, along with industry experts, unanimously agree that Agentic AI will be a key trend dominating the AI market in 2025, supported by industry leaders, academic institutions, major technology companies, and startups. At the World Economic Forum 2025 in Davos, extensive discussions took place regarding the direction of AI, with representatives from various regions conferring with world leaders and policymakers about AI trends. The conversations focused more on practical real-world applications rather than merely on its potential.

In parallel, MIT Sloan Management Review presented an article titled “Five Trends in AI and Data Science for 2025” by Thomas H. Davenport and Randy Bean, which stated that ‘the trend of Agentic AI in 2025 deserves close attention from leaders,’ emphasizing that Agentic AI is one of the most important trends organizational leaders should focus on in 2025.

Another significant difference is how they respond to uncertainty or new information. Gen AI typically requires humans to input new data or adjust commands, whereas Agentic AI can adapt and find new solutions on its own to a certain extent.

Take the example of investment planning: a Generative AI might offer general recommendations on investment strategies based on received data; however, if markets fluctuate or significant news impacts occur, humans would need to input new data to obtain updated recommendations. Meanwhile, Agentic AI can monitor real-time market data, assess impact, and automatically adjust recommendations.

This ability to operate dynamically is especially critical in fast-moving industries like banking and finance, where decisions must be fast, accurate, and responsive to real-time volatility.

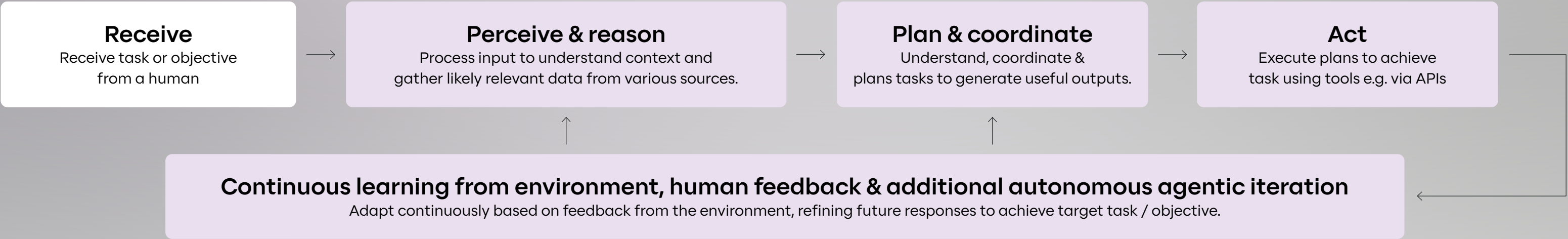


# A Gen AI Approach



# Agentic AI Approach

An Agentic, “Human-like” approach to task completion





# SCBX Agentic AI Financial Advisory Platform

SCBX recognizes the potential of Agentic AI and has developed it into an investment advisory platform that can clearly demonstrate competitive advantages.

## A Gen AI Approach

ChatGPT o3-mini-high

MSFT ผลตอบแทนเป็นยังใจ หลังปีใหม่ นับตั้งแต่วันแรกของเดือนมกรา

ณ วันที่ 25 กุมภาพันธ์ 2025 ราคาหุ้นของ Microsoft Corporation (MSFT) ปิดที่ 404.00 ดอลลาร์สหรัฐ

FINTEL.IO

เมื่อเปรียบเทียบกับราคาปิด ณ วันที่ 2 มกราคม 2025 ซึ่งอยู่ที่ 410.92 ดอลลาร์สหรัฐ 

TH.INVESTING.COM

 ราคาหุ้น MSFT ลดลงประมาณ 6.92 ดอลลาร์สหรัฐ หรือประมาณ 1.68% นับตั้งแต่ต้นปี

อย่างไรก็ตาม ควรสังเกตว่าราคาหุ้นมีความผันผวนตลอดช่วงเวลาดังกล่าว โดยในวันที่ 17 มกราคม 2025 ราคาปิดสูงสุดที่ 429.03 ดอลลาร์สหรัฐ 

MICROSOFT.GCS-WEB.COM

ดังนั้น นับตั้งแต่ต้นปี ราคาหุ้น MSFT มีแนวโน้มลดลงเล็กน้อย

Grok3

Grok 3 beta

MSFT ผลตอบแทนเป็นยังใจ หลังปีใหม่ นับตั้งแต่วันแรกของเดือนมกรา

1 post

15 web pages

จากข้อมูลล่าสุด ณ วันที่ 19 กุมภาพันธ์ 2025 ราคาหุ้นของ Microsoft (MSFT) ปิดที่ 414.77 ดอลลาร์ ซึ่งเพิ่มขึ้น 19% จากต้นปี 2025 (นับตั้งแต่วันที่ 1 มกราคม 2025) โดยราคาหุ้นเริ่มต้นปีที่ประมาณ 348 ดอลลาร์ (คำนวณจากข้อมูลการเติบโต 19% และราคาปัจจุบัน) อย่างไรก็ตาม ผลตอบแทนอาจผันผวนได้เนื่องจากมีรายงานว่าหุ้น Microsoft ลดลง 1.5% นับตั้งแต่ต้นปี 2025 และมีการปรับลดลง 6.2% หลังจากประกาศแนวโน้มรายได้ที่อ่อนแอในเดือนมกราคม ดังนั้น ผลตอบแทนโดยรวมนับตั้งแต่ต้นปีแสดงถึงการเติบโตที่แข็งแกร่ง แต่ก็มีช่วงเวลาที่ผันผวนตามสถานการณ์ตลาดและการรายงานผลประกอบการ กรุณาตรวจสอบข้อมูลล่าสุดจากแหล่งข้อมูลทางการเพื่อการตัดสินใจที่แม่นยำยิ่งขึ้น เนื่องจากราคาหุ้นอาจเปลี่ยนแปลงได้ตลอดเวลา

TS

MSFT ผลตอบแทนเป็นยังใจ หลังปีใหม่ นับตั้งแต่วันแรกของเดือนมกรา

Claude 3.7 Sonnet

ผมไม่มีข้อมูลที่ปัจจุบันเกี่ยวกับผลตอบแทนของหุ้น Microsoft (MSFT) หลังปีใหม่ 2025 นับตั้งแต่วันแรกของเดือนมกราคม 2025

ความรู้ของผมมีข้อจำกัดถึงเดือนตุลาคม 2024 จึงไม่สามารถ ให้ข้อมูลที่เป็นปัจจุบันเกี่ยวกับราคาหุ้นหรือผลตอบแทนของ Microsoft ในช่วงเวลาดังกล่าวได้

สำหรับข้อมูลผลตอบแทนล่าสุดของหุ้น Microsoft ผมแนะนำให้คุณตรวจสอบจาก:

- เว็บไซต์ข้อมูลการเงินอย่าง Yahoo Finance หรือ Bloomberg
- แอปพลิเคชันซื้อขายหุ้นที่คุณใช้งาน
- เว็บไซต์นักลงทุนสัมพันธ์ของ Microsoft โดยตรง

## SCBX Agentic AI Approach

MSFT ผลตอบแทนเป็นยังใจ หลังปีใหม่ นับตั้งแต่วันแรกของเดือนมกรา

Processing ...

Getting symbol information ...

Conducting scenario analysis ...

Your Investment Buddy

Scenario Analysis

MSFT : Performance since the first day of January

Historical Statistics

n_period	count	win_count	hit_rate	avg_return	median_re
5	2294	1333	58.0%	0.0%	0
10	2289	1408	62.0%	1.0%	1
20	2279	1505	66.0%	2.0%	2

Simulation

Date Range

2015-02-23

2025-02-24

2015-02-23

2025-02-24

Stats

MSFT ผลตอบแทนเป็นยังใจ หลังปีใหม่ นับตั้งแต่วันแรกของเดือนมกรา

จากการวิเคราะห์สถิติการเทรดของหุ้น Microsoft (MSFT) ตั้งแต่วันแรกของเดือนมกราคม พบว่า:

1. สัญญาณ 'Buy': มีการแนะนำให้ซื้อเนื่องจากอัตราการทำกำไร (hit rates) สูงกว่า 50% ในทุกช่วงเวลา (5, 10, และ 20 วัน) โดยเฉพาะในช่วง 20 วันมีอัตราการทำกำไรถึง 66.0%

2. ผลตอบแทนเฉลี่ยเป็นบวก: ผลตอบแทนเฉลี่ยในทุกช่วงเวลาเป็นบวก โดยเฉพาะในช่วง 20 วันมีผลตอบแทนเฉลี่ย 2.0%

3. อัตราส่วนความเสี่ยง-ผลตอบแทน (Risk-Reward Ratio): มีค่าเกิน 1 ในทุกช่วงเวลา โดยเฉพาะในช่วง 20 วันมีค่า 1.21 ซึ่งแสดงว่าผลตอบแทนเฉลี่ยที่ได้จากการทำกำไรสูงกว่าการขาดทุน

4. ผลตอบแทนสูงสุดเทียบกับการขาดทุนสูงสุด: ผลตอบแทนสูงสุด (13.0%, 17.0%, และ 28.0%) สูงกว่าการขาดทุนสูงสุด (-14.0%, -22.0%, และ -27.0%) ซึ่งแสดงว่ามีโอกาสทำกำไรสูงกว่าความเสี่ยงที่จะขาดทุน

จากข้อมูลเหล่านี้ สรุปได้ว่าหุ้น MSFT มีแนวโน้มดีในการลงทุนหลังจากวันแรกของเดือนมกราคม

Ask Eureka ...

The platform in question is an investment advisory system powered by Agentic AI. When compared to the use of general-purpose Generative AI tools such as ChatGPT or Claude 3.7 Sonnet for stock analysis, the differences are clearly pronounced.

SCBX conducted a performance comparison between the two systems in analyzing Microsoft (MSFT) stock by posing the same investment-related questions to both a general-purpose Gen AI and SCBX's Agentic AI. The results revealed a significant difference in the quality of analysis and the ability to deliver actionable recommendations.

- The general-purpose Gen AI, when prompted to analyze MSFT stock, typically:
  - Conducts a basic web search but retrieves incomplete information.
  - Fails to provide concrete investment recommendations.
  - Has limited and outdated knowledge.
  - Cannot perform in-depth analytical reasoning.

The outcome from Gen AI often consists of surface-level statements such as "Microsoft had a strong quarterly performance" or "Microsoft is a fundamentally strong company," which are insufficient for informed investment decision-making. Moreover, the provided information may be outdated and lacks the essential context required for investment evaluation.

In contrast, SCBX's Agentic AI approach employs a more complex and structured process that produces insights of far greater value:



### Input Interpretation

Accurately understands the query, such as the request for an analysis of MSFT stock.



### Reasoning and Planning

Designs a tailored analysis plan, identifying necessary data such as historical price trends, financial ratios, market trend analysis, and relevant recent news.



### Tool Execution for Analysis

Leverages financial APIs and tools such as financial statement analyzers, technical signal analyzers, stock screeners, and other related instruments to gather data for strategic analysis. The system generates a concise summary analysis and clearly presents insights from InnovestX (the securities arm of SCBX).



### Real-Time Data Presentation

Displays updated, user-friendly, and actionable information, including:

- Year-to-date performance graphs
- Statistical tables showing win rates across different periods (5, 10, 20 days)
- Average and median returns
- Risk-to-return ratios
- Simulations of potential future returns



### Deep Summary and Recommendations

Provides expert insights from InnovestX, along with practical recommendations based on the complete dataset and analysis.

For example, when asked about MSFT stock, instead of receiving a generic statement like "Microsoft is a good investment," clients receive a data-rich, actionable analysis that facilitates more effective decision-making, e.g., the Analysis states, 'MSFT has a 58% win rate over a 5-day period and 60% over a 20-day period, with a risk-to-return ratio of 1.21, suggesting that the expected return outweighs the associated risk.'

This clearly demonstrates the significant difference in recommendation quality and the ability to manage complex tasks between traditional Gen AI and advanced Agentic AI. Clients of SCBX, such as those served by InnovestX Securities Co., Ltd., benefit from a more comprehensive, accurate, and up-to-date analysis, empowering them to invest with greater confidence.

Nevertheless, investment involves risk. Investors should thoroughly review all relevant information before making any investment decisions.



# ACT IV: Not Quite Human, But Almost There.

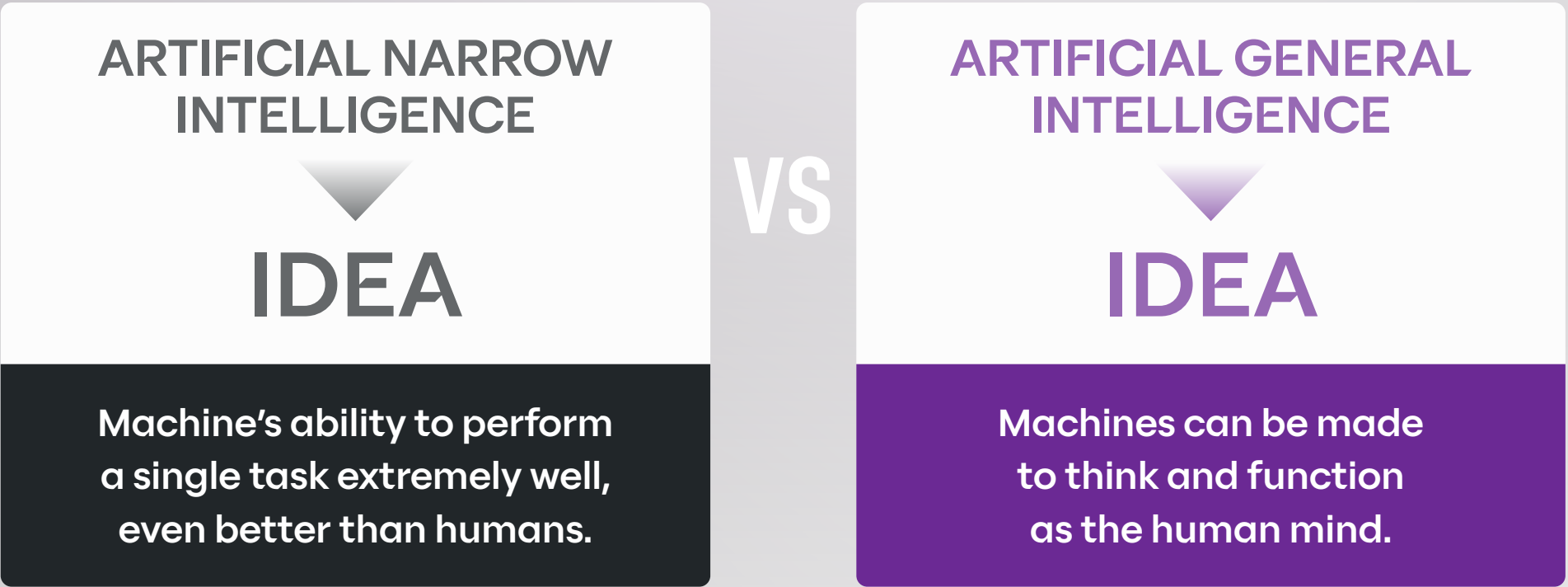




# Artificial General Intelligence (AGI) and the Unresolved Path to Human-Level AI

As artificial intelligence continues its rapid evolution, one concept looms large on the horizon: AGI, or Artificial General Intelligence. Long considered the “holy grail” of AI research, AGI is now capturing renewed attention across the tech and academic communities. Yet despite this growing focus, the path to AGI remains as complex and contested as ever.

## What is AGI?



Artificial General Intelligence (AGI) refers to a form of artificial intelligence capable of thinking, learning, and performing any task at a level comparable to that of a human. While this definition appears straightforward, it conceals a great deal of complexity, particularly because defining what constitutes “human-level cognitive ability” remains a matter of debate among scientists, philosophers, and AI experts.

The key distinction between current AI and AGI lies in the scope and flexibility of their capabilities.

Present-day AI, despite its complexity and power, is primarily categorized as Narrow AI—systems designed to perform specific tasks with high efficiency, such as image recognition, language translation, information retrieval, or financial data analysis.

In contrast, AGI must possess a combination of diverse abilities and be capable of transferring knowledge from one domain to another. Most importantly, it must adapt to new and unfamiliar situations without the need for additional training—much like how humans learn and respond to novel circumstances.

Comparing human intelligence to machine capabilities is inherently complex, as the human brain operates fundamentally differently from computers. Humans possess contextual understanding, creativity, and the capacity to make value-based and emotionally influenced decisions—facets that are profoundly difficult to replicate within computational systems.



# Diverging Voices from AI Industry Leaders

Within the current AI landscape, leaders and experts express significantly differing views regarding the timeline and feasibility of achieving AGI. These contrasting perspectives highlight the inherent uncertainty and the formidable challenge of forecasting a technology as complex and potentially world-changing as AGI.



Has stated that he believes AGI will be achieved sooner than most people expect. His confidence aligns with OpenAI’s mission to develop AGI that is both safe and beneficial.<sup>[1]</sup>

**Sam Altman**  
CEO of OpenAI



Holds a more cautious view, suggesting that the development of AGI may take longer than anticipated—potentially up to ten years.<sup>[2]</sup>

**Mustafa Suleyman**  
CEO of Microsoft AI



Has expressed strong confidence that superhuman AGI could emerge within the next five to ten years. He believes such advancement has the potential not only to enhance but also to revolutionize multiple domains, including healthcare and the fight against climate change.<sup>[3]</sup>

**Demis Hassabis**  
CEO and Co-founder of DeepMind  
(a subsidiary of Google)





# A Framework for Evaluating AGI Capability Levels

The development of Artificial General Intelligence (AGI) is not expected to occur through a sudden leap from current systems to human-equivalent capabilities. Rather, it is a continuous progression through various stages. To better track this advancement and establish clearer developmental milestones, researchers from both academia and industry have collaborated to construct a framework for evaluating AGI capability levels.

This assessment framework categorizes AI progression into multiple levels, ranging from Level 0 (No AI) to Level 5 (Superhuman), with each level defined by distinct characteristics and capabilities.<sup>[4]</sup>

Level  
0

## No AI

Refers to general-purpose computational systems that do not exhibit any AI characteristics. Examples include calculators, compilers, or systems reliant on human input behind the scenes, such as Amazon Mechanical Turk, which represents human-in-the-loop computing.

Level  
3

## Expert

Systems performing at the 90th percentile skill level of an adult. For Narrow AI, examples include grammar and spell-checking tools like Grammarly, and image generation models such as Imagen (Google Brain) or DALL·E 2 (OpenAI). No system has yet achieved this level for Expert AGI.

Level  
1

## Emerging

Systems that are on par with, or slightly better than, unskilled humans. In the case of Narrow AI, these systems rely on simple rule-based mechanisms. For General AI, examples include ChatGPT (OpenAI), Llama 2 (Meta), and Gemini (Google), which are considered examples of Emerging AGI.

Level  
4

## Virtuoso

Systems performing at the 99th percentile skill level of an adult. Notable Narrow AI examples include Deep Blue (Campbell et al., 2002) and AlphaGo (Silver et al., 2016; 2017). To date, no AGI system has attained Virtuoso-level performance.

Level  
2

## Competent

Systems that perform at the level of an average adult with 50th percentile skills. In the realm of Narrow AI, this includes intelligent voice assistants such as Siri (Apple), Alexa (Amazon), and Google Assistant, as well as state-of-the-art large language models (LLMs) used for tasks such as short essay writing or basic coding. As for Competent AGI, no system has yet reached this level.

Level  
5

## Superhuman

Systems that consistently outperform humans in all domains. Examples from Narrow AI include AlphaFold and AlphaZero (Google DeepMind). In the case of Artificial Superintelligence (ASI), no system has yet reached this level.

**Note:** Narrow AI = AI that specializes in specific domains, such as playing chess or recommending products. It performs effectively only within its defined boundaries.  
General AI = Human-level AI capable of learning and solving diverse problems across multiple domains.



# The True Path to AGI



**Yann LeCun**, Vice President and Chief AI Scientist at Meta (Facebook), shared his perspective on the development of true AGI at the AI Action Summit 2025.<sup>[5]</sup>

He emphasized that advancing toward true AGI is not merely a matter of improving existing models for greater efficiency, but may require a fundamental paradigm shift in how AI systems are designed. His key recommendations are as follows:

1

### Abandon Generative Models

Move away from AI systems that generate text, images, or audio based on learned data (such as GPT or Stable Diffusion). Instead of focusing on such generative models, experts suggest shifting toward joint-embedding architectures, which prioritize understanding the relationships between different types of data and delving deeply into the meaning of that data—rather than merely producing content.

2

### Abandon Probabilistic Model

Probabilistic models function by calculating the likelihood of certain events, such as the probability of a word following a given phrase (e.g., predictive text systems). LeCun advocates for the use of energy-based models, which assess the state of a system based on energy formulations. These models may offer a better grasp of complex data structures and more closely emulate how the human brain operates.

3

### Abandon Contrastive Methods

Contrastive learning methods, which teach AI by comparing differences between data points, should be replaced with regularized methods. These approaches focus on model refinement with more stability and comprehensiveness, helping prevent biased or incomplete learning.

4

### Abandon Reinforcement Learning

Reinforcement learning (RL), where AI learns through rewards and penalties, should not be the default approach. LeCun recommends relying on model-predictive control, a technique that uses a predictive model to forecast the outcome of future actions and determines the optimal sequence of actions in real time. RL should only be used as a secondary mechanism—when planning fails to yield the expected results—in order to fine-tune the world model or the critic component for higher accuracy.

5

### Final Advice: If You Are Interested in Human-Level AI, Don't Work on LLMs

Although Large Language Models (LLMs) demonstrate impressive capabilities, they may not ultimately lead to AGI. LLMs face inherent limitations in holistic world understanding, reasoning across domains, and solving complex problems. Therefore, LeCun advises that those who are genuinely focused on building human-level intelligence should explore alternative approaches.

The journey toward AGI remains fraught with challenges and ongoing debate. While AI technology continues to advance rapidly, achieving true artificial intelligence at a human level remains an unresolved and ambitious goal.



# EPILOGUE: The AI Storm – Infinite Impact.

## Case Studies from Inside the Eye of the Typhoon – SCBX Case Studies



### TYPHOON

The SCBX Group, through its innovation arm SCB 10X, is not only observing the AI revolution—it is actively shaping it. From customer experience to financial advisory to national-scale innovation, SCBX has been integrating artificial intelligence into real-world applications that deliver tangible impact.

Among its pioneering initiatives are two featured earlier in this report:

In addition, SCBX, led by SCB 10X, has developed 'Typhoon,' a large language model (LLM) specifically designed to enhance Thai language processing capabilities.

- The Smart RM Training System (see ACT II), a voice-based AI simulation for training relationship managers.
- The Agentic AI Financial Advisory Platform (see ACT III), which empowers intelligent financial decision-making in dynamic market conditions.

But perhaps most emblematic of SCBX's ambition is Typhoon—a Thai-language Large Language Model (LLM) purpose-built to elevate the capabilities of natural language processing in the Thai language. Developed to align with local linguistic nuances and cultural context,

Typhoon enables AI-powered applications to be more accurate, accessible, and impactful for Thai users.

Typhoon has already begun transforming sectors through strategic collaborations with leading national institutions. Key case studies include:

- Siriraj Hospital – Enhancing healthcare documentation and diagnostic workflows with Thai-language AI.
- VISAI – Building advanced legal AI tools, including chatbots and research assistants tailored for the Thai legal system.
- Thailand Development Research Institute (TDRI) – Empowering data-driven labor market policy through large-scale AI analysis.

These implementations underscore SCBX's commitment to responsible, localized, and high-impact AI deployment. Typhoon is more than a model—it's a platform for progress, enabling innovation at the intersection of technology, language, and social value.

As we look ahead, these case studies offer a glimpse into how AI, when paired with a strategic vision and strong ecosystem partnerships, can deliver not just automation—but transformation.





## Public Health Management: Siriraj Hospital and the Enhancement of Medical Information Systems

Siriraj Hospital is a leading medical institution that provides healthcare services, education, and research. Given the massive volume of medical data and documentation, knowledge management within the organization poses a significant challenge.

To address this, Siriraj Hospital has implemented Typhoon to reduce the complexity of managing essential documents such as medical records, and internal documents, all of which require real-time search and retrieval. Typhoon employs Retrieval-Augmented Generation (RAG) to accurately extract relevant information.

The benefits of this case include reducing the burden on medical personnel in reviewing large volumes of documents, accelerating access to critical information, shortening decision-making time, and enhancing the accuracy and efficiency of patient care.

Looking ahead, the hospital aims to expand the use of Typhoon to further support patient care while maintaining data privacy, and to apply Typhoon's capabilities to scanned documents across various departments.



## Legal Sector: VISAI and AI for Legal Education

VISAI is an organization dedicated to developing AI technologies to support legal practice in Thailand.

The organization developed Sommai, a legal chatbot that uses Typhoon to answer questions related to business and commercial law, as well as analyze gaps in legal research. The system supports large document processing and delivers results with internationally comparable accuracy.

Sommai is capable of understanding complex legal content, processing large volumes of legal documents, and delivering precise results. Its key strength lies in its ability to search for information related to Corporate and Commercial Laws (CCL) across 35 legal domains. It also supports legal research, enables comparative legal studies, and facilitates the analysis of differences between Thai and foreign legal systems.

This enhances the effectiveness and accuracy of legal AI in the Thai context and represents a significant advancement in Thailand's legal technology landscape.



## Labor Market Planning: TDRI and Labor Market Trend Analysis

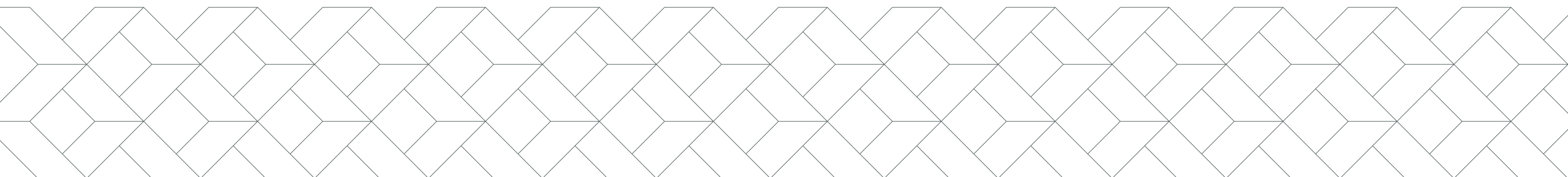
The Thailand Development Research Institute (TDRI) is an independent research organization focused on public policy and national development.

TDRI employs Typhoon to analyze data from online job postings to identify labor market trends in Thailand. The system accurately extracts information on job titles, workplace locations, and in-demand skills.

Typhoon provides faster and more accurate results than previously used systems, enabling timely policy development based on deep labor market insights. This empowers policymakers to better understand labor market dynamics and anticipate the future skill demands of the workforce.

**We invite organizations, industry partners and individuals committed to the growth of the AI Innovation to partner with us:**

SCB 10X Typhoon Team  
contact@opentyphoon.ai  
SCBX R&D Division  
rd@scbx.com



References:

ACT I: Two Philosophies, One Future.  
The Battle Between Open-Source and Closed-Source AI Intensifies

[1] ข้อมูลเปรียบเทียบข้อดี-ข้อเสียของ AI แบบเปิดและปิดในบริบทธุรกิจ : <https://www.multimodal.dev/post/open-source-ai-vs-closed-source-ai>

[2] Closed Models Outperform Open Models, at Staggering Cost : <https://virtualizationreview.com/Articles/2024/04/16/ai-index.aspx>

[3] บทความ Why Open-Source LLMs Are Reshaping the Economics of AI : <https://aithority.com/machine-learning/why-open-source-llms-are-reshaping-the-economics-of-ai/>

[4] อัปเดตอนาคตของ “Open Source AI” จาก SCB 10X : <https://www.scb10x.com/blog/open-source-ai-trends-opportunities>

[5] Zhang, D., Sahai, T., Jarrett, D., Zhao, S., Tamkin, A., Bai, Y., ... & Etchemendy, J. (2024). The AI Index 2024 Annual Report. Stanford Institute for Human-Centered Artificial Intelligence. <https://aiindex.stanford.edu/report/>

[6] กรณีศึกษาจาก World Economic Forum เกี่ยวกับโมเดล DeepSeek : <https://www.weforum.org/stories/2025/02/open-source-ai-innovation-deepseek>

[7] Epoch. (2024, November 28). The state of open foundation models. Epoch. From <https://epoch.ai/blog/open-models-report>

[8] “Should Enterprises Adopt Closed-Source or Open-Source AI Models?” CB Insights, Mar. 2025, [www.cbinsights.com/research/enterprise-adoption-closed-source-open-source-ai-models](https://www.cbinsights.com/research/enterprise-adoption-closed-source-open-source-ai-models)

[9] Artificial Analysis. (2025, February). Artificial Analysis Intelligence Benchmarking Methodology. From <https://artificialanalysis.ai/methodology/intelligence-benchmarking>

[10] Typhoon. (2025, February 7). Introducing Typhoon2-R1–70B: Enhanced Reasoning with DeepSeek R1. Typhoon Blog. From <https://opentyphoon.ai/blog/en/introducing-typhoon2-r1-70b-enhanced-reasoning-with-deepseek-r1-abcdefa9b7ad>

[11] SCB 10X. “SCB 10X Hugging Face Profile.” Hugging Face, <https://huggingface.co/scb10x>

ACT II: Tiny Titans - Small, but Mighty.  
More Versatile, Smaller, and Smarter: 3 Trends of the Next AI Evolution

[1] Google Reveals Gemini 2, AI Agents, and a Prototype Personal Assistant <https://www.wired.com/story/google-gemini-2-ai-assistant-release/>

[2] “Introducing GPT-4o and More Tools to ChatGPT Free Users.” OpenAI, 13 May 2024, <https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free/>

[3] “Gemini 2.5: Our Most Intelligent AI Model.” The Keyword, 25 Mar. 2025, <https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/#gemini-2-5-thinking>

[4] “Alibaba Cloud Releases Qwen2.5-Omni-7B: An End-to-end Multimodal AI Model.” Alizila, 27 Mar. 2025, [www.alizila.com/alibaba-cloud-releases-qwen2-5-omni-7b-an-end-to-end-multimodal-ai-model/](https://www.alizila.com/alibaba-cloud-releases-qwen2-5-omni-7b-an-end-to-end-multimodal-ai-model/)

[5] Veluri, B., Peloquin, B. N., Yu, B., Gong, H., & Gollakota, S. (2024). Beyond Turn-Based Interfaces: Synchronous LLMs as Full-Duplex Dialogue Agents. arXiv preprint arXiv:2409.15594

[6] Phi-4 ทำผลงานได้ดีกว่า LLM อย่าง Gemini Pro1.5: <https://www.techradar.com/computing/artificial-intelligence/microsoft-announced-phi-4-a-new-ai-thats-better-at-math-and-language-processing>

[7] Abdin, M., et al. (2024). Phi-4 Technical Report. Microsoft Research. <https://www.microsoft.com/en-us/research/wp-content/uploads/2024/12/P4TechReport.pdf>

[8] “How IBM CEO Arvind Krishna Is Thinking About AI and Quantum Computing.” TIME, 11 Mar. 2025, <https://time.com/7263217/ibm-ceo-arvind-krishna-interview/>

[9] “How Scaling Laws Drive Smarter, More Powerful AI.” NVIDIA Blog, 12 Feb. 2025, <https://blogs.nvidia.com/blog/ai-scaling-laws/>

[10] อธิบายแนวทางการให้โมเดลลองสร้างและประเมินหลายคำตอบระหว่างการใช้งานจริง เพื่อเลือกคำตอบที่ดีที่สุดใ้ท้ายที่สุด และ บทสัมภาษณ์ Ilya Sutskever ที่กล่าวถึงการเริ่มเห็นขีดจำกัดของการเพิ่มขนาดโมเดลและความจำเป็นต้องหาแนวทางใหม่ ๆ ในการพัฒนา AI : <https://www.reuters.com/technology/artificial-intelligence/openai-rivals-seek-new-path-smarter-ai-current-methods-hit-limitations-2024-11-11>

[11] Mixture of Experts แทนที่จะเป็นแค่ LLM แค่นั้น : <https://www.datacamp.com/blog/mixture-of-experts-moe>

ACT III: AI at Your Fingertips.  
Agentic AI: Rise of the Agents

[1] Stanford Institute for Human-Centered Artificial Intelligence. (2024). HAI at Davos: Key Insights on AI From the World Economic Forum. Stanford HAI. <https://hai.stanford.edu/news/hai-at-davos-key-insights-on-ai-from-the-world-economic-forum>

[2] Davenport, T. H., & Bean, R. (2025). Five trends in AI and data science for 2025. MIT Sloan Management Review. <https://sloanreview.mit.edu/article/five-trends-in-ai-and-data-science-for-2025/>

ACT IV: Not Quite Human, But Almost There.  
Artificial General Intelligence (AGI) and the Unresolved Path to Human-Level AI

[1] Health, A. (2024). Sam Altman lowers the bar for AGI. The Verge. <https://www.theverge.com/2024/12/4/24313130/sam-altman-openai-agi-lower-the-bar>

[2] Davis, W. (2024). Microsoft’s AI boss and Sam Altman disagree on what it takes to get to AGI. The Verge. <https://www.theverge.com/2024/12/9/24316969/mustafa-suleyman-sam-altman-microsoft-openai-agi>

[3] Taylor, J. (2024). AI has the potential to ‘cure all diseases’, says DeepMind chief. The Times. <https://www.thetimes.com/business-money/technology/article/ai-has-the-potential-to-cure-all-diseases-says-deepmind-chief-9hbdp5kpm>

[4] Feng, T., Jin, C., Liu, J., Zhu, K., Tu, H., Cheng, Z., Lin, G., & You, J. (2024). How Far Are We From AGI: Are LLMs All We Need? [Preprint]. arXiv. <https://arxiv.org/abs/2311.02462>

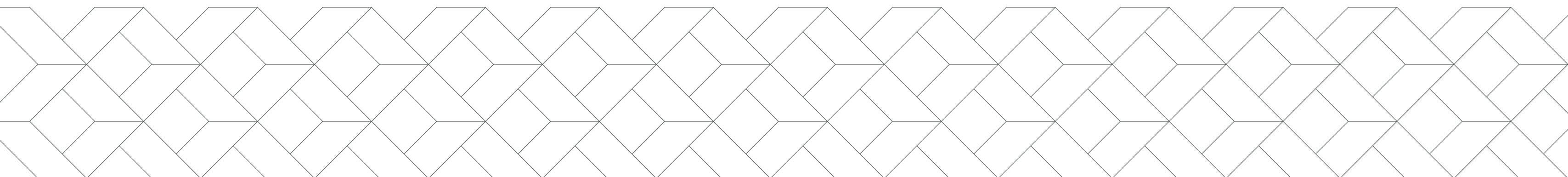
[5] DSAI by Dr. Osbert Tay. (2025). The Shape of AI to Come! Yann LeCun at AI Action Summit 2025 [Video]. YouTube. <https://www.youtube.com/watch?v=xnFmnUOPp-8>



# About SCBX

SCBX is the mothership of the financial technology business group, operating across three key business pillars: Banking Business, Consumer and Digital Finance Business, and Platform and Technology Business. In addition, SCBX also focuses on Climate Technology, aspiring to become 'The Most Admired Regional Financial Technology Group'.

The company conducts its business with flexibility and prudence in governance and risk management and has possesses the potential to compete equally in global competitions.



# Authors



**Tutanon Sinthuprasith, Ph.D.**  
R&D and Innovation Lab LEAD, SCBX



**Kaweewut Temphuwapat**  
Chief Innovation Officer, SCBX

This publication was made possible through the collaboration of





SCB<sup>x</sup>

[www.scbx.com](http://www.scbx.com)