EY安永

Shape the future with confidence
聚信心 塑未来

AI bank white paper

Reconstructing the financial paradigm through intelligent agents

Chinese path, global perspective and implementation roadmap

The better the question.
The better the answer.
The better the world works.

Contents

*Ernst & Young (China) Advisory Limited

Abstract

Artificial intelligence is evolving from an efficiency-enhancing tool to a core productive force driving a paradigm shift in the financial industry. This white paper proposes that the essence of an AI bank lies in using intelligent agents as foundational units and transforming business processes from human-driven to intelligent agent-collaboration-driven. EY Greater China has developed a five-stage maturity assessment model that outlines a clear roadmap for Chinese financial institutions to progress from instrumental applications to ecological-level development. China is demonstrating an accelerated growth in this area, with its prudent and orderly policy environment and rich scenario-based practices providing a unique reference point for the global development of intelligent finance.

Executive summary

The financial industry is at the threshold of a century-old transformation. Artificial intelligence, particularly intelligent agent technology, is no longer a distant future concept but a present-day engine reshaping banks' growth logic, risk frameworks and organizational models.

The core insights of this white paper:

Redefinition: An AI bank is an intelligence-driven growth engine, an autonomously operating risk system and a human-machine collaborative organizational paradigm. It goes beyond digitalization to become an intelligent value unit embedded across industries.

Chinese path: Guided by principles of **safety, prudence and orderliness**, China's banking industry is advancing large-scale AI adoption at remarkable speed. From platform-based empowerment in state-owned banks to journey intelligence in joint-stock banks and deep specialization in regional banks, a distinctive Chinese path of hierarchical innovation and diversified exploration is taking shape.

Assessment and navigation: EY's **five-stage maturity model** – pilot exploration, scenario application, process reconstruction, business integration and super-intelligent prototype – provides a systematic framework for institutions to assess their current position and chart future development. Crossing maturity thresholds requires simultaneous advancement in customer management, risk management and trusted governance.

Action guide: Successful transformation relies on aligning technology strategy with business strategy. This paper provides tailored advancement paths, organizational reform plans and ecosystem collaboration strategies for banks at different stages, centered on building a sustainable innovation system with **measurable value, embedded governance and ecosystem capabilities.**

We believe digital technologies led by AI are reshaping both the structure and essence of financial services. This white paper aims to ignite dialogue and foster collaboration on how to harness this transformation and move toward a more intelligent, inclusive and trusted financial future.

1 Redefining AI banks – Strategic upgrade from tools to ecosystems

1.1 Paradigm revolution: from digital empowerment to AI-native

For the past decade, the banking industry has focused on digitalization – migrating offline processes to online platforms. Yet this shift has not changed the fundamental nature of business being human-driven. Today, technological waves led by generative AI and intelligent agents are propelling us into the **AI-native** era.

The defining feature of this era is that **business operations are designed and driven by data and intelligence from inception.** AI is no longer a tool to optimize legacy models – it is a **production environment** that enables entirely new models.

Large-scale AI adoption in banking has become strategically urgent, driven by four key pressures. First, the digital economy is evolving toward an AI-native form, and banks that remain at the stage of digital tool application risk losing the opportunity to embed themselves into the real economy's value chain. Second, traditional human-driven operation models face rising costs and efficiency bottlenecks, while data silos and redundant processes make it difficult for service capabilities to meet market demands. Third, customer expectations have shifted toward immediacy, personalization and scenario integration, rendering standardized service systems inadequate for the new generation of users. Finally, global financial peers and technology players are accelerating the development of AI ecosystems, transforming industry competition into a race for intelligent capabilities. Any delay in this transformation will weaken banks' market position and diminish their influence, making transformation imperative.

Chinese perspective: China's vast market, diverse scenarios and vibrant digital ecosystem provide a unique global testing ground for AI-native finance. From the leapfrog development in the mobile payment era to today's extensive exploration of AI-driven financial applications, Chinese financial institutions are demonstrating the potential to transform from model learners to **paradigm definers.**

1.2 Core definition: agent-driven value reconstruction

We define the essence of an AI bank as follows:

EY sees an AI bank as a future-oriented paradigm of digital banking built on three core capabilities: an intelligence-driven growth engine, an autonomous risk management system and a human-machine collaborative organizational structure.

In this paradigm, AI banks treat advanced machine learning and large-scale generative models (LLM) as the central drivers of decision-making, and intelligent agents (AI Agents) as functional units that operate across an agent framework covering the front, middle and back offices. By embedding intelligence into product design, risk control, precision marketing and operational execution, AI banks create a full-process business closed loop of real-time perception, automated decision-making and execution.

Meanwhile, AI banks operate under a governance framework grounded in interpretability, auditability and regulatory compliance. They build controllable and compliant AI governance mechanisms and model risk management systems that embed technological innovation and risk management into business processes in parallel to form a secure, controllable and business-led operation model.

1.3 Three-dimensional disruption: Fundamental reconstruction of bank Operations

Dimensions	Traditional banking model	AI bank paradigm	Essential change
Growth logic	Expanding personnel and networks, product-driven, large-scale marketing	Real-time demand triggering, automatic product matching, self-growing customer relationships	From passive response to proactive prediction
Risk management	Reliance on rules and expert experience, post-event response-oriented	Real-time model perception, dynamic strategy optimization, proactive risk prevention and self-healing	From static shield to immune system
Organizational operation	Hierarchical management, departmental silos, process promotion through meetings	Agent-driven processes, human-machine collaborative decision-making, empowerment of front-end agile teams	From pyramid to outpost network

1.4 Challenges and value returns of AI applications

In the practical implementation of artificial intelligence technology, financial institutions face multiple challenges at the technical, data and organizational levels, while also realizing significant gains in terms of operational and customer value. Based on a joint research report by Ernst & Young LLP and the Institute of International Finance (IIF),[1] the core challenges and key benefits of AI applications can be summarized as follows.

Data quality is the most fundamental bottleneck for financial institutions advancing AI solutions, with the vast majority identifying it as a key challenge. Data availability and the adaptability of training data are also cited by most institutions as the second major issue. Organizational and technical constraints are similarly prominent: the lack of support from key internal stakeholders and weak foundational technical infrastructure remain important obstacles to AI implementation. Among the specific pain points related to data quality, almost all financial institutions view "noisy, outdated, inaccurate, and poorly adaptable unstructured data" as the primary barrier to AI deployment, and the lack of labelled data has become a major constraint on technical implementation.

Despite numerous challenges, AI technology continues to deliver highly valuable business returns. The core benefits concentrated on improved operational efficiency, optimized cost structures and in-depth upgrades in customer experience and service reach. From the practical results of AI use cases, financial institutions' investment in developing and deploying AI application scenarios are based on clear expectations of these core returns, which also reinforces the key driving role of AI technology in upgrading financial business models.

2 Development trends – Integration and resonance of global perspectives and Chinese practices

2.1 Global trends: Focus on governance, talent and scalability challenges

As generative AI and intelligent agent technologies mature rapidly, leading global banks have moved into a phase of deep strategic advancement. The industry's focus is shifting from **whether to apply AI** to **how to build sustainable, governable and scalable core AI capabilities**. Current global practices reveal three common trends.

I. Governance first: From applying AI to managing AI

When AI becomes deeply embedded in core business and risk decisions – including credit approval, pricing and fraud prevention – effective enterprise-level governance becomes a prerequisite for scale. International banks increasingly elevate AI governance to the group level.

- **More than 70% of large global banks have integrated AI governance into their group-wide risk management frameworks.**
- **Among banks that have deployed generative AI at scale, more than 80% have established cross-functional AI governance committees or "control towers".**
- **Governance focus has expanded beyond model performance to encompass interpretability, fairness, data traceability and decision accountability.**

Practice shows that after implementing a group-level AI governance framework, one global commercial bank tripled the number of AI applications without a proportional increase in model-related compliance incidents, and internal audit interventions frequency declined by 20-30%.

Conclusion: The key to AI scalability lies less in technological maturity than in governance maturity. Without "regulatable AI," it is difficult to achieve "widely usable AI."

II. Talent reconstruction: From technical teams to business communities

The primary bottleneck in AI implementation has shifted from technology to interdisciplinary talent with both business insight and technical capability:

- In high-return banks, personnel with business backgrounds usually typically account for 40-50% of AI teams.
- Banks with mature AI talent systems see average reductions of 30-50% in the cycle from project approval to launch.
- Conversely, more than half of AI projects stall in the pilot phase, mainly due to disconnection between business and technology.

Case evidence shows that by cultivating AI leaders who understand business, one bank advanced the integration of business strategy and model design, increasing the business conversion rate of relevant AI use cases by about 20-30%.

Conclusion: AI transformation is fundamentally the reconstruction of organization and capability. Success depends less on algorithmic advances than on who applies it, who is responsible for it and who realizes the value.

III. Scalability leap: From pilot success to systematic operation

Numerous pilots but insufficient scalability has become a common industry challenge:

- Approximately 60-70% of banks still struggle to replicate AI applications across businesses after completing more than 10 AI pilots.
- Scalability bottlenecks mostly arise from the absence of a unified platform, operating mechanisms and ROI evaluation systems.
- Successful banks generally establish three foundations – a unified data and model platform, a clear use case value evaluation method and a cross-functional AI Center of Excellence (CoE).

A typical practice: One bank managed more than 20 use cases through a centralized AI operations platform, increasing the model reuse rate to over 50%, reducing the marginal cost per use case by 30-40% and achieving measurable returns on overall AI investment within 18-24 months.

Conclusion: Realizing AI value depends on operating it as a core capability rather than delivering it as an isolated project.

Summary: Implications for China's Banking Industry

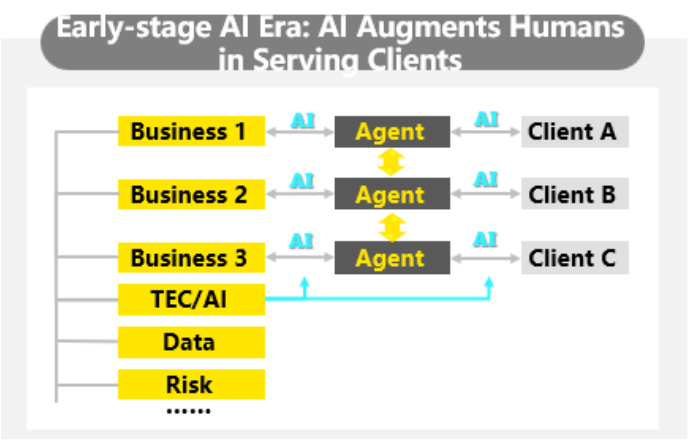
The global construction of AI banks has entered a systematic engineering phase emphasizing governance, talent and scalability, providing important lessons for China's banking industry:

- Governance is a prerequisite for scalability.
- Talent structure sets the upper limit of AI value.
- Without operational and ROI mechanisms, AI struggles to cross the pilot gap.

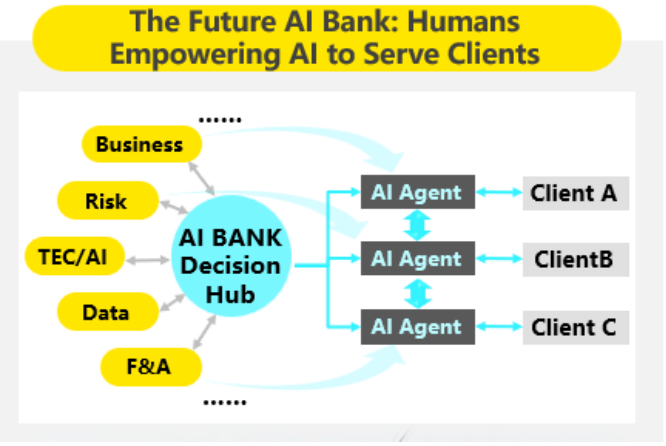
These trends align closely with the regulatory requirements, operating pressures and transformation needs facing China's banking industry and merit full consideration in strategic design and implementation paths.

2.2 Future development: From AI supporting humans in customer service to humans enabling AI to serve customers

The AI services of the banking industry are entering a disruptive transformation – shifting from AI acting only as a tool that assists humans in interacting with customers, to a new stage where humans empower AI and AI becomes the core interface of customer service. From the initial application of AI supporting employees to the in-depth transformation where humans assist AI to serve customers, the intelligent services of banks represent a leap from "tool empowerment" to "subject innovation." Currently, banks remain in a stage where AI enables employees to deliver service; in the future financial landscape, humans will step back to empower AI, allowing AI to directly serve as the core interface of customer engagement. (See Figures 2.2.1 and 2.2.2)



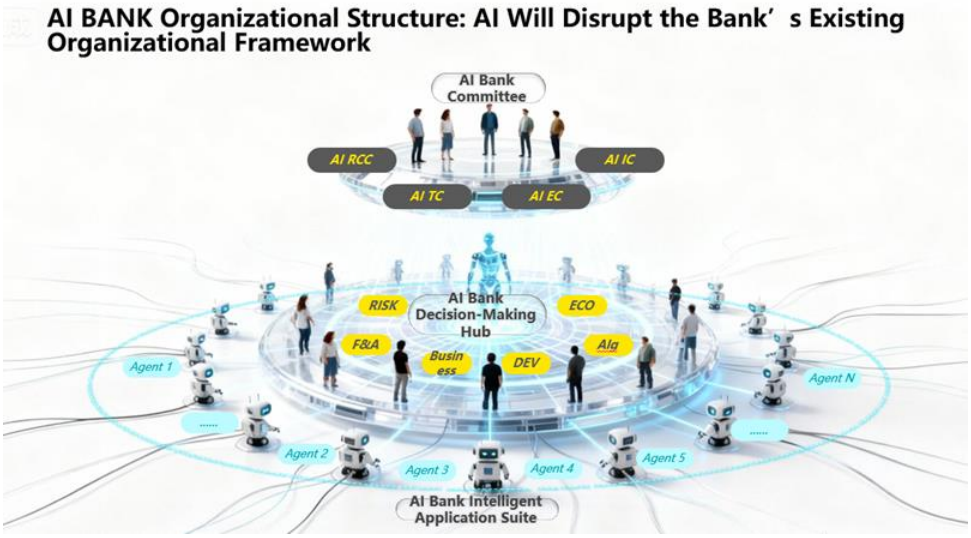
(Figure 2.2.1)



(Figure 2.2.2)

The intelligent evolution of the banking industry ultimately will point to the in-depth integration of artificial general intelligence (AGI). Once AGI technology is adapted to financial scenarios, the hierarchical organizational structure of traditional banks will undergo essential reconstruction. The following AI bank organizational structure (see Figure 2.2.3) provides a forward-looking and concrete depiction of the bank governance and operating model in the AGI era: the structure positions the AI decision-making center as the core brain, supported professional committees such as risk management, technology and ethics, forming a top-level governance system, and then enables front-end execution through a large networks of AI agent applications, constructing an intelligent closed loop of "decision-making - supervision - execution."

This structure not only breaks down traditional departmental boundaries but also outlines a new financial service landscape in the AGI era where AI becomes the core driver of business operations and humans focus on strategic governance and value optimization, providing a clear direction for the organizational evolution required to advance the intelligent transformation of the banking industry.



(Figure 2.2.3)

2.3 Chinese policies: Accelerating "AI +" in a steady and orderly manner

China's banking industry is at a pivotal moment in its evolution from digitalization to digital intelligence. According to recent industry data, AI and generative AI investment in China's financial industry reached RMB19.694 billion in 2024. The market size of generative AI platforms and application solutions stood at approximately RMB914 million and is expected to reach 3.5 billion yuan in 2027.[2] These figures reflect a fundamental shift: AI is no longer an optional enhancement in bank applications, but is rapidly becoming a core engine of business growth.

China's development path for AI-enabled finance is not one of unchecked expansion, but one guided by a rigorous, comprehensive and supportive policy framework that embodies the principle of balancing development and security. This "steady and orderly" approach has provided clear governance direction, ensured risk-aware innovation and created an environment in which financial institutions can accelerate large-scale AI adoption with confidence.

Fundamental principles: Regulators such as the People's Bank of China have clearly emphasized the need to **"steadily, safely and orderly promoting the application of artificial intelligence large models in the financial sector,"** establishing a defined and secure channel for innovation.

Strategic drivers: National-level initiatives – including the "AI +" Action Plan and the focus on data elements and modern industrial systems in **the "15th Five-Year Plan"** – have provided unprecedented opportunities for AI-enabled financial scenarios. Fintech is expected to serve the national strategies of **the "Five Major Articles"** – sci-tech finance, green finance, inclusive finance, pension finance and digital finance.

Ethical bottom lines: Norms such as the "Guidelines for Science and Technology Ethics in the Financial Field" set out essential principles including "upholding integrity and innovation" and "putting people first". These ethical boundaries ensure that innovation does not deviate from responsible and trustworthy development.

At the financial industry regulatory level, central regulatory authorities are focusing on top-level design and risk bottom lines while continuously improving institutional frameworks. At the same time, regional hubs – including Beijing, Shanghai, Shenzhen, Guangzhou and Hong Kong – are actively building innovation testbeds and application scenarios based tailored to their industrial foundations and development needs. Together, these efforts are moving AI-enabled financial services from policy-level guidance to practical implementation.

- The People's Bank of China's Fintech Development Plan (2022-2025) is now approaching its final stage, laying the groundwork for the sector's transition from digitalization to digital intelligence. The central bank has also introduced a series of standards and guidelines – including the *Evaluation Specifications for the Financial Application of Artificial Intelligence Algorithms*, the *Guidelines for Science and Technology Ethics in the Financial Field*, and the *Information Disclosure Guidelines for the Financial Application of Artificial Intelligence Algorithms*. These documents establish clear requirements for science and technology ethics and algorithm governance in the application of AI models within the financial sector.

- In 2024, the National Financial Regulatory Administration (NFRA) issued the *Guiding Opinions on Banking and Insurance Institutions Doing a Good Job in the 'Five Major Articles' of Finance*, calling on banking institutions to apply AI and other technologies to optimize financial service processes under sound risk control. The document also sets out clear timetables and roadmaps for implementation.

- The NFRA has also maintained a high vigilance over emerging risks associated with AI. Its 2024 *Measures for the Management of Data Security of Banking and Insurance Institutions* introduce specific requirements for information disclosure and risk mitigation when banking and insurance institutions develop and deploy AI models, reinforcing the regulatory foundation for safe and trusted AI in finance.

Beijing, Shanghai, Shenzhen, Guangzhou and Hong Kong are showing a pattern of "adapting to local conditions and letting a hundred flowers bloom," promoting the precise implementation of policies and resources through the integration of finance and technology.

As the national financial management center, Beijing issued the *Opinions on Promoting the High-Quality Development of Digital Finance in Beijing* and, building on the *Several Measures for Promoting the Innovative Development of General Artificial Intelligence in Beijing*, is prioritizing support for financial large-model R&D and major scenario demonstrations.

As an international financial center, Shanghai emphasizes original technological research – including AI and privacy-preserving computing – in the *14th Five-Year Plan for the Construction of Shanghai International Financial Center* and the *Action Plan for Shanghai to High-Quality Promote the Construction of a Global Fintech Center*. It encourages Shanghai Clearing House to use AI to enhance cross-border payment and settlement efficiency, with a focus on foundational technology breakthroughs and financial infrastructure development.

As a hub for cross-border finance and supply chain finance, Shenzhen has issued *Implementation Opinions on Promoting Digital Finance to Support the Development of a "Digital China" Model in Shenzhen* and the *Action Plan for Jointly Building a Hong Kong-Shenzhen Global Fintech Center (2025-2027)*. These policies encourage financial institutions to leverage AI, big data and cloud computing, while proposing forward-looking initiatives such as joint incubators with Hong Kong and cross-border central bank digital currency (CBDC) bridge research.

As the first city in China to introduce local policy documents for the investment advisory industry, Guangzhou issued the *Several Measures for Guangzhou to Vigorously Promote the Development of AI Investment Advisory*. This policy aims to address bottlenecks in scaling AI-based investment advisory services, advance investment-side reforms in the capital market and support the shift of the wealth management industry toward a buyer-advised model powered by fintech, fostering a high-quality investment advisory ecosystem.

The Hong Kong Monetary Authority (HKMA) launched the industrywide *Future Bank Capacity Building* initiative in July 2019 to support future business development needs. In 2024, the HKMA updated its *Supervisory Policy Manual* on talent development, recommending that banks use the *Banking Practitioners' Competency Enhancement Framework* as a competency benchmark. In 2025, it released the *Future Bank Capacity Building 2026-2030* report, which assesses potential talent gaps and identifies three core future skill domains for banking professionals: technical skills in AI, data and cybersecurity; soft skills in human-machine collaboration and human-centered service; and banking knowledge in areas such as sustainable finance risk management and emerging markets.

2.4 Chinese practices: AI application exploration and agent leapfrogging

The exploration of AI in China's banking industry has evolved from early single-point assistant tools to systematic application practices covering the front, middle and back offices, and is gradually moving towards a new stage centered on the construction and collaboration of intelligent agents, striving to reshape structures and processes and explore new value networks and business models.

2.4.1 Panorama of AI application system in China's banking industry

Based on in-depth observation of industry practices, this report constructs an AI application system centered on **Grow the Bank, Protect the Bank and Run the Bank,** starting from the core mission of the banking industry. The framework expands into four key areas and aims to provide a panoramic reference for the AI strategic layout of financial institutions:

- **Customer management and growth:** Focus on **grow**, reconstructing end-to-end customer journeys from precise reach to personalized experience.
- **Risk and compliance management:** Focus on **protect**, building a real-time, proactive, and penetrative intelligent risk control and compliance system.
- **Operation and treasury management:** Based on **run**, driving leapfrog improvements in operational efficiency and decision-making quality through process intelligence.
- **Technology and data infrastructure:** As the foundation, consolidating the intelligent foundation across computing power, data, models and R&D to support Grow, Protect and Run.

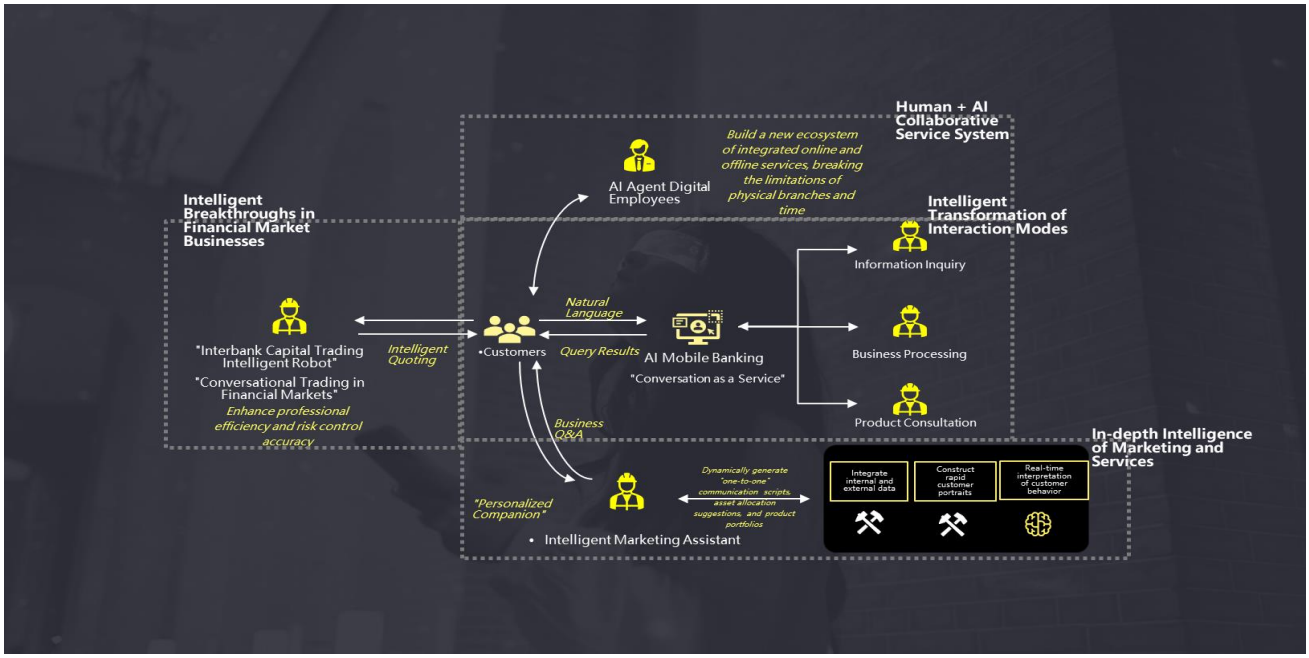
Customer Operation and Growth COO, CMO	Risk and Compliance Management CRO, CCO	Operation and Treasury Management COO, CFO	Technology and Data Infrastructure CTO, CIO, CDO
Intelligent Marketing Assistant	End-to-end intelligent credit approval	Enterprise-level Knowledge Q&A Platform	Full-stack self-developed financial large model technology
Marketing script generation	Credit due diligence report assistant	Corporate Contract Management	Intelligent agent development platform
Intelligent Wealth Manager	Intelligent Credit Pre-Approval	Meeting planning and minutes generation	AI computing power management platform
Customized Financial Product Recommendation	Dynamic risk monitoring and early warning	Automated Generation of Press Releases and Training Materials	Intelligent Financial AI Inference Acceleration Applications
Intelligent Customer Service and Call Handling Summary	Supply Chain Finance Risk Control and Early Warning	Intelligent writing and official document polishing	Enterprise-level knowledge base management
Customer portrait sketching	Intelligent Detection and Early Warning of Market Risks	Email Assistant	Data Asset Retrieval Optimization
Product portfolio and innovation	Anti-money Laundering Intelligent Monitoring	Business Intelligent Review	AI + BI (Intelligent Analysis and Q&A)
Product Comparison and Benefit Evaluation	Intelligent Anti-Fraud Monitoring	Intelligent Service Training Assistant	R&D Intelligent Assistant
AI Mobile Banking ("Conversation as a Service")	Intelligent Compliance Review	Financial document automatic review	Intelligent Code Review and Completion
Interbank Fund Trading Intelligent Robot	Intelligent Policy Management	Financial Intelligence Report and Q&A	Codebase Migration Workflow
Conversational Trading in Financial Markets	Intelligent Contract review	Intelligent Prediction and Allocation	Synthetic Data Generation
Automated Generation of Market Investment Research Reports	Audit Report Assistant	Dynamic Risk Hedging	Test Case Generator

Source: Public information, EY analysis

1) Customer management and growth: From transaction processing to value symbiosis

AI is profoundly changing the interaction model between banks and customers, promoting the paradigm of customer management and growth to leap from standardized services to a personalized, proactive and scenario-based value symbiosis model.

- **Intelligent transformation of interaction models**: Conversation as a Service, intelligent customer service and digital employees are breaking the constraints of traditional menu- and form-based interactions.
  - *A city commercial bank launched an AI mobile banking app featuring "Conversation as a Service," allowing customers to complete information queries, business handling, and product consulting through natural language.*[3]
- **Human + AI collaborative service systems**: Built a new ecosystem of integrated online and offline services, breaking the limitations of physical branches and time.
  - *A certain joint-stock bank launched an AI-powered digital employee that handles 80% of the group's customer service volume, increasing the AI resolution rate from 38% to 92%.*[4]
- **Intelligent upgrades to marketing and services:** Intelligent marketing assistants integrate internal and external data to construct customer portrait sketches. AI can real-time interpret customer behaviors, dynamically generate one-to-one communication scripts, asset allocation suggestions and product combinations, promoting mass marketing to personalized companionship. AI is extending beyond financial scenarios, deeply embedding bank services into customers' production and living scenarios through vertical domain agents, upgrading from financial experts to industry partners and life assistants and building new customer value propositions.
  - *Some banks are leveraging large models to upgrade their retail intelligent service systems: transforming customer service bots into proactive financial companions that intelligently analyze and anticipate wealth management needs, and into expert advisors capable of not only answering questions but also explaining reasons and providing actionable recommendations.*
- **Intelligent breakthroughs in financial market business:** Interbank fund trading intelligent robots and conversational trading in financial markets show how AI is accelerating automation and intelligence in highly specialized areas such as interbank markets and fund trading. This enhances professional efficiency and improves the accuracy of risk control.
  - *Some banks offer intelligent pricing for small transactions, significantly improving customer transaction efficiency and inquiry volumes, unlocking new service reach and profit potential in the long-tail market.*

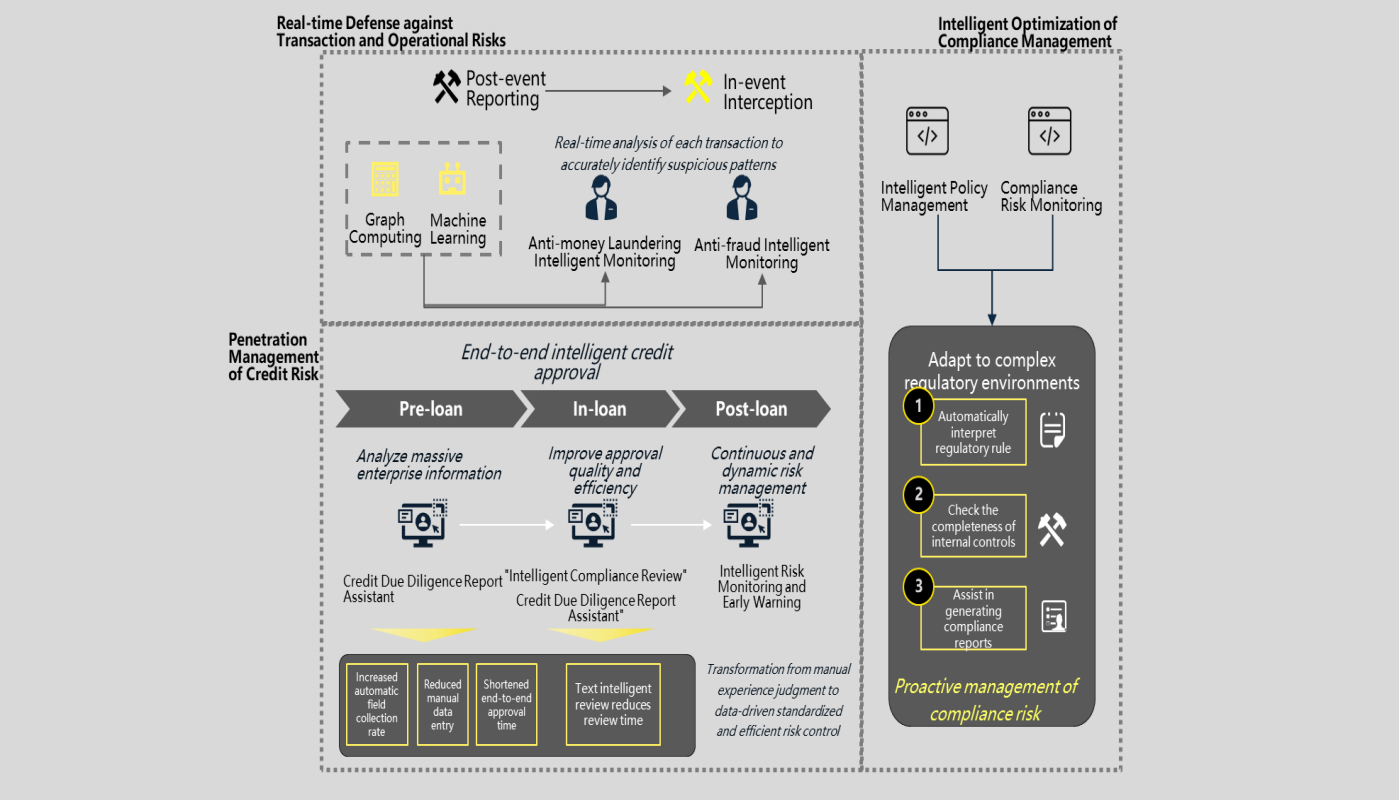


(Figure 2.4.1)

2) Risk and compliance management: A solid line of defense from human prevention to intelligent prevention

In a highly regulated and high-risk industry, the application of AI in risk control and compliance has become the lifeline and core competitiveness of banks' stable operations, promoting risk management from post-event response to pre-event early warning and in-event intervention.

- **Penetrative management of credit risk:** End-to-end intelligent credit approval runs through pre-loan, in-loan and post-loan stages. Credit due diligence report assistants can quickly analyze massive enterprise information; intelligent compliance review and credit approval drafting assistance improve approval quality and efficiency; intelligent risk monitoring and early warning enable continuous and dynamic risk management.
  - *A certain bank is leveraging large models to drive intelligent upgrading across the entire credit approval process. In the pre-loan stage, the rate of automatic data field collection has significantly increased, reducing manual data entry by nearly 80% and cutting the overall approval cycle by half. In the mid-loan stage, intelligent text review has reduced the time required for each review by 15%, enabling a transition from manual, experience-based decision-making to data-driven, standardized, and efficient risk management.*[5]
- **Real-time defense of transaction and operational risks**: Intelligent anti-money laundering monitoring and intelligent anti-fraud monitoring can analyze each transaction in real time, using graph computing and machine learning to accurately identify suspicious patterns from billions of transaction data, promoting the prevention and control of risks such as money laundering and fraud from post-event reporting to in-event interception.
  - *Some banks have reduced their anti-money laundering (AML) screening workloads by implementing end-to-end AI-powered AML detection, while significantly increasing AML reporting rates, successfully establishing a new, efficient, and precise paradigm centered on AI.*
- **Intelligent optimization of compliance management**: Applications such as intelligent system management and compliance risk monitoring help banks cope with the increasingly complex regulatory environment. AI can automatically interpret regulatory rules, check the completeness of internal controls, and assist in generating compliance reports, enabling banks to shift from passively meeting regulatory requirements to proactively managing compliance risks.

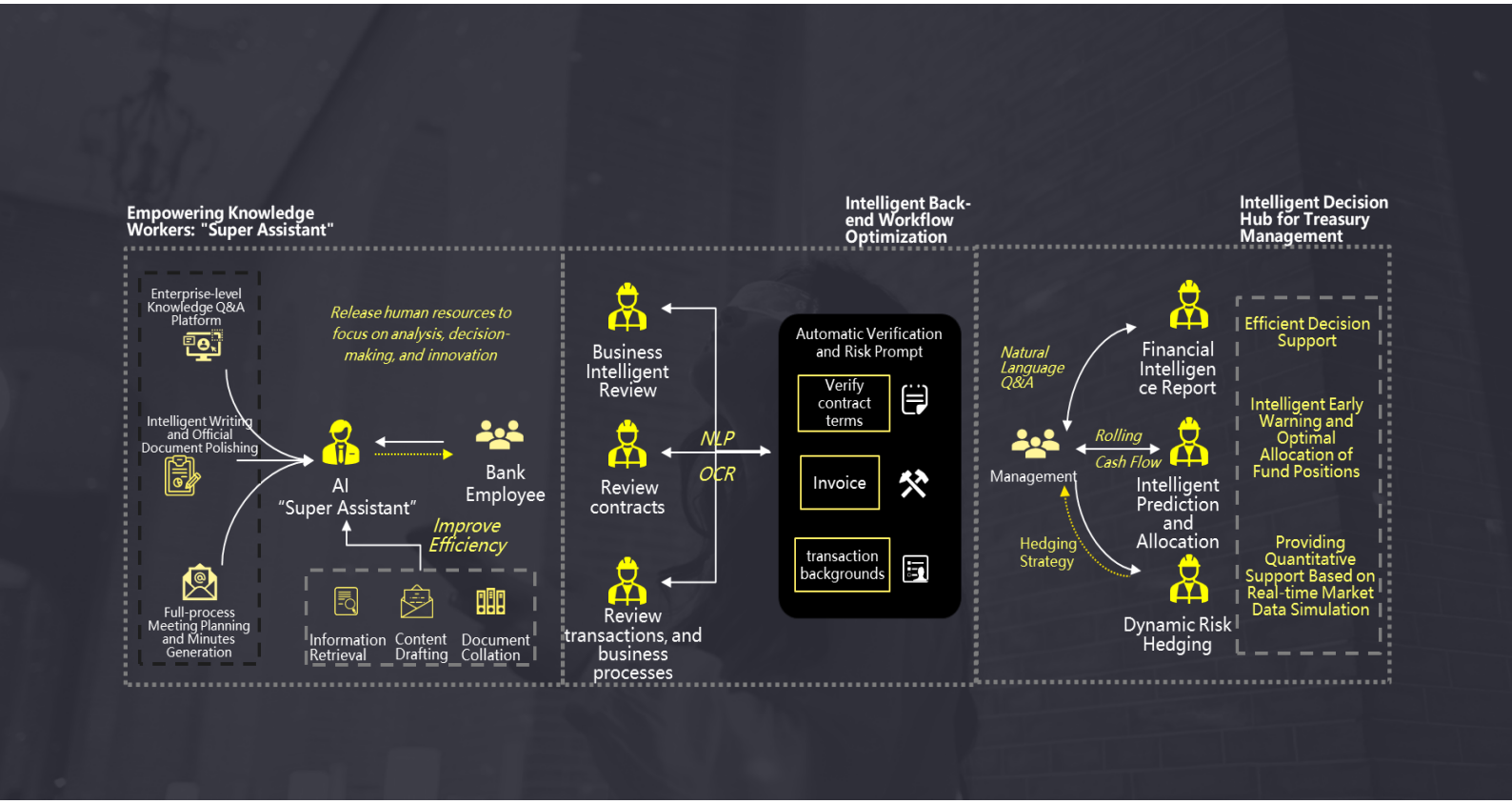


(Figure 2.4.2)

3) Operation and treasury management: From process optimization to intelligent decision-making

Operational efficiency and financial stability are the engines of banks' sustainable development. By improving the automation and intelligence of business operations, AI is building a more agile, accurate and cost-effective intelligent organizational support system.

- **Super assistants for knowledge work**: Applications such as enterprise-level knowledge Q&A platforms, intelligent writing and document polishing and full-process meeting planning and minutes generation directly empower banks' core knowledge workers. AI has become a super assistant for each employee, taking over many repetitive and time-consuming tasks such as information retrieval, content drafting and document sorting, freeing up human resources to focus on analysis, decision-making and innovation.
  - *A major state-owned bank has integrated large models into its corporate banking operations, empowering staff to efficiently handle repetitive tasks such as knowledge retrieval, document review and client survey report generation, resulting in average time savings of over 30%. This enables employees to focus on higher-value activities such as client consulting, product promotion and marketing services, significantly enhancing the overall effectiveness of corporate banking services.*[6]
- **Intelligence of back-office workflows**: Intelligent business review, enterprise contract management and automatic review of financial documents realize automatic verification and risk prompting of contract clauses, invoice vouchers and transaction backgrounds through NLP and OCR.
  - *A certain bank has enhanced its financial document image text recognition product through large model technology, expanding the number of supported bill types by over 40% and achieving recognition accuracy exceeding 99% for multiple bill categories, significantly improving operational quality and efficiency.*[7]
- **Intelligent decision-making center for treasury management**: Intelligent treasury reports support senior management to conduct efficient interactive Q&A through natural language, assisting decision-making. Intelligent forecasting and allocation conducts rolling forecasts of the bank's cash flow, realizing intelligent early warning and optimal allocation of funds. Dynamic risk hedging simulates in real time the impact of interest rate and exchange rate fluctuations on the balance sheet based on market data, providing quantitative support for hedging strategies.

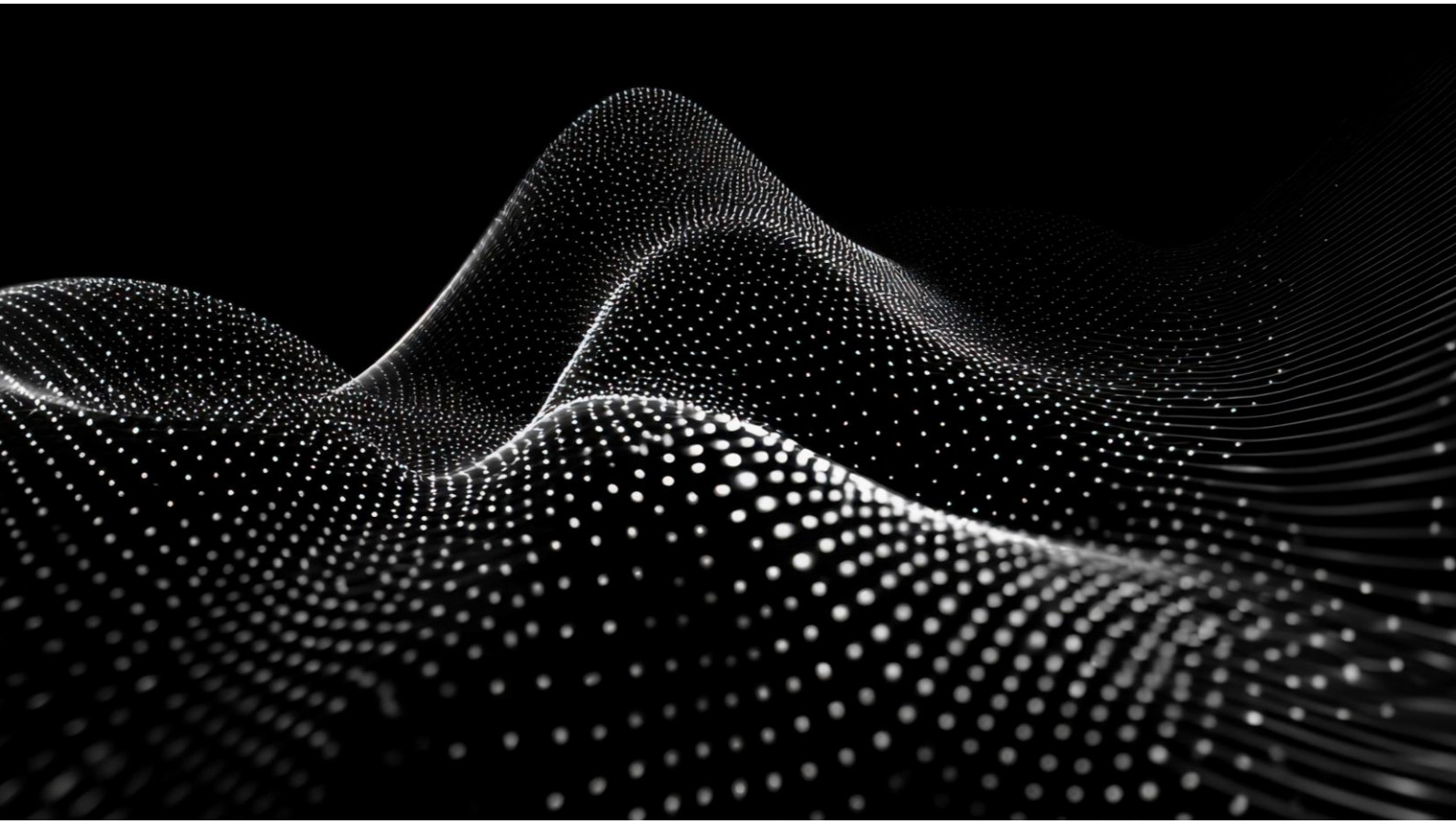


(Figures 2.4.1.4)

4) Technology and data infrastructure: From support system to core driving engine

Large-scale innovation of front-end applications relies on a solid, agile and intelligent technology and data foundation. AI itself is also profoundly changing the way technology systems are built and operated, and is precipitating into the bank's strategic basic capabilities.

- **AI-native infrastructure**: Large banks are shifting from purchasing external models to building independent and controllable underlying AI capabilities, including the construction of full-stack self-developed financial large model technology systems, AI computing power management platforms and agent development platforms. These platforms provide high-performance, low-cost, safe and compliant model training and inference services for upper-layer intelligent applications and are the operating systems for banks' intelligent transformation.
- **In-depth activation of data value**: Enterprise-level knowledge base management, synthetic data creation and data asset retrieval optimization directly address two core challenges of the AI era: high-quality data supply and efficient knowledge utilization. Banks are transforming unstructured documents scattered everywhere into structured knowledge components, and solving data privacy and scarcity problems through synthetic data technology, making data truly a resource understandable and callable by large models.
- **R&D paradigm shift**: Intelligent code review and completion, test case generators and R&D intelligent assistants deeply integrate AI into the software development life cycle. Developers pair program with AI, improving both efficiency and code quality, and transforming technology delivery capabilities.
  - *One bank is advancing coding intelligence, with the platform's active users accounting for over 90% of the bank's total developers, and monthly active users reaching nearly 75%. This has fundamentally transformed the development process and team collaboration model.*[7]



2.4.2 Agent leapfrogging – From AI empowerment to AI-native

Building on the systematic advancement of AI applications across the four major domains, leading Chinese banks are now moving towards a new stage: shifting from using AI tools to optimize processes to reconstructing business models with intelligent agents as the core. At the heart of this AI-native paradigm is the creation of super digital employees capable of independent planning and collaborative execution. The critical enablers are two foundational capabilities: multi-agent collaboration and AI operating system-level control.

- Multi-agent collaboration: Future agents are not isolated tools, but digital legions that can automatically combine and collaborate to achieve complex goals. Their value lies in breaking departmental barriers and forming group intelligence oriented towards complex objectives.
  - *One internet bank, in collaboration with the Ministry of Agriculture and Rural Affairs, launched a farmer-specific AI agricultural assistant. Through multi-agent collaboration, it integrates various agents such as meteorological experts, agricultural technology experts and financial experts, and accurately provides planting plans, policy interpretations and financial services according to farmers' locations and crop cycles, realizing full-cycle intelligent companionship from production to funds and creating a new model for integrated agriculture-plus-finance services.*[8]

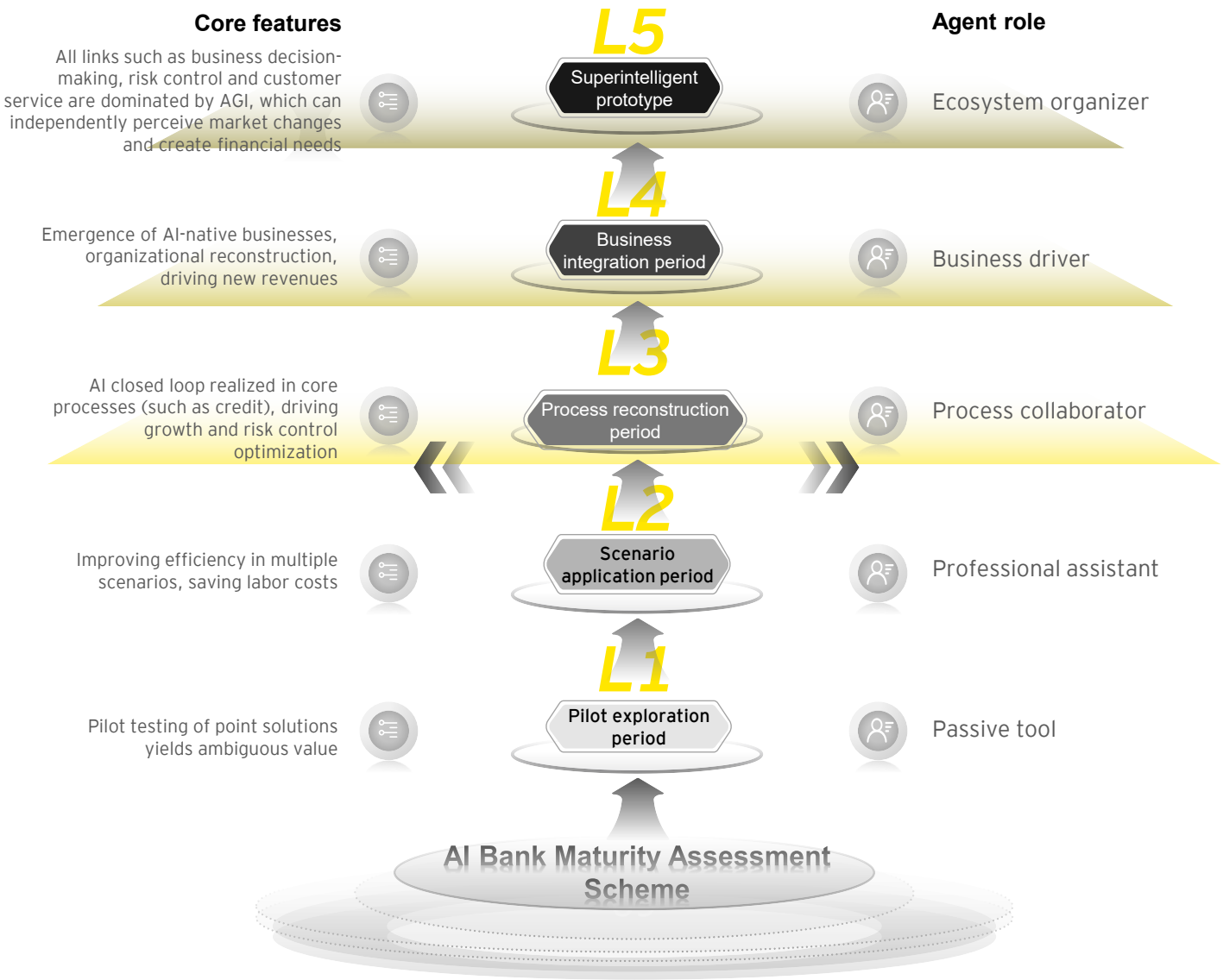
- AI operating system-level control: The large-scale application of agents requires a unified AI operating system to achieve ecological-level control over the agent life cycle, task scheduling, and compliance.
  - *A major state-owned bank has built an enterprise-level large model technology system based on a high-performance AI computing foundation and trillions of financial data points. Using a 1+X large-scale application paradigm – one intelligent center collaborating with multiple professional agents – and a self-developed intelligent central control scheduling model, the bank had, as of October 2025, deployed intelligent capabilities across more than 400 business scenarios, achieving efficient transformation from technical capability to enterprise-wide business empowerment.*

In summary, the agent-led AI-native paradigm derives its fundamental value from enabling AI to become the core engine of high-quality and sustainable bank development. Through multi-agent collaboration, it reshapes the depth and breadth of service to the real economy, and through AI operating-system control, it enables highly refined and agile operations and management. The goal of this transformation is not technological advancement for its own sake, but a fundamental leap in value logic – building a development model driven by data and intelligence that positions banks to remain resilient and competitive in the future landscape.

3 EY AI Bank Maturity Assessment Scheme

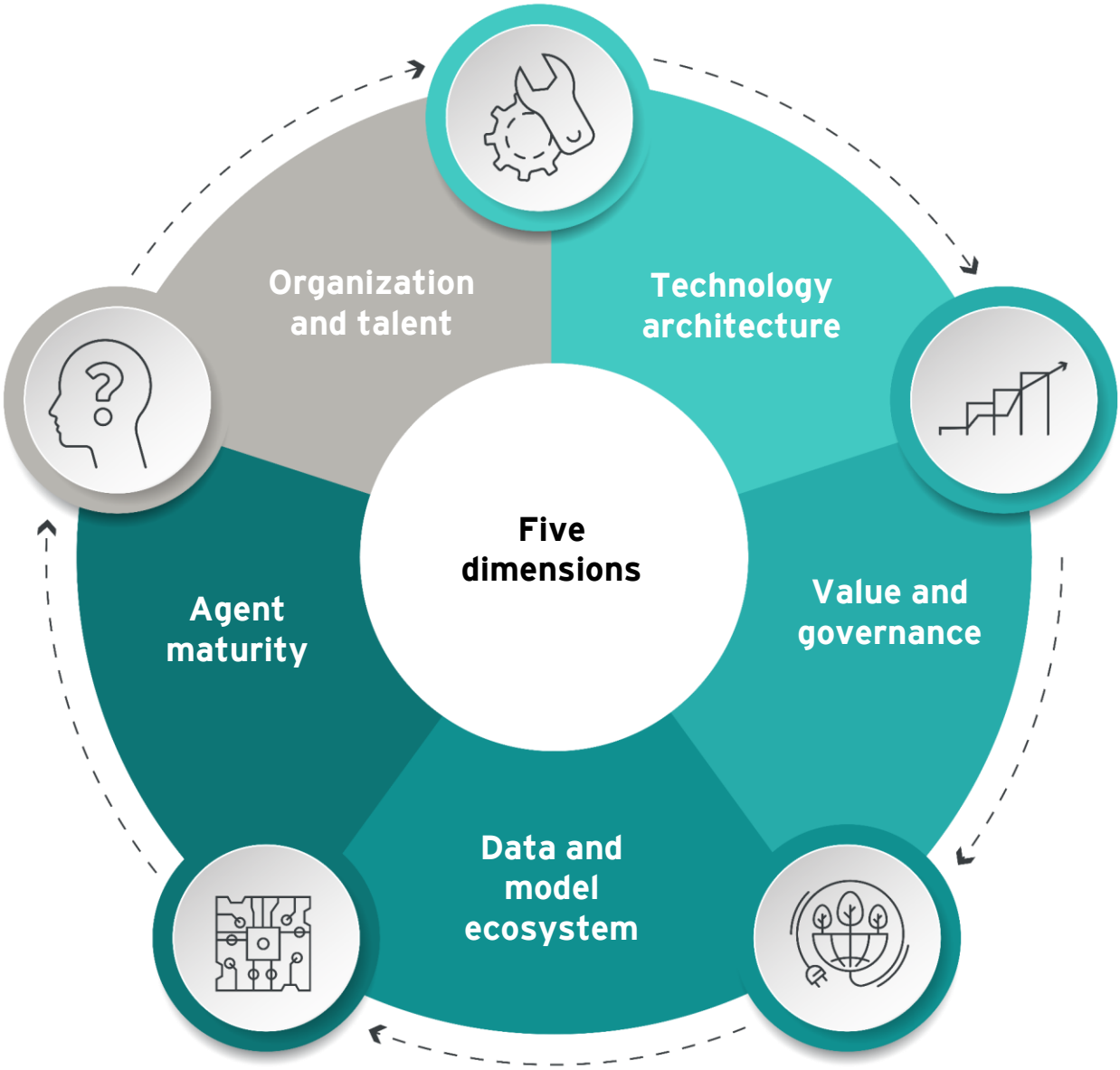
3.1 Model introduction: Five-stage evolution – defining the future

To help institutions accurately position themselves, we have constructed the EY AI Bank Maturity Assessment Scheme. Taking the evolution of intelligent agents as the axis, this model divides the bank's AI transformation process into five stages:



3.2 Assessment dimensions: Five perspectives – comprehensive insight

> The assessment of the bank's AI transformation process is carried out across five dimensions: agent maturity, data and model ecosystem, organization and talent, technology architecture as well as value and governance.



The assessment of the bank's AI transformation process

1 Agent maturity: Capability leapfrogging from rule execution to complex reasoning and independent evolution

Agent maturity is the core measurement dimension of AI bank transformation. Its evolution from L1 to L5 represents a capability leapfrogging from passive rule-based tools to superintelligent prototypes. The core characteristics of each stage can be summarized as follows:

L5 (Superintelligent prototype): Agents evolve into superintelligent prototypes. As the core organizers of the financial ecosystem, they can deeply connect multi-source data and the capabilities of external industrial ecosystems, and independently complete ecological resource allocation, value distribution and complex risk deduction. At this stage, agents assume roles such as the bank's chief data scientist and chief strategy officer, discovering deep patterns that human experts cannot detect in the bank's historical operational data, forming innovative and in-depth industry insights, proposing disruptive business strategies and becoming the intelligent hub that enables AI banks to integrate across industries.

Super-intelligent prototype

L4 (Business integration period): Agents become engines of business innovation, capable of exploring customers' potential needs, designing AI-native financial products and enabling cross-departmental capability collaboration. At this stage, the agent system begins to exhibit self-governance. While completing tasks, it can monitor its own operating status in real time, identify potential issues early and take corrective and optimization actions, driving an upgrade of the bank's revenue structure through innovation.

Business integration period

L3 (Process reconstruction period): Agents gain cross-link process collaboration and dynamic decision-making capabilities, and can be embedded into core processes such as credit and operations to achieve end-to-end intelligent closed loops that drive business process reconstruction. The core of this stage is collaboration and decision-making. Through collaboration, the agent connects business entities, semantic relationships, state changes and executable actions, significantly improving decision-making efficiency and enabling end-to-end automation of the entire process.

Process reconstruction period

L2 (Scenario application period): Agents develop basic data perception and simple reasoning capabilities, and can serve as professional assistants in specific scenarios – such as wealth recommendations and preliminary credit review – reducing the workload of manual data processing. The core of this stage is the development of a tool ecosystem. Agents need to master a range of capabilities, including structured queries, unstructured analysis, rule engines and external APIs, but cross-scenario collaboration and complex decision-making still rely on humans.

Scenario application period

L1 (Pilot exploration period): Agents can execute fixed rules but have no independent perception and reasoning capabilities, functioning as mechanical tools attached to a single link. The core value of this stage is replacing basic manual operations and achieving foundational efficiency improvements.

Pilot exploration period

Data and model ecosystem: Evolution from data silos to real-time assets and from single models to model factories

In the digital transformation of banks, data and models are core strategic assets. The evolution of their management and application capabilities determines the level of intelligence and business innovation a bank can achieve. Traditional data silos and chimney-style AI applications can no longer meet the needs of agile response and real-time risk control. Across the five stages, banks' data and model ecosystems follow a clear evolutionary path – from local pilots to global intelligence, from static data to real-time assets and from single-point models to large-scale production.

Super-intelligent prototype

L5 (Superintelligent prototype): This is the highest stage, where banks build an open intelligent financial ecosystem defined by ecological interconnection and continuous self-evolution. On the data front, banks function as nodes that collaborate with external partners under secure frameworks to form a cross-domain panoramic data view, which drives innovation at the societal level. On the model front, the model factory evolves into an open AI ecosystem platform that supports deployment, joint modeling and federated learning across internal and external models, and also outputs capabilities through model marketplaces. Models within the ecosystem can continuously and independently optimize based on real-time feedback, forming a dynamic intelligent network and ultimately realizing the qualitative shift from AI banking to banking built on AI.

Business integration period

L4 (Business integration period): This stage enables deep integration of AI and core business, marked by AI nativeness and value creation. AI is inherently embedded into business processes, directly driving large-scale value. At the data level, the enterprise-level data middle platform supports real-time services, transforming data into assets that flow securely through APIs. At the model level, the enterprise-level model factory provides full lifecycle automated management, supporting assembly-line production and continuous optimization. Breakthroughs in large-model technology give rise to next-generation applications such as superintelligent customer service. Banks develop a dual engine of data assets and model factories, moving services toward decision intelligence.

Process reconstruction period

L3 (Process reconstruction period): AI begins to drive deep process reconstruction, supported by a cloud-native, microservices-based AI middle platform oriented toward process optimization. It enables efficient model development, deployment and monitoring, and forms tight coupling with business systems. Computing infrastructure allows multi-model parallelism and large-scale inference, meeting requirements for high-concurrency, low-latency real-time decision-making with elastic scaling. MLOps is introduced to achieve semi-automated operations with autonomous fault diagnosis. The goal at this stage is to support process intelligence and automated decision-making.

Scenario application period

L2 (Scenario application period): Following initial pilot success, AI moves into deeper application within specific scenarios, characterized by scenario-driven, partial enablement. On the data side, department-level or scenario-specific data platforms begin to emerge, connecting data across relevant business lines to support intelligent risk control and precision marketing, but cross-departmental data movement remains limited. On the model side, independent models are developed and deployed for key scenarios with initial performance monitoring, but development, deployment and operations remain fragmented with low reusability.

Pilot exploration period

L1 (Pilot exploration period): This is the embryonic stage of bank AI, characterized by point-like experimentation and technical validation. Data environments remain highly siloed, with scattered data, low integration, inconsistent quality and mostly offline batch processing. On the model side, only single-model explorations are conducted around isolated use cases such as anti-fraud and churn prediction. Model lifecycle management and risk management mechanisms are almost nonexistent, relying heavily on manual oversight. The focus at this stage is to validate technical feasibility and accumulate experience, but scale remains limited.

3 Organization and talent: Cultural and organizational transformation for business-technology integration

The essence of AI reshaping banks is not only the iteration of technology, but also a profound transformation of the integration paradigm of productive forces and production relations. As a core dimension for measuring the transformation of AI banks, the evolution of organizational and talent maturity from L1 to L5 reflects the progression of the integration model between business and technology (AI).

L5 (Superintelligent prototype): Ecological organizational model, with an extremely flat structure, forming a dynamic value network across enterprises. A large number of decisions and executions are independently completed by AI agents, and humans step back to serve as ethical and strategic governors. The talent structure is a pine tree-shaped, high-density structure, retaining only top strategists and ethical experts. Organizational expansion no longer relies on hiring people, but on the growth of computing power, achieving true human-machine symbiosis.

Super-intelligent prototype

L4 (Business integration period): Liquid organizational model, dynamically forming teams based on goals, with blurred organizational boundaries. The traditional hierarchical system is dissolved, and innovation is no longer driven by instructions, but emerges from the bottom up by front-line employees augmented by AI capabilities. The talent structure is an inverted diamond shape, with primary operational positions largely replaced, and experts with AI thinking becoming the main force. AI becomes the primary productive force driving business growth. Performance appraisal is no longer limited to individuals, but reflects the comprehensive effectiveness of humans + AI.

Business integration period

L3 (Process reconstruction period): Federal collaborative model, forming cross-functional agile teams. The traditional hierarchical system is broken, and each team has end-to-end decision-making power. Interdisciplinary talents rise, and AI product managers and business architects become the hub connecting technology and scenarios. AI is embedded into core processes as a collaborator. The talent structure transitions to an inverted diamond shape, and the assessment mechanism begins to cover the joint output of cross-functional teams.

Process reconstruction period

L2 (Scenario application period): Centralized empowerment model, starting to establish a centralized AI team or Center of Excellence (CoE) as a public functional department to output general capabilities to various business lines. The boundary between technology and business organizations is still clear, showing an order-taking service relationship. The professional division of talent has initially emerged: algorithm engineers are introduced on the technology side, and AI translators appear on the business side. The role of AI is upgraded to a professional assistant. The talent structure still maintains a pyramid shape, but the proportion of middle-level technical backbones has slightly increased.

Scenario application period

L1 (Pilot exploration period): Decentralized model, with no dedicated and centralized AI functional department established. AI capabilities are scattered within various technology teams, in an isolated state of each fighting on their own. Talent are mainly part-time traditional technology personnel, lacking professional algorithm engineers and AI product managers. Business personnel generally lack AI thinking, and AI is treated as a passive tool, and the talent structure presents a pyramid shape.

Pilot exploration period

4 Technology architecture – from passive support to AI-native openness

As the core foundation of AI banks, the evolution of technology architecture is always deeply bound to scenario iteration and business needs, centering on three main lines: support upgrade, elasticity improvement and open collaboration. It has experienced a step-by-step evolution from passive adaptation tools to active support middle platforms, and then to open AI native. Initially, it supports pilot implementation through lightweight transformation; in the middle stage, it builds an AI-native architecture to achieve capability reuse; in the later stage, it forms an open ecosystem built on cutting-edge technologies. Its maturity directly determines the deployment efficiency, scenario coverage, innovation speed and ecological collaboration capabilities of AI applications, and is the core support and key constraint for banks' intelligent transformation.

Super-intelligent prototype

L5 (Superintelligent prototype): Banks evolve into AI-native super-intelligent prototypes. The technology architecture is open and boundless, breaking internal and external barriers. It builds a distributed computing power network, integrates cloud-edge-end resources, achieves ultra-large-scale scheduling and supports continuous learning and self-evolution of models. The architecture is highly scalable, able to quickly adapt to ecological scenarios and independently iterate. It has intelligent operation and maintenance capabilities, enabling autonomous optimization and self-repair of the architecture. The goal is to support cross-domain collaboration and ecological operation.

Business integration period

L4 (Business integration period): Achieve deep integration of AI and business and establish a highly elastic and fast-innovating AI-native architecture, following the principle of AI first. It supports agent orchestration and collaboration. It builds a distributed heterogeneous computing power pool, with unified scheduling to support large-scale training and commercial inference. The architecture is elastically scalable, supporting the rapid launch of AI-native businesses. A mature MLOps system achieves full lifecycle automation and fault self-healing. The goal is to support the rapid integration of AI-native businesses.

Process reconstruction period

L3 (Process reconstruction period): AI begins to drive process reconstruction, forming a cloud-native and microservice-based AI middle platform oriented towards process optimization. It supports efficient model development, deployment, and monitoring, and realizes deep coupling with business systems. Computing power supports multi-model parallelism and large-scale inference, meeting the needs of high-concurrency, low-latency real-time decision-making and providing elastic expansion. MLOps is introduced to achieve semi-automated operations and maintenance, with independent fault diagnosis capabilities. The goal is to support process intelligence and automated decision-making.

Scenario application period

L2 (Scenario application period): AI applications begin to deepen, and banks build basic AI platforms to achieve centralized resource management and platform-based support. A computing power pool is established to support scenario fine-tuning and inference. AI capabilities are encapsulated as standard APIs and integrated into auxiliary links of core processes through service buses. A basic development and deployment pipeline is established to enable version management. The goal is to deploy AI applications at scale, but architecture coupling remains high, and new scenarios still require customized development.

Pilot exploration period

L1 (Pilot exploration period): Banks initially attempt AI with no unified planning for technology architecture. AI is connected to existing systems as plug-ins or embedded components. Models are mainly purchased externally or drawn from open-source communities, connected through interfaces and dependent on third-party support, with no independent computing power. The architecture is rigid, with high expansion costs and long cycles, making it difficult to support high-concurrency scenarios. The goal is to verify the feasibility of single-point tools, but scalability is poor and broader expansion remains difficult.

Value and governance: Value deepening from cost saving to ecological revenue, and guarantee upgrade from compliance audit to embedded governance

The application of artificial intelligence in the banking industry is driving dual transformations in value creation and governance models. In terms of value, AI has gradually evolved from an auxiliary tool to improve efficiency to a core engine driving revenue growth and shaping differentiation, and ultimately to a value hub capable of reconstructing the financial ecosystem. At the same time, the AI governance paradigm is evolving from passive control to active leadership and from partial response to global co-governance.

L5 (Superintelligent prototype): The core characteristic of AI governance is ecosystem-level collaborative governance and adaptive operation. The scope of governance objects expands from single models and business processes to cross-institutional, cross-scenario ecosystem systems. Banks have established a governance system covering the entire lifecycle of models and achieve dynamic adjustments through data-driven and automated mechanisms. The AI system can support ecological resource allocation, value collaboration and complex risk analysis within the established governance framework. Governance is integrated into ecological operation logic in an endogenous, systematic way, supporting the stable and collaborative development of banks across a wider range of industrial scenarios.

Super-intelligent prototype

L4 (Business integration period): Banks establish a regular, automated and refined AI governance system, with the core characteristic shifting to governance rules embedded in development platforms and business processes. Governance requirements are converted into system capabilities and embedded into the entire process of model development, deployment and operation. The scope of governance objects expands to model platforms and business processes, and rules are written into the system in the form of parameters and thresholds, making compliance an inherent attribute of system operation and supporting high-frequency iteration and deep business integration of AI within a highly trusted boundary.

Business integration period

L3 (Process reconstruction period): Banks form a relatively complete AI governance framework, with the core characteristic of independent model audit and ethical review. AI is systematically incorporated into the existing risk governance framework, and governance activities are model-centered, covering risk identification, performance verification, fairness assessment and ethical evaluation after launch. Governance achieves a transition from disordered to rule-based, but still focuses on post-event control, making it difficult to provide continuous constraints on dynamic risks and large-scale applications.

Process reconstruction period

L2 (Scenario application period): With the gradual implementation of AI across multiple scenarios, banks begin to build an initial governance framework and clarify basic responsibilities, but the governance system remains fragmented. Risk identification is mainly based on case-by-case evaluation and passive response, and only limited compliance checks are carried out in some projects. There is no unified standard and management mechanism, and the overall governance capacity is still difficult to support large-scale applications.

Scenario application period

L1 (Pilot exploration period): Banks have not yet formed an AI governance system and mainly respond passively. AI is not incorporated into the risk and compliance framework, and there is a lack of clear responsibilities, systems and standards. Issues such as model risks, algorithmic bias and privacy security are usually exposed after the event and handled manually, and governance remains in a primary, event-driven state.

Pilot exploration period

3.3 In-depth analysis: case study

Taking the most representative **leap from L3 (process reconstruction) to L4 (business integration)** as an example, four core capabilities must undergo qualitative transformations:

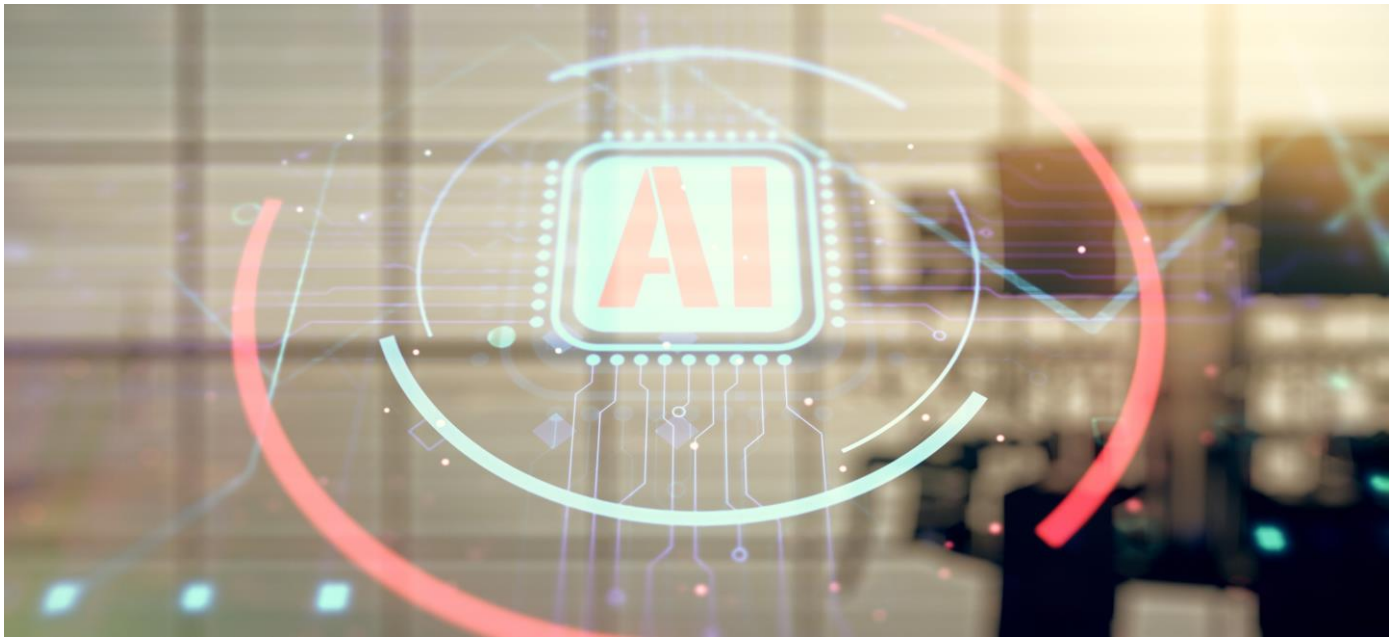
3.3.1 Leapfrogging of customers' digital and intelligent management capabilities

The leap of banks' digital and intelligent customer management systems from L3 (journey reconstruction and intra-departmental connection) to L4 (value co-creation and cross-domain collaboration) marks a capability transformation from intelligent interaction and product matching on specific customer journeys to comprehensive financial solutions based on customers' panoramic scenarios and lifelong value. The core of this evolution is to build an intelligent management network that can actively perceive customers' diverse scenarios, anticipate complex needs, calculate using customers' full lifecycle value, integrate the bank's overall product and resource capabilities and make dynamic responses, promoting the bank's transformation into customers' full lifecycle financial partners and ecological value integrators.

In the L3 stage, banks achieve a significant advancement of digital and intelligent management centered on customer journeys: decision-making models and generative models are deeply integrated through a journey strategy engine, enabling accurate matching of people, products and scenarios at each node of specific journeys, guiding customers to move smoothly along the journey and upgrading product and service experiences. For example, in the retail wealth management business, for individual customers with salary payments, the system identifies customers' occupational situations and predicts their demand for financial products and services; for new employees of universities and hospitals, it gradually recommends credit cards, flexible wealth management and fixed-investment products, and provides personalized fund-arrangement reminders before monthly salary payment to cultivate financial management habits; for customers about to receive year-end bonuses, it provides differentiated asset-allocation recommendations based on customers' asset levels, consumption patterns and historical bonus arrangements to support smooth wealth growth.

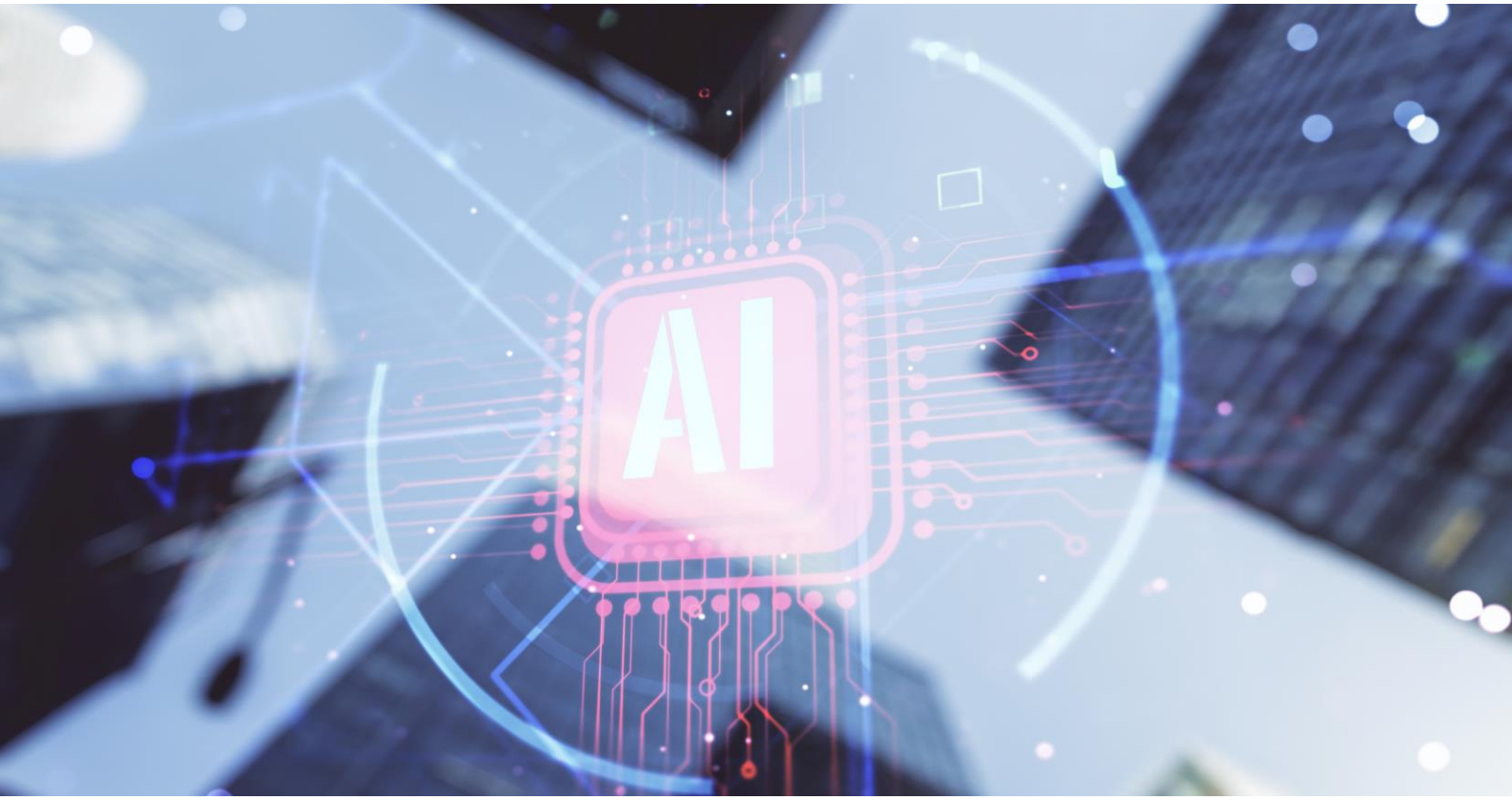
However, the essence of the L3 model is still local optimization. When customer needs span corporate and retail businesses or involve long-term financial planning across life stages, the intelligent strategies of a single journey reveal inherent limitations.

The L4 stage pursues scenario perception and global optimization, with core capabilities that include actively perceiving customers' diverse real-time scenarios, predicting cross-cycle complex needs and dynamically integrating the bank's overall resources through a multi-agent collaborative network to co-create solutions with customers. This leap can be illustrated through the following typical scenarios:



- **From salary payment services to comprehensive financial solutions for enterprises and employees:** In the L3 stage, the salary payment journey focuses on employees' personal fund inflows and wealth management recommendations. Entering L4, the main agent simultaneously perceives changes in the salary payment model of corporate accounts and announcements of enterprises' equity incentive plans, and, combined with macroeconomic and industry trends, proactively issues collaborative instructions: the corporate agent provides human capital analysis and optimized bulk procurement solutions for enterprise HR, while the retail wealth agent synchronously generates differentiated, comprehensive solutions for employees at different tiers, including tax planning for equity incentive exercise, long-term incentive trusts and children's education fund reserves. This realizes a leap from salary payment services to talent value management partnership.
- **From wealth management product recommendations to dynamic companion of family's full lifecycle asset allocation:** In the L3 stage, wealth management sorts out different customer journeys for different scenarios, customer groups and products, forming an intelligent strategy closed loop for each journey. However, due to the complexity of implementation, collaboration between different journeys often cannot be achieved. In the L4 stage, the system constructs a comprehensive family financial view by integrating multidimensional real-time data such as family members' accounts, real estate holdings, insurance policies, future education and pension planning. When predicting that customers' children are about to enter the university application season, the main agent will dispatch the education fund planning agent to evaluate the existing reserve gap, the insurance agent to re-examine family protection and the investment agent to adjust asset allocation suggestions based on market outlook. Generative AI integrates these professional analyses into a dynamically updated family wealth health report and action recommendations, realizing the transformation from single journey service to customers' long-term professional companion and planner.

In summary, the evolution of banks' digital and intelligent customer management systems from L3 to L4 is more than just a technological upgrade – it is a profound transformation of banks' management philosophy and value creation model. By building a cross-domain intelligent collaborative network, banks can closely link customers' full lifecycle value, transform their professional services into financial modules that can be seamlessly embedded into customers' key production and living scenarios and establish a cooperative relationship with customers based on deep trust and value symbiosis, thereby forming their own unique core competitiveness.



3.3.2 Trusted governance

With the deep integration of AI into banks' core businesses, the governance model relying solely on independent model audits and ethical reviews has gradually become unable to adapt to the surging number of models and their high-frequency iterations. The external governance represented by the process reconstruction period (L3), although able to control model risks to a certain extent, has clear limitations in supporting large-scale innovation and real-time risk prevention and control. Entering the business integration period (L4), some leading international financial institutions have begun to embed governance rules into model development platforms and business processes, achieving a step the leapfrogging of trusted governance capabilities through automated and systematic methods.

- Some of the banks widely apply artificial intelligence models in retail credit, anti-fraud, and compliance monitoring. For example,In the L3 stage, its governance mainly relies on the expanded model risk management framework. An independent model validation team reviews model stability, fairness and interpretability. High-risk models must be submitted to the Model Risk Committee for approval, and manual review is conducted regularly after launch. This model emphasizes professional independence, but governance activities are relatively separated from business systems, and there is a time lag in risk identification. Entering the L4 stage, the bank directly codes fairness thresholds, data drift indicators and model confidence requirements into the model release pipeline – models that do not meet the rules are automatically blocked from launch. During model operation, the system monitors in real time differences in customer group rejection rates and changes in input distribution, and automatically downgrades model permissions and switches to manual review once thresholds are triggered, thereby realizing continuous enforcement of governance rules on model behavior.

- Some of the banks in generative AI customer service also reflect the leapfrogging of AI governance maturity. For example,in the L3 stage, the institution established trusted AI principles and ethical review mechanisms, conducting ethical impact assessments and data compliance reviews on generative AI use cases, but still relies mainly on manual sampling during operation. With the high-frequency use of generative AI in customer service, the bank embedded governance capabilities into the model call link in the L4 stage, constructing a multi-layer governance mechanism: first, real-time scanning of model outputs to automatically intercept or rewrite non-compliant statements; second, assessing whether the dialogue exceeds the authorized business scenario and transferring to manual processing if it does; third, automatically upgrading and reviewing conversations that trigger risk rules multiple times. Through mechanized compliance control, compliance has transformed from post-event inspection to automated processing.

- Some of the bank groups reflect a platform-level governance approach. For example, In the L3 stage, the group established a group-wide trusted AI governance framework, implementing strict pre-event approval for AI use cases. Entering the L4 stage, the group shifted governance forward to the technical architecture level, establishing a unified AI access platform. All model calls must be completed through the platform, with has built-in permission control, data isolation and full-link log audit functions. For high-risk scenarios involving credit approval and compliance judgment, the system implements mandatory manual review mechanisms for risk control.

The above cases illustrate that the key to the leapfrogging of AI governance from L3 to L4 does not lie in adding more manual review links, but in converting governance rules into system capabilities, making compliance an inherent attribute of model operation and business processes. Through platform-based governance, real-time monitoring during operation and embedded governance rules, banks can support high-frequency iteration and large-scale application of AI within a trusted boundary, thereby achieving a dynamic balance between innovation efficiency and risk control.

3.3.3 Risk management

From "full-process automated monitoring and approval" (L3) to "self-optimization of risk strategies and API-based output of risk capabilities" (L4), banks are moving toward intelligent, adaptive risk control. For example, agents can simulate risk transmission in supply chain scenarios in real time, dynamically adjust quotas at each link and even output risk control modules to industrial platforms.

The leap of banks' risk management systems from L3 to L4 marks a fundamental transformation from passive response to active evolution. The core of this evolution is to build an intelligent risk control system with self-learning, dynamic adjustment and ecological empowerment capabilities, promoting banks' transformation from risk control executors to strategy planners and ecological enablers.

In the L3 stage, banks' risk management systems have basically realized an end-to-end automated closed loop of "access – monitoring – disposal." For example, in the pre-loan stage, the system enables intelligent access judgment and automated approval decisions through the AI pre-inspection center. In the in-loan and post-loan stages, relying on the real-time risk early warning network – integrating transaction portraits, repayment behavior tracking and dynamic quota control – the system can identify risks and trigger response mechanisms within seconds, completing the initial shift from manual intervention to independent system operation.

Entering the L4 stage, banks' risk management systems have further evolved from automated execution to intelligent decision making, with perception, cognition and independent governance capabilities. Banks' risk management systems can not only monitor in real time their own operating status, but also actively identify abnormalities and implement dynamic optimization. For example:

- By monitoring macroeconomic indicators in real time (such as GDP, CPI, interest rates and industry policies), the system can predict the risk evolution trends of specific industries, simulate their transmission paths in the supply chain and then submit dynamic optimization plans to the risk control decision-making layer in advance. For example, it may automatically lower the credit rating thresholds of upstream and downstream enterprises that are highly associated with the affected industry and tighten new credit policies. Conversely, when identifying supportive signals for emerging industries, it will recommend designing more attractive risk pricing models to help seize market opportunities.
- When detecting that the false positive rate of an anti-fraud model is abnormally high on weekends, the system can automatically adjust the sensitivity threshold or switch to a backup model, then restore the original configuration on working days, realizing adaptive calibration of algorithms.
- It treats the optimization of risk strategies as a continuous learning process. Through simulation testing and related algorithms, it achieves a dynamic balance between maximizing returns and minimizing risks.
- By building an offensive and defensive confrontation simulation environment, banks can actively simulate new attack techniques used by black industries, conduct stress tests on existing rules and quickly generate, verify and deploy protective strategies after discovering vulnerabilities, forming an adaptive defense system with self-evolution capabilities.

At the ecological empowerment level, in the L4 stage, banks can realize a model of risk control as a service. For example, in supply chain finance scenarios, banks can encapsulate dynamic credit risk control modules into standard APIs and embed them into the ERP systems of core enterprises. When suppliers initiate financing applications based on accounts receivable, the risk control API can retrieve in real time multidimensional data such as orders, logistics and acceptance records, dynamically assess transaction risks and verify credit quotas. If the system detects that the payment cycle of core enterprises is lengthening, it can proactively alert suppliers to potential risks and recommend switching to more suitable products such as order financing, realizing seamless collaboration between risk early warning and business guidance.

In summary, the evolution of banks' risk management systems from L3 to L4 is not only an upgrade of technical capabilities but also a strategic transformation of risk control systems from closed operation to open empowerment, and from post-event disposal to pre-event insight. Banks not only strengthen their own risk resistance capabilities but also transform risk control into an exportable core competitiveness, thereby promoting the collaborative development and stable operation of the entire industrial ecosystem.

3.3.4 Back-office technology

- Some of banks are actively building a digital ecosystem spanning all business domains, aiming to achieve dual empowerment through internal efficiency enhancement and external comprehensive outreach. This drives the bank's transformation toward a smarter, more efficient, integrated, innovative and inclusive digital intelligence model, gradually advancing from L3 to L4 maturity. For example, Its AI technology platform has evolved from isolated, point-based applications to a comprehensive foundation that supports organization-wide efficiency gains and end-to-end customer engagement, fostering deep integration and synergistic development between AI technologies and core business processes. In terms of computing infrastructure, the bank has established an AI computing resource pool based on large-scale domestically produced chip clusters, upgrading from standalone computing units to a scalable cluster architecture. It was also among the first in China's financial industry to initiate the construction of a domestic super-node intelligent computing cluster. On model capability development, the bank adopts a dual-track strategy combining external acquisition and in-house innovation – actively integrating advanced industry-wide large models while developing proprietary, enterprise-grade large models. This approach enables the gradual establishment of a comprehensive AI model system covering multiple domains and capabilities. Through a systematic model development roadmap, the bank has built a mature AI capability framework, achieving full lifecycle management of models – from introduction to retirement – supporting the rapid incubation and iteration of over 200 large-model application scenarios.

- Some of banks are advancing the deep integration of AI with financial services, continuously driving digital and intelligent transformation and rapidly evolving toward L4 maturity. For example, Its AI technology platform is built on an enterprise-grade large model platform, applying AI across multiple business scenarios such as intelligent coding and intelligent customer service. The platform is further evolving into a support system capable of developing and running intelligent agents, enabling the creation and operation of sophisticated agents that can understand customer intent and assist with complex product operations. In computing infrastructure, the bank has established a centralized computing pool, progressing toward an intelligent cloud architecture with elastic scheduling and cost optimization, providing on-demand computing resources for intelligent agents and large-model applications to support high-concurrency, low-latency AI services. At the same time, by building a hybrid model matrix that combines large and small models, the bank enables model ensemble applications in key areas such as trade banking and capital management, significantly enhancing the intelligence level of business processing. The bank also emphasizes a continuous improvement mechanism based on business feedback – using real-world operational data generated by intelligent agents to refine and optimize models, forming a closed-loop enhancement path from business to technology. Intelligent agents not only deliver seamless, end-to-end services to customers but also integrate deeply into core business areas such as capital management, driving business model innovation.

In general, the evolution from L3 to L4 is essentially a transformation of the technology architecture from supporting processes to business intelligence. Compared with AI-native banks, traditional banks have already demonstrated typical L4 characteristics in their digital transformation journeys, but they are still in the stage of deepening L3 while exploring L4.

4 Implementation roadmap – Precision navigation based on maturity

4.1 General principles: Value leadership, ecosystem co-construction, governance escort

Successful transformation requires adhering to three principles:

1. **Taking clear business value (revenue increase, cost reduction, risk control) as the starting point and measurement criterion**
2. Making up for capability shortcomings through open **ecological cooperation** to avoid working in isolation
3. Integrating **trusted governance** throughout as a safety belt for innovation

4.2 Advancement roadmap: Key tasks from L2 to L4

For most institutions building AI banks that are in L2 and moving toward L3/L4, progress needs to be made around the dual main lines of data and model ecosystem upgrading + business scenario implementation. The core tasks are divided into two stages as follows:

Scenario application period

Process reconstruction period

Business integration period

First stage: L2 → L3 (Process reconstruction period, 6-12 months)

1. Data side: Build an enterprise-level data middle platform to realize data assetization

- Sort out data sources for the entire customer management chain (internal transactions, external cooperation, real-time behaviour data) and formulate unified data governance specifications covering data cleaning, label system and permission control.
- Access real-time data streams such as the Internet of Things inputs and government service platforms, and build a 360° customer portrait data asset library for customer management scenarios to support real-time data retrieval and value quantification.

2. Model side: Build a multi-model collaborative architecture to implement end-to-end intelligent processes

- Select high-value processes in customer management (such as wealth management marketing and consumer loan recommendation) and build a multi-model collaborative system of "portrait modelling + product matching + reach decision-making."
- Introduce automated feature engineering tools, establish model monitoring mechanisms and enable real-time early warning of model drift to ensure decision accuracy.

3. Organization side: Establish a cross-departmental CoE (Center of Excellence) to break business barriers

- Form a cross-departmental team of "business + technology + data" responsible for the AI reconstruction and implementation of customer management processes, ensuring data and model construction are not divorced from business needs.
- Establish an agile working mechanism and promote process optimization following the rhythm of "pilot - verification – promotion." Synchronize business feedback weekly and iterate models accordingly.

Second stage: L3 → L4 (Business integration period, 12-24 months)

1. Data side: Upgrade the data brain to realize the external output of data value

- Develop data mining algorithms based on customer management data assets to independently identify customers' potential needs (such as study abroad and pension financial needs), enabling proactive insight discovery.
- Promote data asset pricing and output desensitized customer behavior data to cooperative ecological partners (such as education and cultural tourism platforms) and obtain ecosystem-level value returns.

2. Model side: Build an MLOps platform to establish a primary model factory

- Realize full-process automation of "model development – training – deployment – iteration," supporting the integrated application of large models and traditional machine learning models.
- Design AI-native products for customer management scenarios (such as dynamically priced consumer loans and intelligent investment advisory portfolios) and enable the model factory to independently complete the development and iteration of product models.

3. Business side: Promote organizational reconstruction and implement the "business + agent" integration model

- Break the traditional hierarchical structure and form an integrated team of "customer management agents + front-end business teams," with agents undertaking data analysis and decision-making suggestions and business teams focusing on customer services.
- Establish an assessment system for AI-native businesses, incorporating ecosystem revenue and customer LTV improvement into core indicators to drive business model upgrading.

4.3 Special recommendations: Agile breakthrough path for small and medium-sized banks

I. Market overview: The AI advantage of small and medium-sized banks comes from focus rather than scale

From the practices of North American community banks, European specialized banks and Asian regional banks, the most common and effective successful paths of AI in small and medium-sized banks share the following characteristics:

- They do not pursue the construction of a bank-wide, unified AI platform.
- They clearly focus on one to two high-value business scenarios.
- They compensate for data, models and technical capabilities through external ecosystems.
- They form local leadership within segmented customer groups or regional markets.

This is fundamentally different from the approach of large banks that promote comprehensive intelligence relying on scale advantages.

II. Path selection: small and medium-sized banks should not copy the AI-bank model of large banks

The AI-bank construction model of large banks typically aims for a "unified data base + central model factory + full-process automation," which depends on long-term high investment, strong governance and strong technical organizational capabilities.

If regional banks simply adopt this model, they often encounter three major challenges:

- **Mismatch between investment intensity and revenue structure**
- **Technical complexity that exceeds organizational carrying capacity**
- **Amplified AI risks and compliance costs**

EY recommends that the regional banks prioritize serving clear business goals rather than pursuing technical completeness when building AI-banks capabilities.

III. Path recommendations: Strategic focus + ecological empowerment

Based on global practices and aligned with the requirements of China's 15th Five-Year Plan and related financial regulatory policies, EY recommends that small and medium-sized banks adopt a path of strategic intelligence, with three core principles:

Strategic focus: Clarify that the primary goal of AI is not to "improve the efficiency of the entire bank," but to strengthen the core business capabilities with the strongest regional or customer-group advantages, such as:

- Micro, small and medium-sized enterprise (MSME) and inclusive finance
- Sci-tech innovation and industrial chain finance
- Agriculture, county and characteristic industrial finance

Scenario-driven: Build relatively independent and quickly implementable AI capability units around a single core scenario, rather than transforming the entire bank's systems all at once.

Ecological collaboration: Proactively deepen cooperation with local government platforms, local industrial platforms and third-party technology companies with strong credit and technical capabilities. Through capability combination rather than independent R&D, small and medium-size banks can shorten construction cycles and reduce trial-and-error costs.

IV. Value positioning: AI as a capability amplifier rather than a cost center

Successful experiences from global regional banks show that the optimal role of AI in small and medium-sized banks is not to replace human resources or achieve full automation, but to:

- Significantly improve risk identification and pricing capabilities
- Enhance service capabilities for complex, small-scale customers
- Reduce reliance on individual experience and realize standardized, replicable capabilities

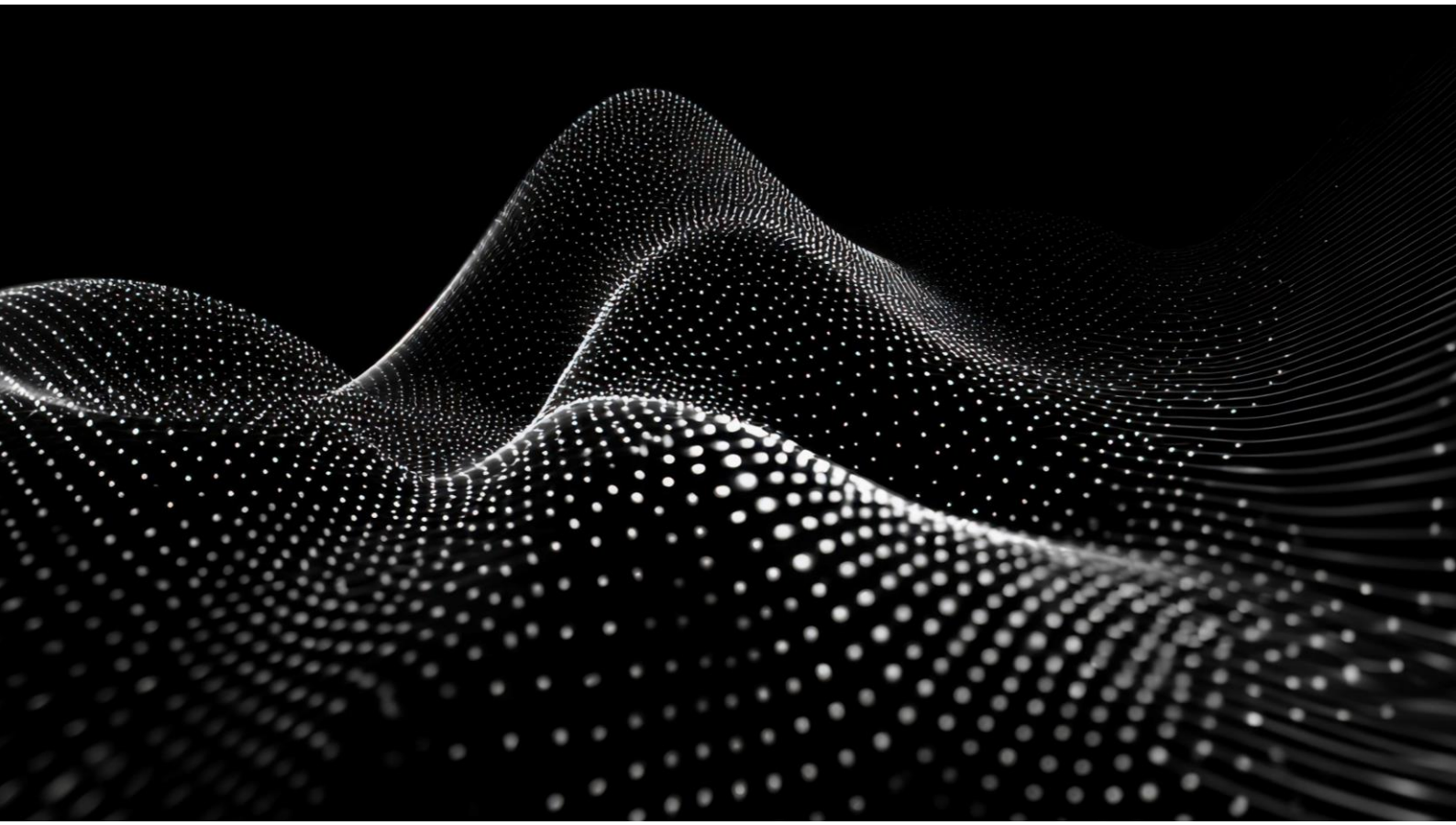
When **AI is used to amplify existing business advantages** rather than to create new business forms, its input-output ratio and sustainability are significantly improved.

V. Trend prediction: Evolution direction of AI-banks for small and medium-sized banks

EY predicts that in the next 3-5 years, the AI-bank models of regional banks will evolve along the following directions:

- From technology-oriented to strategy and value-oriented
- From platform-based construction to modular and scenario-based capability accumulation
- Gradually expanding application boundaries under strict frameworks of interpretability, auditability and regulatory compliance

For regional banks, becoming an AI-bank does not mean achieving full intelligence. Instead, it means being smart enough and continuously leading in key business areas.



4.4 Organization and talent: An indispensable engine for change

Talent strategy: Implement the “golden seed" program

Through a **“learning – training – practice"** model, banks can quickly cultivate a group of interdisciplinary talent who understand business, data and AI.

Learning

Cognitive reconstruction and general knowledge popularization: Establish a hierarchical AI general knowledge curriculum system. For senior management, focus on AI leadership training to strengthen their understanding of technical boundaries and ethical responsibilities; for business backbones (the “golden seed" reserve team), conduct training on prompt engineering and basic data governance to help them master the new ways of interacting with AI agents and eliminate fears of black-box technologies.

Training

Scene simulation: Build a financial AI training sandbox. Different from traditional theoretical training, create a simulated business environment based on desensitized data. Trainees need to complete the entire process from data cleaning and feature engineering to model optimization in the sandbox. Through scenario-based learning, reduce trial-and-error costs, cultivate the ability to translate business pain points into technical requirements, and incubate the first batch of "citizen developers".

Practice

Practical delivery and two-way secondment: Incorporate the “golden seeds" into the bank's strategic key projects. Implement a two-line assessment system in practice: business personnel act as AI product managers and are responsible for the realized value of models; technical personnel go deep into network or approval frontlines to gain insight into the pain points of the last mile and are responsible for the final business conversion rate. Through deeply coupled practical experience, enable the “golden seeds" to complete their ability transformation by solving specific problems.

Organizational transformation

Gradually evolve from the traditional hierarchical system to an agile network structure of **“outposts (business teams) + hub (capability platform)".**

Build an intensive enterprise-level capability middle platform:

To avoid redundant construction by outposts, banks need to integrate scattered data assets, algorithm models and computing power resources to build a unified AI middle platform, providing capability assets packaged for front-end use. Establish an AI ethics committee focused on model risk governance, responsible for building a bank-wide AI governance framework, coordinating the full lifecycle risk monitoring and compliance early warning of models, and ensuring that technological innovation operates within the regulatory compliance system.

Deploy all-round agile teams:

Break departmental barriers and form cross-functional agile teams that include business personnel, data analysts, algorithm engineers and other roles in core business lines such as retail finance, corporate finance and financial markets, and give them sufficient operational decision-making power. In the L4 stage, AI agents can be introduced as team members to handle high-frequency and standardized tasks.

Establish a market-oriented internal collaboration mechanism:

Build a settlement mechanism based on internal transfer pricing between the middle platform and the front office. The front office purchases computing power and model services from the middle platform by creating actual business value (such as AUM growth and lower non-performing rates), quantifies technical value through input-output ratios, and drives the organization to shift from administrative, order-driven to value-contract-driven operations, forming a virtuous closed loop of demand-driven supply and supply-empowered demand.

4.5 Governance framework: Establish a dynamically evolving risk governance model aligned with maturity levels to ensure controlled innovation

The development of AI in banks requires a dynamically evolving governance model that matches their technical and business maturity. By deeply integrating governance requirements into the entire lifecycle of AI, banks can achieve the transformation from passive compliance to active empowerment. While fully releasing technical potential, they must also systematically manage the complexity, uncertainty and new risks associated with innovation, ensuring that AI always operates on a safe, reliable and trusted track.

Firstly, the dynamic risk governance model needs to match AI maturity. AI systems at different maturity stages have significant differences in technical stability, application complexity and organizational governance capabilities, and their capacity to absorb risk control also varies. The governance model should take maturity as the core basis for configuring governance intensity and delegation boundaries: emphasize risk isolation and manual control in the early stage and gradually introduce automated governance and flexible constraints in the high-maturity stage, avoiding hindering innovation due to uniform governance.

Secondly, the key to realizing the parallel development of governance and innovation lies in building a governance mechanism that links maturity and risk levels. On the basis of maturity classification, banks should further distinguish the risk levels of different AI application scenarios and accurately match governance measures with risk assessment results. For low-risk and high-maturity applications, governance can rely more on operational monitoring and post-event correction; for high-risk or low-maturity applications, it is necessary to strengthen pre-event review, manual intervention and usage restrictions. Through the linkage configuration of "maturity – risk – control intensity", banks can realize the precise allocation of governance resources.

In addition, the effective operation of the dynamic risk governance model relies on the continuous embedding and feedback adjustment of governance mechanisms in the entire lifecycle of AI. Risk governance should cover the entire process of design, development, deployment, operation and change, and dynamically identify model behavior deviations and risk accumulation through operational monitoring and indicator feedback. On this basis, governance results should feed back into maturity assessment and governance intensity configuration, enabling delegation and constraints to be dynamically adjusted with governance performance. By combining lifecycle governance with feedback mechanisms, risk governance is no longer a one-time institutional arrangement but an endogenous mechanism that continuously optimizes with the evolution of AI capabilities.

In summary, the dynamic risk governance model matching the maturity of bank AI models is based on maturity classification, centered on risk stratification and supported by full lifecycle management and continuous feedback. By dynamically adjusting governance intensity, this model provides a stable and flexible institutional environment for the innovative application of bank AI, ensuring controllable risks and clear responsibilities, and realizing the long-term balance between innovative development and sound governance.

5 Conclusion – Toward a people-centered intelligent financial future

5.1 Core conclusion: The essence of competition lies in ecosystem and trust

The ultimate battlefield of AI bank competition is not the simple stacking of technologies, but the ability to build a new financial ecosystem based on trust, centered on intelligent agents and supported by open platforms. The winners in the future will be those institutions that can most effectively convert technical potential into customer trust, ecological collaboration and sustainable business value.

5.2 Future outlook: China's contributions and global insights

The value of China's exploration in AI finance lies not only in the speed of application, but also in providing a practical model for balancing innovation and safety amid rapid development. China's huge real-economy scenarios, unique digital ecosystem and rich data resources are expected to nurture an intelligent financial model distinct from Western pathways, contributing to greater global financial inclusiveness and resilience.

5.3 EY's commitment: driving responsible innovation with you

As the financial industry enters a new era shaped by AI, EY brings the full strength of the globally unified EY.ai platform together with deep experience across audit, consulting, tax, strategy and transactions to help clients accelerate AI transformation with confidence. We go beyond assessing technology to guiding its practical, scalable and responsible deployment so it delivers measurable business outcomes. Our shared goal: embrace AI responsibly and build a more effective, inclusive and resilient financial system.

Contact us

Effie Xin

Chief Operating Officer, Greater China
Managing Partner, Financial Services, Greater China
Managing Partner, FinTech & Innovation, Asia Pacific
Ernst & Young (China) Advisory Limited
effie.xin@cn.ey.com

Jason Zhang

Partner, Financial Services Mainland China
Consulting Leader, Greater China
Ernst & Young (China) Advisory Limited
jason.zhang@cn.ey.com

Keith Yuen

Partner, Financial Services Technology
Consulting Leader, Greater China
Ernst & Young (China) Advisory Limited
keith.yuen@cn.ey.com

Bing Kun Wang

Partner, Financial Services Consulting
Ernst & Young (China) Advisory Limited
bing-kun.wang@cn.ey.com

Yuan Yuan

Partner, Financial Services Consulting
Ernst & Young (China) Advisory Limited
yuan-yy.yuan@cn.ey.com

Floyd Qian

Partner, Financial Services Consulting
Ernst & Young (China) Advisory Limited
floyd.qian@cn.ey.com

Bruce Wang

Partner, Financial Services Consulting
Ernst & Young (China) Advisory Limited
bruce.zh.wang@cn.ey.com

Kenneth Cen

Partner, Financial Services Consulting
Ernst & Young (China) Advisory Limited
kenneth.cen@cn.ey.com

Cannie Xu

Senior Manager, Financial Services Consulting
Ernst & Young (China) Advisory Limited
cannie.xu@cn.ey.com

Ashley Xu

Senior Manager, Financial Services Consulting
Ernst & Young (China) Advisory Limited
ashley-bq.xu@cn.ey.com

EY | Reference

1. 《IIF-EY Annual Survey Report on AI Use in Financial Services》 https://www.iif.com/Publications/ID/6322/2025-IIF-EY-Annual-Survey-Report-on-AI-Use-in-Financial-Services
2. IDC. (2025年6月).《中国金融行业生成式AI平台及应用解决方案市场份额，2024：风云初启》报告
3. 上海银行. (2025年7月29日). 上海银行发布数字化转型最新成果. 取自 https://www.bosc.cn/m/home/shnews/xwkd/202507/t20250729_264940.shtml
4. 央广网. (2025年7月31日). 人工智能大会：金融壹账通首席架构师吴磊发表演讲. 取自 https://www.cnr.cn/jrpd/mxhq/20250731/t20250731_527292968.shtml
5. 中国银行保险报网. (2025年9月1日). 江苏银行：双轮驱动助中小银行数字化转型. 取自 http://www.cbimc.cn/content/2025-09/01/content_554214.html
6. 《金融电子化》. (2024年11月上半月). “方舟计划”智绘数字金融新篇章. 取自 https://k.sina.com.cn/article_7072825535_1a592c0bf001018nis.html
7. 建设银行. (2025年). 建设银行2025年半年度报告. 取自 https://ccb.cn/chn/attachDir/2025/09/2025092617063696179.pdf
8. 经济参考报. (2025年9月28日). 网商银行上线AI农事助手. 取自 http://jjckb.xinhuanet.com/20250928/23c12741c91a4a02be4b7c9455587883/c.html

All in，聚信心，塑未来。

EY | Building a better working world

EY is building a better working world by creating new value for clients, people, society and the planet, while building trust in capital markets.

Enabled by data, AI and advanced technology, EY teams help clients shape the future with confidence and develop answers for the most pressing issues of today and tomorrow.

EY teams work across a full spectrum of services in assurance, consulting, tax, strategy and transactions. Fueled by sector insights, a globally connected, multidisciplinary network and diverse ecosystem partners, EY teams can provide services in more than 150 countries and territories.

All in to shape the future with confidence.

APAC no. 03025081
ED None

ey.com/china

Follow us on WeChat
Scan the QR code and stay up to date with the latest EY news.