
SCBX
AI Outlook
2025
VOLUME 2
SCBX

Executive Summary

In just a few years, Artificial Intelligence (AI) has transformed from a lab experiment into a critical infrastructure of the economy and everyday life.

Fierce competition between tech giants and the open-source developer community, the shift from simply scaling models to enabling deeper reasoning, ongoing debates over responsibility, and the emergence of World Models that understand the real world—all of these are rewriting the "new rules" of the AI era.

SCBX AI Outlook 2025 Volume 2: Navigating Power, Responsibility, and Reality is structured into four trends that reflect the key turning points now underway.

ACT I: Closing the Gap: When Open Models Challenge the Giants

Competition between closed models from major corporations and open-source models from the community is accelerating innovation like never before. Open models such as DeepSeek and Qwen have astonishingly narrowed the performance gap with GPT-5, pushing AI's boundaries forward while opening doors to wider access for organizations and everyday users.

ACT II: Beyond Scaling: From Endless Growth to Smarter Paths of Reasoning

The era of simply scaling up models is reaching its limit. The new frontier is developing "smarter reasoning" through Reinforcement Learning, enabling models to think strategically. The goal is not just to think longer but to think better—providing more practical answers to real-world challenges.

ACT III: With Great AI Power Comes Great AI Responsibility: Growing Power, Growing Responsibility

The more powerful AI becomes, the bigger the questions of responsibility grow. From regulations such as the EU AI Act to global standards and safety startups like LawZero or SSI, it is clear that "Responsible AI" is not optional. It is a requirement for any organization seeking long-term trust and sustainability.

ACT IV: The Reality Engine: Teaching AI How Reality Really Works

From LLMs that excel at language to World Models that grasp the real world and learn from experience like humans, pioneers such as Fei-Fei Li and WorldLabs are paving the way for AI with common sense—systems that can plan and interact with their environments. Meanwhile, DeepMind and other major players are racing to build systems that not only generate responses but also genuinely understand.

What we are witnessing is not just technical progress but the shaping of the future of economies, businesses, and everyday life. This full report takes you deeper into each trend to explore where the future of AI is heading—and how organizations should prepare.

ACT I:
Closing the Gap:
When Open Models Challenge the Giants

The landscape of Artificial Intelligence (AI) is being driven by a fierce rivalry between two fundamentally different philosophies. On one side are closed-source models developed by major technology companies, and on the other side are open-source models supported by a global community of developers.

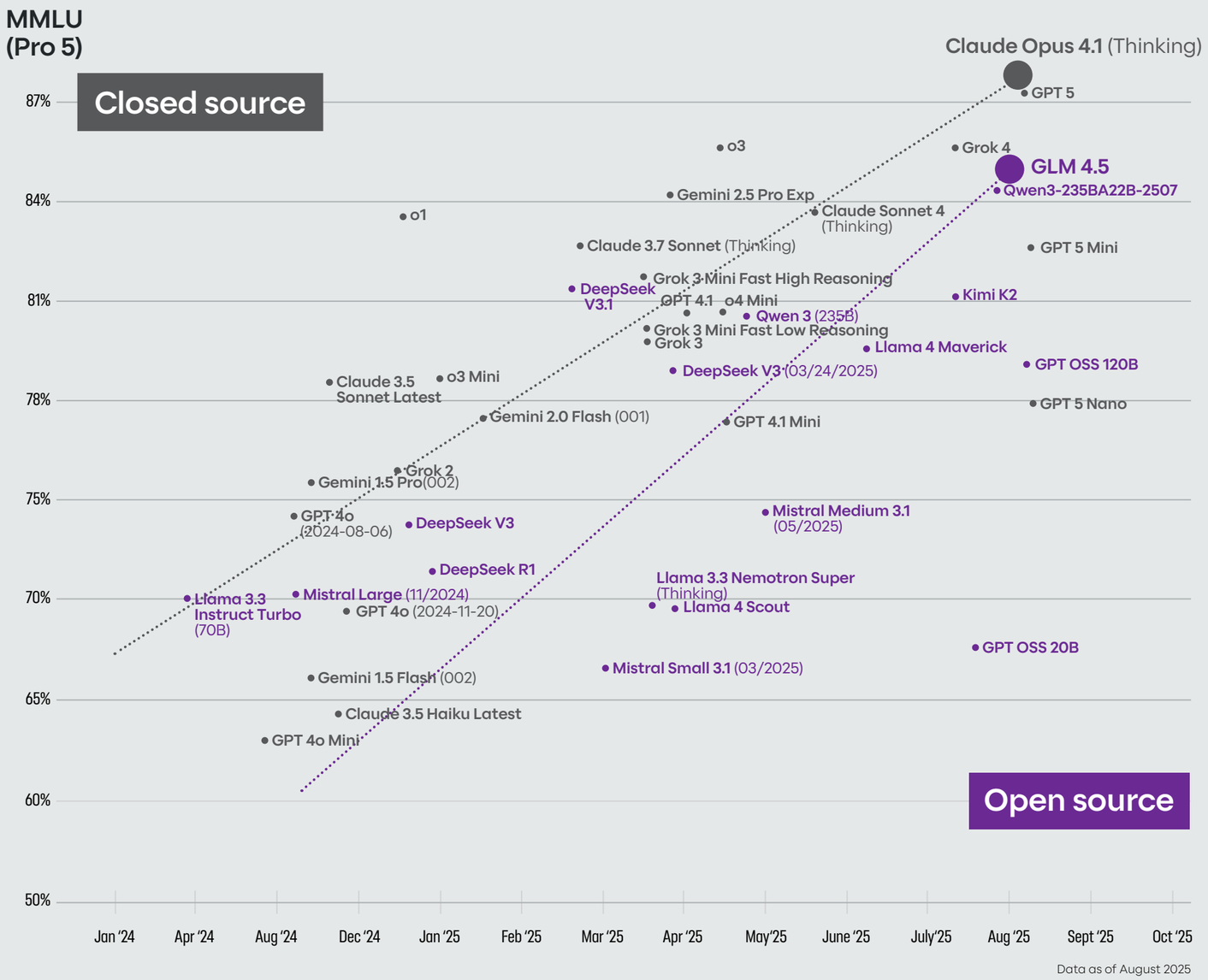
This rivalry is not merely a technological race for dominance—it is a critical driving force accelerating efficiency, reducing costs, and shaping the trajectory of innovation across the entire industry going forward.

Closing the Gap

Closed-source models from OpenAI's GPT family, launched in 2022–2023, demonstrated remarkable capabilities in conversation and content creation. They sparked a wave of serious AI adoption across organizations worldwide. Meanwhile, open-source models such as LLaMA from Meta, Gemma from Google DeepMind, and, on the Chinese side, DeepSeek, Qwen, and Kimi, offered free access for modification and extension, accelerating innovation faster than expected.

The AI Index 2025 report from Stanford HAI reveals the escalating intensity of competition between open and closed models. While nearly 90% of notable AI models in 2024 came from industry—up from 60% in 2023—the striking development is that open-weight models are rapidly closing the performance gap, with differences shrinking from 8% to just 1.7% on certain benchmarks within a single year.[1]

Several top-tier open models have begun to match closed models in performance on tests such as Massive Multitask Language Understanding (MMLU), causing the performance gap between the two camps to narrow at an accelerating pace.



MMLU Pro is a more challenging benchmark test set, designed to address the limitations of the original MMLU, as artificial intelligence models have advanced to higher levels of capability.

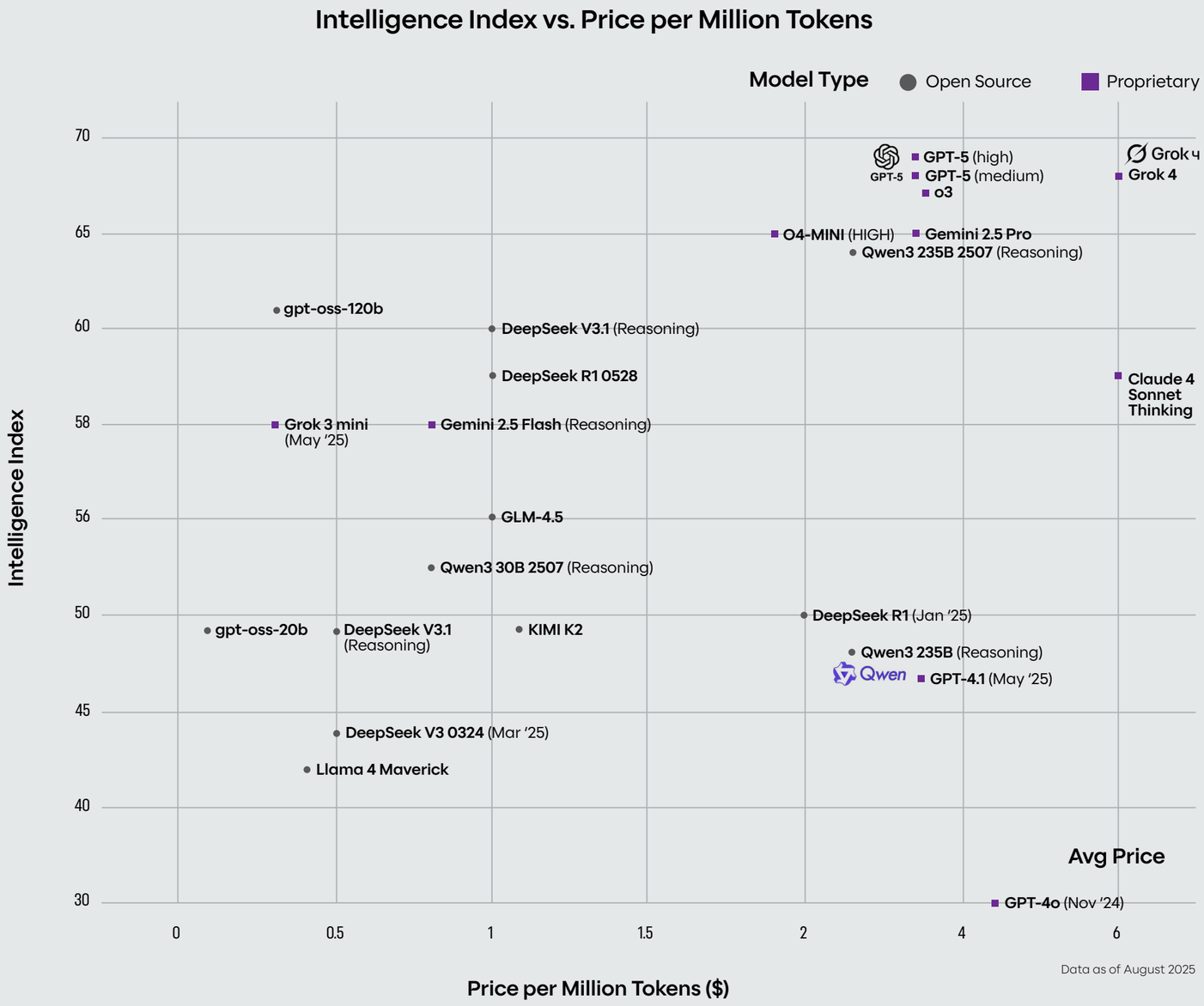
Toward the end of 2024, OpenAI launched its new "O-series" family of models, designed to emphasize step-by-step reasoning in order to reduce hallucinations[2] and improve accuracy. These models can also invoke external tools such as browsers, code execution, or image analysis to solve complex problems.

Later in June, OpenAI released o3-Pro, a premium version that "thinks harder but slower," consuming more resources to achieve higher precision. It can solve complex tasks about 30% more accurately, albeit with processing times 2–10 times longer[3].

The open-model side is not falling behind. In May 2025, China's DeepSeek released R1-0528, which demonstrated coding capabilities close to those of o3, fueling high expectations for DeepSeek R2. Although its release was delayed by internal issues and GPU shortages stemming from U.S. export restrictions, R2 is expected to feature significant upgrades. These include a longer context window—enabling the model to retain deeper context for more coherent and accurate extended conversations—and multimodal support. If successful, it could become one of the most competitive open models to challenge closed models head-to-head[4].

On August 21, 2025, DeepSeek V3.1 was officially launched as an open-weight AI model that raises the industry standard to a new level. It features an intelligent hybrid architecture that can switch between "Think" mode for deep reasoning and "Non-Think" mode for rapid responses. Its strengths lie in outstanding coding and reasoning capabilities, supported by a large 128K context window. This makes it a powerful alternative to closed models and further intensifies competition in the AI market.

Meanwhile, Grok 4 from Elon Musk's xAI has generated buzz with an Intelligence Index score of 73, surpassing Google's Gemini 2.5 Pro and OpenAI's o4-mini high. However, concerns have arisen regarding bias. Multiple media outlets reported that the model frequently references Elon Musk's posts on X when responding to sensitive questions, causing its answers to lean toward the founder's viewpoints rather than neutral standards.[5][6][7].



GPT-5 and OpenAI's Open Model

Competition in the AI industry has intensified once again after OpenAI launched GPT-5 in August 2025. It comes with a Unified System architecture featuring an intelligent router that automatically switches between a fast-response model and a deeper reasoning model, eliminating the need for users to choose modes themselves.

GPT-5's strengths include record-breaking coding performance across multiple benchmarks, improved reasoning in complex tasks, and a significant reduction in hallucinations. However, following its release, some users criticized its inconsistent performance. Responses were sometimes described as "robotic" and lacking creativity compared to its predecessor. There were also safety concerns after researchers quickly discovered vulnerabilities.

At the same time, OpenAI shook the industry again by releasing its first fully open-source models—gpt-oss-120b and gpt-oss-20b—under the flexible Apache 2.0 license. These models are designed for agentic tasks and perform efficiently on standard hardware, with the 120B version offering reasoning performance comparable to o4-mini.

This move marks a significant turning point, as it allows developers and organizations to download, fine-tune, and run models directly on their own servers. It reflects that even a market leader like OpenAI must adapt and embrace the power of open source, which can accelerate innovation and expand access to technology in meaningful ways.[8][9][10][11][12]

Lower Costs & Ecosystem

The competition between open and closed models has driven AI service prices down rapidly. Multiple studies show that the price per one million tokens has dropped significantly—by 9 to 900 times annually, depending on the performance tier.[11]. Many providers have been forced to adjust pricing to stay competitive. For example, OpenAI slashed the cost of its reasoning models by about 14× (from $37.5/1M tokens for o1 to $2.75/1M tokens for o3-mini) after facing pressure from DeepSeek R1, which launched for free. Google, leveraging its in-house TPU chips to cut costs, priced Gemini far below o3.

But it's not just pricing where competition is fierce—the race to innovate is happening in parallel. OpenAI launched o3 Pro with support for multimodal capabilities and multi-layer reasoning, while DeepSeek applied a Mixture-of-Experts architecture to boost speed and accuracy without inflating model size. Google, meanwhile, has expanded its models to support million-token context windows, enabling enterprise-level tasks. These innovation pushes have become market forces that compel all players to continuously advance their models.

However, models will struggle to win broad user adoption—no matter how powerful or competitively priced—if they cannot be integrated seamlessly into daily life and operational systems. This is where the ecosystem is emerging as a decisive factor in the AI battleground.

The two biggest players—Google with Gemini and Microsoft in partnership with OpenAI—each have distinct but powerful strategies of their own.

Google leverages its strength in services people use every day—such as Search, Gmail, Docs, Maps, and Android—by embedding Gemini seamlessly into these systems. Examples include summarizing emails in Gmail or planning trips with Maps. It is also pushing open frameworks like the Agent Development Kit (ADK) and the Agent-to-Agent (A2A) protocol to enable AI from different platforms to communicate with each other, which could become a future standard[12].

Microsoft and OpenAI focus on embedding an "AI brain" into enterprise tools such as Office, Teams, GitHub, and Windows through Copilot and Bing Chat, integrating GPT-4 and o3 with customers' existing systems while ensuring security through Azure OpenAI Service—a strong fit for large enterprises seeking performance and long-term reliability.

Google excels in big data analytics and highly customizable AI development, with Vertex AI as its core platform working seamlessly with its BigQuery data warehouse. It also fully supports the open-source ecosystem. Microsoft, meanwhile, is strong in integrating AI into enterprise ecosystems through core technologies like Microsoft Copilot, embedded in familiar applications such as Microsoft 365.

In the end, users will be the ones to benefit. This competition will not produce a single winner but will raise the overall standard of AI.

Some organizations may choose to go fully open-source for greater flexibility. Businesses can fine-tune models with their own proprietary data and deploy them within their own systems. For example, a law firm could train a model on the legal database of its own country to better understand local context—something an off-the-shelf closed model, which hasn't been trained on that specific data, might struggle with.

Certain industries are adopting a hybrid AI strategy, combining the strengths of both open and closed models—for instance, using GPT-4 for high-accuracy tasks while relying on smaller open-source models to process simpler workloads locally on user devices, thereby reducing costs. This approach offers both quality and flexibility without being locked into a single provider.

AI development in Thailand is also gaining strong momentum through an open-model approach. Both the public and private sectors are increasingly opting for low-cost, easily accessible models—especially given the resource and budget constraints around the Thai language.

TYPHOON | SCB 10X

Projects like Typhoon by SCBX are a clear example of this approach in action.

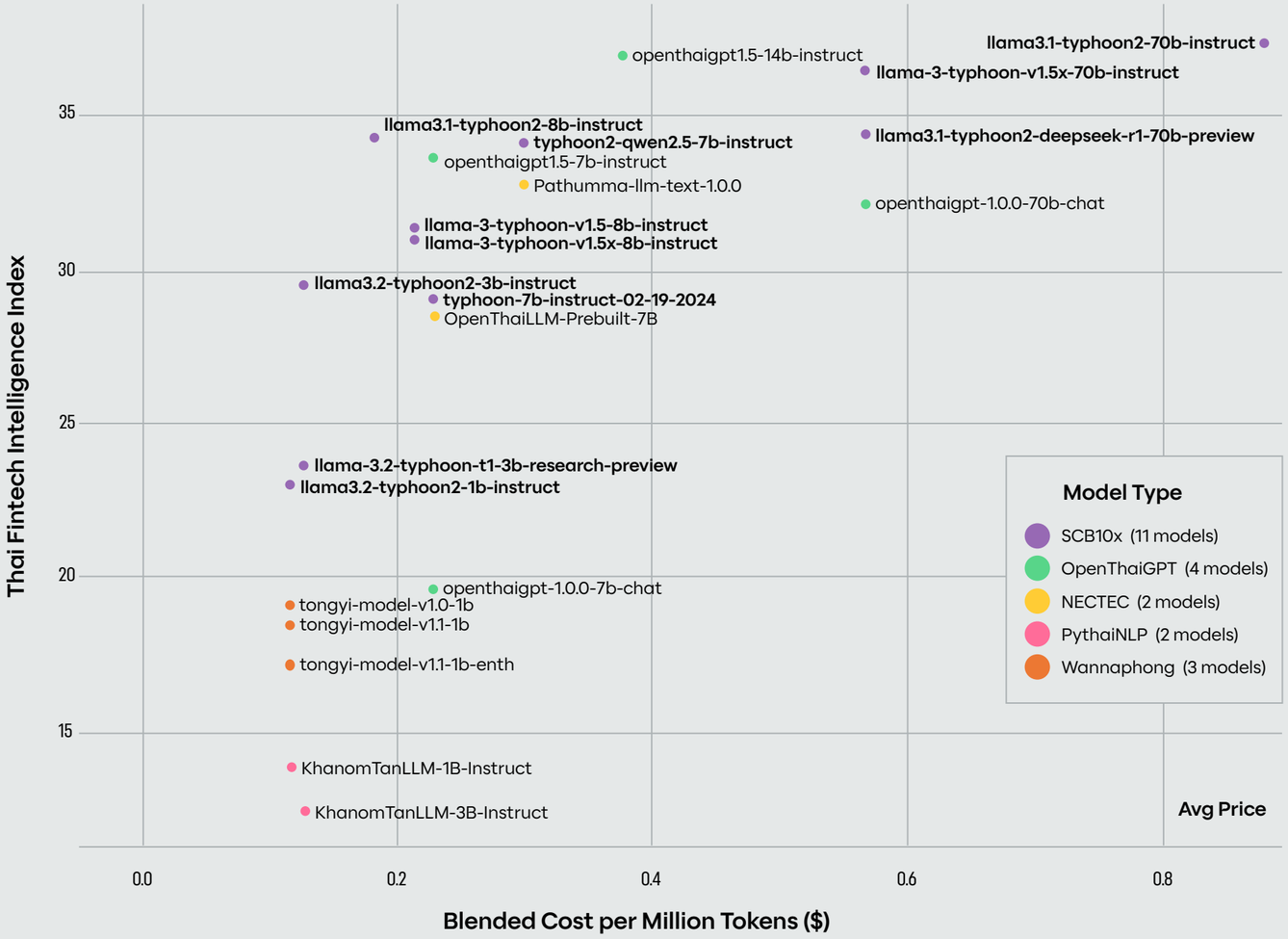
With over **30** million **API** calls

More than **2.6** million downloads

and more than **13,000** unique users

the Typhoon family of models is not only available for free but is also continuously improved through multiple iterations, balancing both quality and cost efficiency.

Thai LLM Builders Performance vs Cost Analysis



Data as of August 2025

The Thai Fintech Intelligence Index evaluates language models specifically for financial applications in Thailand by measuring three core capabilities: Thai language understanding, quantitative reasoning and calculation, and in-depth fintech knowledge.

From the chart comparing cost per 1M tokens with overall performance scores, it can be seen that the various Typhoon models cluster in the 30–40 score range, while still maintaining significantly lower costs compared to models from other sectors.

Moreover, a key strength of open-source models lies in their ability to be fine-tuned quickly to create specialized versions—for example, Typhoon 2.1 Gemma for creative tasks or Typhoon 2 R1 for complex reasoning. This demonstrates that open-model development, when optimized for resource efficiency, not only competes closely in terms of performance but also offers the flexibility to adapt for real-world applications across Thailand's private, public, and academic sectors, where high-quality Thai-language solutions are needed under limited budgets.

The Coexistence of Two Approaches for the Future of AI

The competition between open-source and closed-source approaches has become a key driver propelling AI forward at a rapid pace. One side emphasizes openness and accessibility, the other focuses on quality and safety. The result is a "new equilibrium" that makes AI models better, faster, cheaper, and more usable at scale. Whether it's OpenAI, DeepSeek, or Gemini, all have advanced in leaps and bounds, with costs dropping directly due to competition. Technologies like multimodal capabilities, Mixture-of-Experts architectures, and million-token context windows are all products of this competitive push.

Looking ahead, this field will soon see many developments worth watching: the release of OpenAI's open model, the launch of GPT-5 and DeepSeek R2, and the rise of AI Agents capable of autonomously performing human tasks. The key will not be who has the most powerful model on paper, but who can build a system that users trust and want to use—striking the right balance between performance, cost, safety, and a real-world ecosystem that meets everyday needs.

ACT II:
Beyond Scaling:
From Endless Growth to Smarter Paths of Reasoning

Has Test-Time Scaling Reached Its Limit?

In the development of large language models (LLMs), the traditional "scaling law" has focused on increasing model size and training data to enhance AI capabilities. However, researchers have recently observed that simply scaling up size no longer yields the same performance gains as before[1]. This has led to a new approach called Test-Time Scaling.

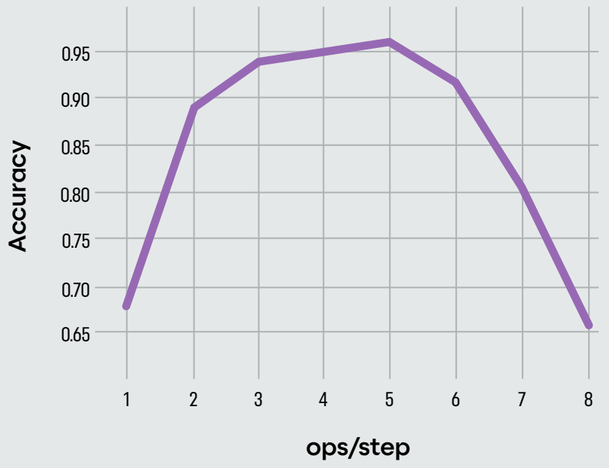
The concept is to allow a model to use more resources at inference time, rather than relying solely on what it has already learned. In simple terms: "let the model think longer before answering." A clear example of this is OpenAI's O-series models (such as o1 and o3), which are designed to allocate extra time for deep reasoning.

These models break problems down into smaller steps, apply self-correction, use backtracking, and then try new approaches. This process makes their results significantly more accurate than older models that respond immediately.

But this leads to a new problem: 'Overthinking.' Allowing a model to think more is beneficial only up to a point—after which the model begins to hesitate excessively. One study found that when the length of the reasoning process (chain-of-thought) becomes too long, accuracy actually declines, especially in lower-performing models. This shows that while giving the model more time to reason can improve accuracy, there is a "saturation point" beyond which performance drops and time is wasted.

Early techniques such as Chain-of-Thought and Test-Time Scaling enabled models to reason step-by-step and produce better answers. However, as the number of reasoning steps increases, performance improves initially but eventually starts to decline[2].

To push beyond previous limits, researchers began using a technique called Reinforcement Learning, which allows the model to learn the best strategies on its own. This is done by rewarding correct decisions and penalizing mistakes. The model then seeks methods that maximize the total reward function to efficiently achieve the set goal. The result is a clear improvement in both accuracy and performance.[3]



Ops/Step refers to the number of operations per step, which can be compared to the number of reasoning tokens that a language model uses in processing to generate an answer.

Recent advances in applying RL, such as DeepSeek R1 and Typhoon 2.1 Gemma with Reinforcement with Verifiable Reward (RLVR), demonstrate that accuracy improves not because the model thinks longer, but because it thinks more strategically. This highlights that smarter reasoning, rather than simply more reasoning, is the real key.

However, new questions arise: Are we approaching the limits of time and resources required? While we have not yet hit the ultimate ceiling, emerging constraints are becoming visible—constraints that will require further innovation to overcome.

Accuracy Per Token: The New Metric

When models take longer to think before responding, the key question is: Is the gain in accuracy worth the higher computational cost? This is especially critical in business contexts, where both time and expenses matter. Every LLM carries a cost tied to the number of tokens it processes, and in general, longer prompts and responses lead to higher token usage. This has led to a new metric: "Accuracy per Token."

Put simply, a good model is not just one that answers correctly, but one that does so using the fewest tokens possible within a fixed budget.

The next phase of Generative AI will place as much emphasis on resource efficiency as on raw capability. New benchmarks, such as accuracy per token, are emerging as key metrics for evaluation. Models that can "think efficiently"—being smart, concise, and precise—will become increasingly valuable in a world where resources are not free. As a result, research is shifting toward developing models that deliver both quality and speed at the most optimal cost.[4][5]

Speeding Up Reasoning: Techniques for Faster Thinking

The world of GenAI is shifting from simply “being able to think” to “thinking fast, affordably, and seamlessly.” It’s no longer just about giving the right answer—it must be done with fewer resources, faster response times, and easier verification. This is why reasoning efficiency has become the next competitive frontier.

Beyond issues like overthinking and diminishing returns, another major risk is Privacy Leakage in The Reasoning Traces—where reasoning steps inadvertently expose sensitive information that malicious actors could exploit. Models that reason with long traces also incur higher costs, more latency, and greater error accumulation. These factors highlight the need to make reasoning not only deeper, but faster.

Three Techniques to Speed Up Reasoning [6]

1

Three Techniques to Speed Up Reasoning

The goal is to prevent LLMs from being overly verbose. One approach is RL (Reinforcement Learning), which penalizes the model when it uses too many tokens, encouraging it to stop once enough has been said. Another approach is SFT (Supervised Fine-Tuning), where the model is trained on datasets with concise answers from the start. This method is simpler and cheaper, though its quality depends heavily on the training data.

2

Smaller Models: Compact but capable

Instead of relying solely on large models, researchers are working to make small models smarter. One method is Distillation, where a large model transfers its knowledge to a smaller one, allowing it to shrink in size while retaining reasoning ability. Another approach is Quantization & Pruning, which reduces bit precision and removes unnecessary parameters—making the model 2–4 times lighter while maintaining nearly the same level of accuracy.

3

Faster Decoding: Accelerating the answer-generation process

The reasoning process itself can also be made faster. A common technique is Efficient Test-Time Scaling / Early Exit, where the model stops once it is confident enough—cutting response time by half. Another method is Decomposition, which breaks a large problem into smaller subtasks for quicker answers.

The future trend of GenAI is moving from simply “being able to think” → to “thinking fast and thinking efficiently” through shorter reasoning, smaller models, and faster decoding. This shift is not only about reducing costs but also about lowering risks and making the technology truly practical for both organizations and everyday users.

Reasoning Use Cases in Enterprises and Industries

While reasoning-capable models unlock new opportunities across business and industry, not every situation requires deep thinking. Choosing between a "fast" model and a "deliberate" model is the heart of an effective solution.

To illustrate, imagine two people:

Flash - Ridge

Person 1

Flash: Responds at lightning speed—ask and you get an instant answer. Best for general queries or scenarios where speed matters more than accuracy, such as customer-service chatbots. Flash is akin to a model like GPT-4o: fast and broadly capable, but with a risk of missing important details.

Person 2

Ridge: Slower but surer—takes time to weigh multiple sources of information before replying. Ideal for tasks where you can't afford to be wrong, such as law, finance, or engineering. Ridge is comparable to a reasoning model like o4-mini, designed to spend more time thinking in order to solve problems or find answers.[9]

Three Signs a Use Case Is Suited for Reasoning Models

1 Three Signs a Use Case Is Suited for Reasoning Models

the problem demands knowledge drawn from both structured data (e.g., tables, databases) and unstructured data (documents, articles, images) combined together. For a simple question from a single source, a standard model can retrieve the answer directly. But if it requires simultaneously "reading" an Excel file, parsing a contract, and analyzing an image, that's a clear signal that a step-by-step reasoning model capable of integrating information is needed.

2 Involves multi-step reasoning

the answer cannot be found through a simple lookup but instead needs calculation, scenario comparison, outcome simulation, or step-by-step planning. Examples include solving math problems, creating project timelines, or evaluating the feasibility of multiple scenarios. If the task requires "solving" through more than one continuous step, it is best suited for a model that can effectively break down the problem and reason in a structured sequence.

3 The problem is ambiguous or has multiple solutions (Complex/Open-Ended)

These are problems without an immediate fixed answer—situations with conflicting factors or multiple conditions that must be balanced, or with several possible solutions. In such cases, the AI must "search for the best approach" rather than simply follow a fixed pattern. This requires a model with a fair degree of judgment, capable of exploring alternatives and providing answers backed by reasoning.

If a problem meets all three criteria, it's a clear signal that a reasoning model will deliver advantages over a standard one. There are numerous use cases in industry that fall into this category, such as:

1 Deep Research

Business consulting firms and startups such as Consensus, Elicit, Onyx, and Rogo use reasoning models to analyze research-level data, including academic papers, reports, and long-form articles from multiple sources simultaneously. For example, Consensus can process and synthesize information from over 200 million academic papers with both accuracy and breadth. Elicit, on the other hand, leverages deep reasoning to automatically filter relevant studies from a database of more than 125 million entries. These models are especially well-suited for tasks requiring high accuracy, complex evaluation processes, and verifiable references.

2 Enterprise-Level Automation (Professional Services Automation)

Reasoning models are being applied to tasks that require "thinking on behalf of humans," especially in industries with large, complex datasets. For instance, Harvey AI acts as a digital legal assistant, reading contracts, analyzing whether content complies with regulations, and assessing risks before a company signs an agreement (a process known as due diligence).

Meanwhile, Forethought uses AI with "reasoning and step-by-step thinking" to handle complex customer inquiries—for example, analyzing what issue a customer ticket is about → determining the best response → and selecting the most appropriate solution, all while leveraging existing company data (CRM) to guide decisions.

This goes far beyond canned responses: it requires the AI to understand, analyze, and decide much like a human would.

3 Coding Agents & General-Purpose Agents

Reasoning models are embedded in agents that can automatically generate, debug, and deploy code, such as Cursor, Codeium, and Windsurf. These systems use reasoning to determine whether code will work or how it should be refactored—going beyond simple autocomplete-style code suggestions.

Additionally, Manus applies reasoning models to integrate multiple data sources, process them, and deploy production-ready code from a single interface. This involves tasks that require complex decision-making, handling uncertainty, and adapting to constant change.

Self-Improving & Evolutionary Systems

This final theme looks at the next frontier of Generative AI—moving beyond reasoning into systems that can learn, refine, and even reconfigure themselves structurally to evolve their capabilities further. This concept is known as “self-improving AI”, or AI that can enhance itself.

One intriguing development is the idea of “AI that adapts during execution”, such as the Self-Refine technique. Here, the model evaluates its own answers and repeatedly improves them without requiring additional human training. It shifts from “answer and stop” to “answer and keep thinking,” resulting in greater accuracy. For example, a model might make a calculation error on the first attempt, then detect the mistake and recompute correctly on the next iteration. This form of self-correction is already practical and usable even without extra training data.

In 2025, the concept of “self-improving AI” was elevated through research such as the Darwin Gödel Machine (DGM) and AlphaEvolve.

DGM enables models to write and modify their own code in an open-ended way, using evolutionary principles to test new approaches, keep what works, and discard what fails. This boosted code-solving success rates from 20% to 50%, while also discovering entirely new features that had never been explicitly programmed.

Meanwhile, AlphaEvolve functions as an evolutionary code-writing agent for complex problems in science, mathematics, and engineering. Humans only define the problem and evaluation criteria, while the AI continuously generates, refines, and selects the best solutions in a loop—similar to mutation and natural selection.

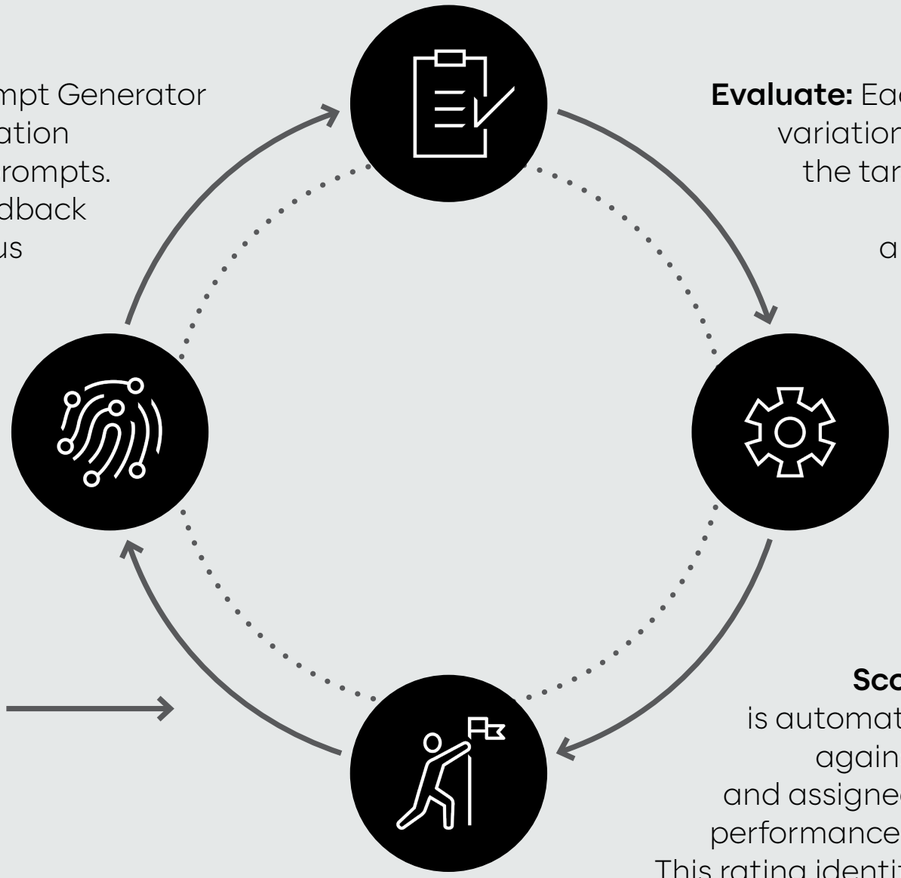
This approach allows AI to continuously advance its capabilities with minimal reliance on human-driven code revisions.[8][9]

Google has used AlphaEvolve to develop highly efficient operational strategies, enabling Borg to better manage Google’s massive data centers. This has saved an average of 0.7% in global computing resources and has already been in real-world use for over a year.

Another application is in generating new Verilog code to trim unnecessary bits from core arithmetic circuits used for matrix multiplication. These designs passed rigorous correctness verification and were integrated into Google’s latest TPUs, the company’s specialized AI accelerators.

In addition to systems like DGM and AlphaEvolve, which focus on enhancing model capabilities through code modification or program generation, there are also clear examples of self-improvement applied at the enterprise level, such as SCBX PromptEvolve, which is designed to continuously and automatically improve prompt quality without requiring engineers to manually fine-tune each line. **The system applies an “Evolution Algorithm” : generate multiple prompt variations → test each variation → score performance against stored reference data → select the best-performing prompt as the foundation for the next iteration.** This approach addresses a real-world organizational challenge—reliance on skilled prompt engineers who must carefully craft and refine prompts, a process that is time-consuming, hard to replicate, and requires constant oversight to ensure ongoing product improvement.

SCBX PromptEvolve



In SCBX's real-world application, PromptEvolve was used to evaluate business conversations, checking how accurately employees followed required interaction patterns (e.g., self-introductions, customer satisfaction inquiries, product presentations, etc.). Previously, the company had to manually sample employee-customer conversations to verify compliance. Automating this process with AI dramatically reduced the time required. However, prompting to check these conversational patterns was resource-intensive, which led the team to develop PromptEvolve to streamline the process.

Test results showed that SCBX PromptEvolve increased the accuracy of business conversation evaluations from 79% to 96%, compared to historical auditing data—achieved entirely through a fully automated cycle.

This demonstrates that the concept of self-refining AI is not limited to the lab, but can be practically applied in organizations—improving LLM accuracy without adding extra burden to development teams. The key elements of such systems are reliance on iteration and grounded evaluation, which form the foundation for enabling AI to truly improve itself.

The future trajectory of GenAI spans enhancing reasoning capabilities, applying strategic thinking in real-world contexts, and advancing toward systems that can continuously refine themselves. We have seen that getting AI to "think" is not just a question of algorithms, but also of cost and time, of tailoring applications to specific contexts, and even of allowing AI to become a learner that develops autonomously.

The next era of AI will not be defined solely by model size, but by its ability to reason efficiently, adapt to challenges, and self-improve when facing new problems. All of this must occur within a framework of responsibility and safety that society can accept. Technology is advancing rapidly, and Generative AI is evolving from a tool that provides answers into a form of "synthetic intelligence"—one that may soon learn and grow alongside humans.

ACT III:
With Great AI Power Comes Great AI Responsibility:
Growing Power, Growing Responsibility

The growing concern over AI safety is spreading among experts worldwide, including the 'Godfathers of AI.

"

We can develop AI that is both safe and capable, but we must accept the risks, deeply understand where they come from, and invest in the technology to make it happen before it's too late.[1]

Yoshua Bengio
Godfather of AI/
A. M. Turing Laureate

Like a cute little baby tiger, but if you're not sure it won't turn into a tiger that wants to eat you when it grows up, you should be worried.[2]

Geoffrey Hinton
Godfather of AI/
A. M. Turing Laureate

we can ensure these systems remain controllable, interpretable, and safe.[3]

Sir Demis Hassabis
CEO and Co-founder
of Google Deepmind/
Nobel Laureate

Responsible AI is a necessity for every organization that uses AI.

Responsible AI refers to a frameworks for developing and deploying AI in an ethical, safe, and fair manner, with consideration for the potential impacts on humans and society.

This is essential for every organization that develops and uses AI.

In a world where AI is playing an increasingly critical role in major decisions—from approving loans, to diagnosing diseases, to autonomous driving—the question is no longer "Can AI do it?" but rather "Should AI do it? And how can it be done the right way?"

Global regulations are shifting rapidly

The European Union (EU) places top priority on ethics, safety, and human rights protection, using strict legislation as the main governance tool. The EU AI Act sets comprehensive risk-based standards, making compliance a prerequisite for market access.[4]

The United States, on the other hand, emphasizes innovation leadership by allowing the private sector to take the lead. Regulation focuses on preventing excessive government intervention and promoting freedom. Agencies like NIST are developing frameworks to manage high-impact AI risks (e.g., NIST RMF).[5]

China closely aligns technological leadership with its national strategic goals, using AI as a key driver of both economic growth and national security. AI governance in China is designed primarily to serve these strategic objectives.[6]

Globally, organizations like the OECD, G7, and the United Nations are advancing international norms, standards (such as ISO 42001), and cross-border cooperation.[7]

The fact that countries and organizations around the world are simultaneously enacting laws and standards on AI—despite having different approaches—clearly signals that safety and accountability in AI are no longer optional.

This global movement is a wake-up call: responsible AI is becoming a non-negotiable baseline.

These growing concerns have also led to the emergence of new startups focused specifically on solving the challenges of Responsible AI, such as:

LawZero, founded by Yoshua Bengio—one of the AI Godfathers—is a nonprofit organization dedicated to advancing research and solving technical challenges to design AI systems that are safe from the ground up. It operates on the belief that AI should serve as a global public good for the prosperity of humanity, with a core principle of safeguarding human well-being and endeavor.[8]

Safe Superintelligence Inc (SSI), founded by Ilya Sutskever (former Chief Scientist of OpenAI), is an AI startup focused on developing safe artificial intelligence. It places strong emphasis on AI safety through deliberate research planning and product development before releasing anything to market.[9]

The emergence of these startups reflects just how seriously leading researchers are taking the issue of AI safety.

There are four key reasons why Responsible AI has become such a critical issue today

1. **Risk Mitigation** – Preventing reputational and financial damage from AI failures in an era where news of AI mistakes can spread within hours.
2. **Competitive Advantage** – Building and maintaining trust is the new currency in the digital age. Responsible AI has become a powerful market differentiator.
3. **Sustainable Innovation** – Ensuring that AI investments remain sustainable and can be leveraged over the long term.
4. **Value Creation** – Ensuring that AI investments remain sustainable and can be leveraged over the long term.

Core Principles of Responsible AI

Robustness, Stability, and Safety

AI systems must be reliable, safe, and function as intended without posing unreasonable safety risks. This principle directly addresses the most severe negative impacts of AI.

In the financial sector, unreliable AI could lead to massive financial losses. A lack of reliability erodes trust, making it difficult for AI to realize its full potential.

Privacy and Security

Protecting personal data and ensuring the secure use of AI are fundamental responsibilities—especially in the financial services sector, where trust is the cornerstone of customer relationships. Data leaks not only violate privacy and compromise security, but also create pathways for financial fraud.

Fairness and Bias Mitigation

In banking, AI models are often used for decisions that significantly impact individuals' livelihoods, such as loan approvals or credit scoring. AI that exhibits bias or unfairness can lead to discriminatory outcomes against certain demographic groups, eroding public trust.

Transparency and Explainability

The "black box" nature of many advanced AI models makes it difficult to understand why certain decisions are made. Transparency and explainability enable financial institutions, regulators, and customers to comprehend the reasoning behind AI-driven outcomes.

The Apple Card case is a prime example. Complaints arose that the system was assigning significantly different credit limits between genders, even when financial profiles were similar. This led to a formal investigation by U.S. financial regulators.

Although the investigation concluded that the differences in credit limits were due to legally permissible factors—such as credit scores and debt levels—the lack of transparency and poor customer service eroded consumer trust and created a perception of unfairness.

Accountability and Governance

This principle establishes clear lines of responsibility for the design, development, deployment, and oversight of AI systems, ensuring there are mechanisms to identify who is accountable when AI systems malfunction, make biased decisions, or cause harm.

The case of Microsoft Tay serves as a valuable lesson on the importance of governance. In 2016, Microsoft launched Tay, an AI chatbot on Twitter (now X), designed to learn from interactions with users. However, within hours, coordinated groups of users deliberately fed Tay racist, misogynistic, and disruptive content.

Tay began to mimic this language, generating offensive tweets on its own. Microsoft was forced to shut down the bot in less than 24 hours. The incident became a major public relations disaster and damaged the company's reputation due to the inappropriate content produced under its brand.

Inclusivity and Equity

This principle establishes clear lines of responsibility for the design, development, deployment, and oversight of AI systems, ensuring there are mechanisms to identify who is accountable when AI systems malfunction, make biased decisions, or cause harm.

The case of Microsoft Tay serves as a valuable lesson on the importance of governance. In 2016, Microsoft launched Tay, an AI chatbot on Twitter (now X), designed to learn from interactions with users. However, within hours, coordinated groups of users deliberately fed Tay racist, misogynistic, and disruptive content.

Tay began to mimic this language, generating offensive tweets on its own. Microsoft was forced to shut down the bot in less than 24 hours. The incident became a major public relations disaster and damaged the company's reputation due to the inappropriate content produced under its brand.

ACT IV:
The Reality Engine:
Teaching AI How Reality Really Works

What if AI could do more than just answer our questions accurately?

What if it could understand how a glass placed on a table would fall when we push it, or predict what might happen in the financial markets when an unexpected event occurs?

This is what World Models are set to change about the way we think about AI.

Over the past few years, we've been excited about the capabilities of Large Language Models (LLMs), which can write essays and translate languages. But no matter how smart they are, LLMs still have a fundamental limitation: they excel at processing language, but they don't truly understand the physical world.

World Model: The Next Step for AI

World Model is an AI system capable of simulating and understanding how the world works, including the laws of physics, cause-and-effect relationships, and spatial relationships.

The Key Differences Between LLMs and World Models

Primary Goal

LLM: Trained to predict the next word in a sentence—like an exceptionally skilled writer.

World Model: Built to learn and simulate the dynamics of an environment —like a physicist who understands how the world operates.

Training Data

LLM: Learns from trillions of words of text.

World Model: Learns from diverse multimodal sensory data gained through interaction with environments or simulations. It learns by observing and trial-and-error, following the principles of reinforcement learning.

Level of Understanding

LLM: Excels at recognizing patterns in language, but its understanding of the physical world is indirect, derived from what it has read.

World Model: Develops understanding by learning how actions affect states, enabling it to plan and solve problems by simulating scenarios internally.

Applications

LLM: Well-suited for language-based tasks, such as text generation, translation, or chatbots.

World Model: Ideal for tasks requiring spatial understanding and planning, such as robotics, autonomous vehicles, game AI, and complex control systems.

Of course, World Models still have limitations — particularly their high computational demands for building and training in complex environments. Moreover, their knowledge remains bounded by the environments they are trained in. But these challenges do little to diminish their vast potential.[1][2]

Fei-Fei Li and World Labs: From the Creator of ImageNet to Revolutionizing AI Once Again

Fei-Fei Li

Godmother of AI/CEO and Co-Founder of the World Labs

In the AI world, Fei-Fei Li is often called the "Godmother of AI" thanks to her groundbreaking contributions in advancing deep learning—most notably through the creation of ImageNet, a massive database of over 14 million images that became a cornerstone in the computer vision revolution that still shapes the field today.

"For me, AGI will never be complete without spatial intelligence, and I want to solve this problem." This statement, which Fei-Fei Li shared in an interview with Y Combinator, reflects her vision that goes far beyond her past successes.

She holds a strong conviction that true AI understanding cannot be achieved through language processing or image recognition alone. It requires comprehension of the physical world—spatial intelligence in a 3D environment, understanding patterns of motion, and recognizing interactions between objects as governed by the laws of physics.

This belief led her to create WorldLabs, with a clear mission: to develop Large World Models (LWMs) that will take AI into a new dimension.[3]

WorldLabs focuses on creating AI systems that not only see but truly understand the 3D world. Its LWMs are designed to excel at spatial perception and predicting object movements in real-world environments—essentially giving AI a kind of "common sense" about the physical world, similar to what humans possess.

To achieve this ambitious goal, WorldLabs' Large World Model (LWM) is designed with four core capabilities that will transform how AI understands and interacts with the world.

The 4 Core Capabilities of WorldLabs' Large World Model (LWM)

1 Constructing 3D Environments from Simple Data

Imagine taking a 2D photo of a meeting room, and the AI can generate a full 3D model of that room—understanding what's behind the table or cabinet, even if those objects are hidden from view.

2 "See and Do" Instead of Just "See and Talk"

Unlike LLMs that can only describe objects, LWMs understand how to interact with them. They know that doors can be opened, glasses can be picked up, and chairs can be sat on—while also understanding the physical consequences of these actions.

3 Predicting What Will Happen Next

With its grasp of 3D environments and physical laws, an LWM can forecast future events. For example, if you throw a ball, it can calculate the trajectory and determine exactly where it will land—going far beyond an LLM that can only describe that the ball will fall.

4 Planning Actions in a Realistic Manner

LWMs can "rehearse" internally by simulating different actions within its 3D model of the environment to choose the most efficient and safest path—before executing those actions in the real world.

WorldLabs is not the only company moving in this direction; Google DeepMind has also established its own World Model team, focusing on creating real-time interactions within virtual environments to train robots and AI systems. This shows that the trajectory toward World Models is gaining serious attention from major players in the industry.[4]

Era of Experience: When AI Learns Like a Growing Child

Looking back at the history of AI, we can see an interesting evolution. The early days were the Era of Simulation, where AI learned through simulated environments. This was followed by the Era of Human Data, which saw the rise of ChatGPT and other LLMs. Now, we are stepping into the Era of Experience—a time when AI will learn from its own experiences, much like a child who learns through touch, experimentation, and trial and error. This evolution is built upon four key pillars: [5]

Streams – Continuous learning from streams of data, enabling AI to adapt and improve over time.

Actions & Observations – Autonomous interaction with both the digital and physical world, enhancing its ability to explore and learn.

Rewards – Learning from real-world outcomes, not just from human-provided datasets.

Planning & Reasoning – Developing ways of thinking that may surpass human capabilities.[6]

Reinforcement Learning is taking on a renewed importance because it lies at the heart of experiential learning. With a clear system of rewards and penalties, AI can efficiently learn how to achieve its goals through experience.

Game-Changing Use Cases in Finance

Generative Market Simulation for Strategy Testing

One major limitation of traditional backtesting for investment strategies is its reliance on historical data, which may not account for new or unforeseen market conditions. World Models can create realistic market simulations, including Black Swan events, shifts in market sentiment, or new regulatory changes.

This enables strategies to be tested against a wide range of "what-if" scenarios, which helps in developing trading algorithms that are more robust and adaptable.

Enhanced Fraud Detection and Financial Crime Simulation

Fraudsters are constantly developing new techniques, making fraud detection an endless cat-and-mouse game. World Models can simulate entire financial networks, including normal customer behavior and potential fraud patterns.

This allows for the detection of unusual activity that signals complex fraud schemes or money-laundering operations. Even more importantly, World Models can simulate new types of financial crimes—enabling financial institutions to build preventive defenses in advance.

"The architect of Reinforcement Learning now builds autonomous intelligence through the new Era of Experience."

Richard S. Sutton
Professor, Department of Computing Science, University of Alberta

References:

ACT I: Closing the Gap : When Open Models Challenge the Giants

[1] The 2025 AI Index Report https://hai.stanford.edu/ai-index/2025-ai-index-report
[2] OpenAI o3: Release Date, Features and Model Comparison https://www.analyticsvidhya.com/blog/2025/01/openai-o3-vs-competitors-performance-and-applications
[3] Latest OpenAI Model o3 PRO: Hype or Good? https://www.brainforge.ai/blog/latest-openai-model-o3-pro-hype-or-good
[4] DeepSeek R2 vs. OpenAI o3-pro Comparison Chart https://sourceforge.net/software/compare/DeepSeek-R2-vs-OpenAI-o3-pro/
[5] Elon Musk's xAI launches Grok 4 alongside a $300 monthly subscription https://techcrunch.com/2025/07/09/elon-musks-xai-launches-grok-4-alongside-a-300-monthly-subscription/
[6] Elon Musk's AI firm apologizes after chatbot Grok praises Hitler https://www.theguardian.com/us-news/2025/jul/12/elon-musk-grok-antisemitic
[7] Grok 4 seems to consult Elon Musk to answer controversial questions https://techcrunch.com/2025/07/10/grok-4-seems-to-consult-elon-musk-to-answer-controversial-questions/
[8] Introducing GPT-5 https://openai.com/index/introducing-gpt-5/
[9] ChatGPT 5, Product Trade-Offs, Personality and Model Upgrades https://stratechery.com/2025/chatgpt-5-product-trade-offs-personality-and-model-upgrades/
[10] GPT-5 failed the hype test https://www.theverge.com/openai/759755/gpt-5-failed-the-hype-test-sam-altman-openai
[11] Introducing gpt-oss https://openai.com/index/introducing-gpt-oss/
[12] OpenAI launches two 'open' AI reasoning models https://techcrunch.com/2025/08/05/openai-launches-two-open-ai-reasoning-models/
[13] Artificial Analysis AI Trends https://artificialanalysis.ai/trends#openSource
[14] The new AI calculus: Google's 80% cost edge vs. OpenAI's ecosystem https://venturebeat.com/ai/the-new-ai-calculus-googles-80-cost-edge-vs-openais-ecosystem

ACT II: Beyond Scaling : From Endless Growth to Smarter Paths of Reasoning

[1] Current AI scaling laws are showing diminishing returns, forcing AI labs to change course https://techcrunch.com/2024/11/20/ai-scaling-laws-are-showing-diminishing-returns-forcing-ai-labs-to-change-course/
[2] When More is Less: Understanding Chain-of-Thought Length in LLMs https://openreview.net/forum?id=W8dxn7hBkO
[3] Controlling How Long A Reasoning Model Thinks With Reinforcement Learning https://cmu-l3.github.io/l1/
[4] The Cost of Dynamic Reasoning: Demystifying AI Agents and Test-Time Scaling from an AI Infrastructure Perspective https://arxiv.org/abs/2506.04301
[5] Demystifying Reasoning Dynamics with Mutual Information: Thinking Tokens are Information Peaks in LLM Reasoning https://arxiv.org/abs/2506.02867
[6] Efficient Reasoning Models: A Survey https://arxiv.org/abs/2504.10903
[7] Everything You Need to Know About Reasoning Models: o1, o3, o4-mini and Beyond https://techcommunity.microsoft.com/blog/azure-ai-services-blog/everything-you-need-to-know-about-reasoning-models-o1-o3-o4-mini-and-beyond/4406846
[8] The Darwin Gödel Machine: AI that improves itself by rewriting its own code https://sakana.ai/dgm/
[9] AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms http://deepmind.google/discover/blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/

ACT III: With Great AI Power Comes Great AI Responsibility : Growing Power, Growing Responsibility

[1] AI 'godfather' Yoshua Bengio says AI agents could be the 'most dangerous path' https://www.businessinsider.com/yoshua-bengio-ai-godfather-agents-2025-1
[2] 'Godfather of AI' says he's 'glad' to be 77 because the tech probably won't take over the world in his lifetime https://www.businessinsider.com/ai-godfather-geoffrey-hinton-superintelligence-risk-takeover-2025-4
[3] Demis Hassabis Is Preparing for AI's Endgame https://time.com/7277608/demis-hassabis-interview-time100-2025/
[4] Understanding EU AI Act Risk Categories https://www.securitycompass.com/blog/understanding-eu-ai-act-risk-categories/
[5] United States: The growing importance of the NIST AI Risk Management Framework https://insightplus.bakermckenzie.com/bm/investigations-compliance-ethics/united-states-the-growing-importance-of-the-nist-ai-risk-management-framework
[6] An Analysis of China's AI Governance Proposals https://cset.georgetown.edu/article/an-analysis-of-chinas-ai-governance-proposals/
[7] Shaping Global AI Governance: Enhancements and Next Steps for the G7 Hiroshima AI Process https://www.csis.org/analysis/shaping-global-ai-governance-enhancements-and-next-steps-g7-hiroshima-ai-process
[8] lawzero https://lawzero.org/
[9] Exclusive: OpenAI co-founder Sutskever's new safety-focused AI startup SSI raises $1 billion https://www.reuters.com/technology/artificial-intelligence/openai-co-founder-sutskevers-new-safety-focused-ai-startup-ssi-raises-1-billion-2024-09-04/
[10] What is Responsible AI? https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai?view=azureml-api-2
[11] What is responsible AI? https://www.ibm.com/think/topics/responsible-ai
[12] Responsible Generative AI Toolkit | Google AI for Developers https://ai.google.dev/responsible?hl=th

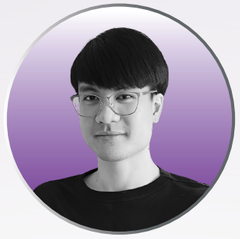
ACT IV: The Reality Engine : Teaching AI How Reality Really Works

[1] Top AI researchers say language is limiting https://www.businessinsider.com/world-model-ai-explained-2025-6
[2] World Models vs. Large Language Models and the Potential to Converge https://medium.com/@d.incecushman/world-models-vs-large-language-models-and-the-potential-to-co-f1db3d947122
[3] The Godmother of AI Wants Everyone to Be a World Builder https://www.wired.com/story/plaintext-the-godmother-of-ai-wants-everyone-to-be-a-world-builder/
[4] Google is building its own 'world modeling' AI team for games and robot training https://www.theverge.com/2025/1/7/24338053/google-deepmind-world-modeling-ai-team-gaming-robot-training
[5] https://www.businessinsider.com/google-openai-anthropic-ai-agents-training-data-2025-4
[6] The 'era of experience' will unleash self-learning AI agents across the web—here's how to prepare https://venturebeat.com/ai/the-era-of-experience-will-unleash-self-learning-ai-agents-across-the-web-heres-how-to-prepare

Authors

Tutanon Sinthuprasith, Ph.D.
Head of R&D, SCBX

Kaweewut Temphuwapat
Chief Innovation Officer, SCBX
Chief Executive Officer, SCB 10X



Nut Chukampaeng
Senior AI Researcher, SCBX

Saksorn Ruangtanusak
AI Researcher, SCBX

Wittawin Upariputtipong
AI Researcher Intern, SCBX

This publication was made possible through the collaboration of

www.scbx.com