Charting the Future:
How Digital, Intelligent & Green Tech are Reshaping Chinese Banking

Chen Sichong

School of Finance, Zhongnan University of Economics and Law

July 8th, 2025

陈思翀

博 文 明 理 · 厚 德 济 世

Instructor's CV

- **Chen Sichong**, Professor, School of Finance, Zhongnan University of Economics and Law; Director of the Finance Department; Executive Head, Virtual Teaching and Research Section for Banking Management Courses of Ministry of Education of China; Executive Director, ESG Research Institute.

- Ph.D. in Finance; Assistant Professor, Graduate School of Commerce and Management, Hitotsubashi University; Visiting Professor, Warrington College of Business, University of Florida; Deputy General Manager of the Personal Credit and Housing Finance Department, China Construction Bank (Hubei).

- Current Research Focus: Banking and financial markets, emerging risk management in International Finance.

陈 思翀（Chen Sichong）

· Email: sichongchen@zuel.edu.cn
· Homepage: https://finance.zuel.edu.cn/2020/0804/c2003a248224/page.htm

数字技术与
现代银行业导论

AN INTRODUCTION TO
DIGITAL TECHNOLOGY
AND MODERN BANKING

本书具有以下特点：

◎ **全面系统**。对现代银行业经营中数字技术的底层逻辑及其应用场景进行了系统梳理。不仅阐述了数字技术的原理、特性及其应用，还从经济学理论角度揭示出数字技术如何重塑银行业的规模经济和范围经济。此外，书中探讨了数字技术在银行业务的多个关键领域—包括营销、授信和风控等—的全过程应用，并提供了丰富的银行数字化转型实践案例。

◎ **紧跟实践变革**。银行数字化转型中，"人"尤为重要。本书紧跟大数据＋人工智能等数字技术给现代银行业带来的深刻变革，聚焦商业银行数字化经营中复合型人才的培育，旨在培养学生根据银行数字化经营的业务场景，匹配和应用"数字素养"。

◎ **理论与实践并重**。几位作者深耕相关学术领域多年，其深厚的研究功底为分析提供了坚实的理论支撑。同时，他们都具有国有大行的挂职工作经历，并多次开展银行内部培训，对现代银行业数字化经营场景进行了深入的观察和思考，确保本书内容与银行业实践紧密结合。

数字技术与现代银行业导论

Scan it to buy
in JD.COM

长按或扫描查看

"北京大学经管书苑"
微信公众号

"北京大学出版社"
微信公众号

上架建议：数字金融

定价： .00 元

数字技术与现代银行业导论

陈思翀　白小滢　董志华　主编

AN INTRODUCTION TO
DIGITAL TECHNOLOGY
AND MODERN BANKING

数字技术与
现代银行业导论

陈思翀　白小滢　董志华　主编

北京大学出版社
PEKING UNIVERSITY PRESS

作者简介

陈思翀，男，中南财经政法大学金融学院教授、博士生导师、文澜青年学者，日本一桥大学商学金融博士；曾任日本一桥大学商学院讲师、美国佛罗里达大学惠灵顿商学院访问教授，挂职中国建设银行湖北省分行住房金融与个人信贷部副总经理；主持多项国家自然科学基金、教育部及国际合作研究项目；研究成果发表于国内外权威期刊；出版两部专著、一部译著；获湖北省社会科学优秀成果奖等荣誉。

白小滢，女，经济学博士，中南财经政法大学金融学院副教授，曾任美国圣路易斯华盛顿大学访问学者。曾挂职中国建设银行湖北省分行公司业务部副总经理，兼任国家电网、中交集团、中铁建集团等多家央企国企的财务咨询顾问，主持国家自然科学基金、国家社会科学基金、教育部人文社科基金等。出版专著 2 部，研究成果发表于国内权威期刊，获得财政部一等奖、国企创新二等奖等荣誉。

董志华，男，中南财经政法大学金融学院副教授、硕士生导师，路易斯安那大学拉斐特分校数学专业博士；曾就职于美国知名咨询公司任商业数据分析师，挂职于中国建设银行湖北省分行任金融科技部副总经理；主持或参与多项国家级与省部级研究项目，发表多篇论文；曾获全国青教赛二等奖、湖北省青教赛一等奖，湖北省青年教学能手等荣誉。

An Introduction to Digital Technology and Mordern Banking

1. Comprehensive and Systematic Analysis

It systematically examines the underlying logic, technical principles and application scenarios of digital technologies in modern banking.

It reveals how digitalization reshapes banking economies of scale and scope from an economic theory perspective.

2. Multi-Domain Coverage

Covers digital applications in core banking functions (marketing, credit granting, risk control) with real-world cases.

Explores innovative practices in rural revitalization, ESG, green/low-carbon development, and modernizing social governance.

3. Focus on Human-Centric Transformation

Addresses profound impacts of digital technologies (e.g., big data + AI) on banking.

Emphasizing cultivation of interdisciplinary talent to trains students to apply digital literacy within business contexts of bank digitalization.

4. High Practical Utility

Serves as both an accessible textbook for university economics and finance programs and an essential training reference resource for banking professionals navigating digital transformation.

博文明理 · 厚德济世

ESP

碳金融和碳交易系列丛书

气候变化
与可持续发展银行
CLIMATE CHANGE AND
SUSTAINABLE BANKING

庄子罐 陈思翀 编著

气候变化与可持续发展银行

（书脊）气候变化与可持续发展银行　庄子罐 陈思翀 编著

**气候变化
与可持续发展银行**
CLIMATE CHANGE AND
SUSTAINABLE BANKING

庄子罐

中南财经政法大学金融学院教授、博士生导师，中南财经政法大学碳交易与碳金融研究中心主任；武汉大学金融学博士、北京大学博士后；主持国家社科基金、国家自然科学基金、教育部人文社科基金等项目，参与国家社会科学基金重大项目和教育部哲学社会科学研究重大课题攻关项目。在宏观经济政策、碳市场与碳金融、低碳转型与绿色发展等领域积累了丰富的学术研究、政策咨询和项目研究经验。

陈思翀

中南财经政法大学金融学院教授，博士生导师，中南财经政法大学碳交易与碳金融研究中心副主任；日本一桥大学商学金融博士；曾任日本一桥大学商学院讲师、美国佛罗里达大学惠灵顿商学院访问教授；挂职中国建设银行湖北省分行住房金融与个人信贷部副总经理；主要研究领域为国际金融、资产定价、金融机构、金融领域的政治经济学以及碳金融。主持国家自然科学基金、国际合作科研基金、教育部留学回国基金等项目；主要成果发表于中英文权威期刊以及《财经》、《经济日报》、《证券日报》等大众传媒

Scan it to buy
in JD.COM

京东

长按或扫描查看

My working experience in CCB

CCB's Production Park near ZUEL across the south lake

(Direction from Wentai Building to the Park)

博文明理 · 厚德济世
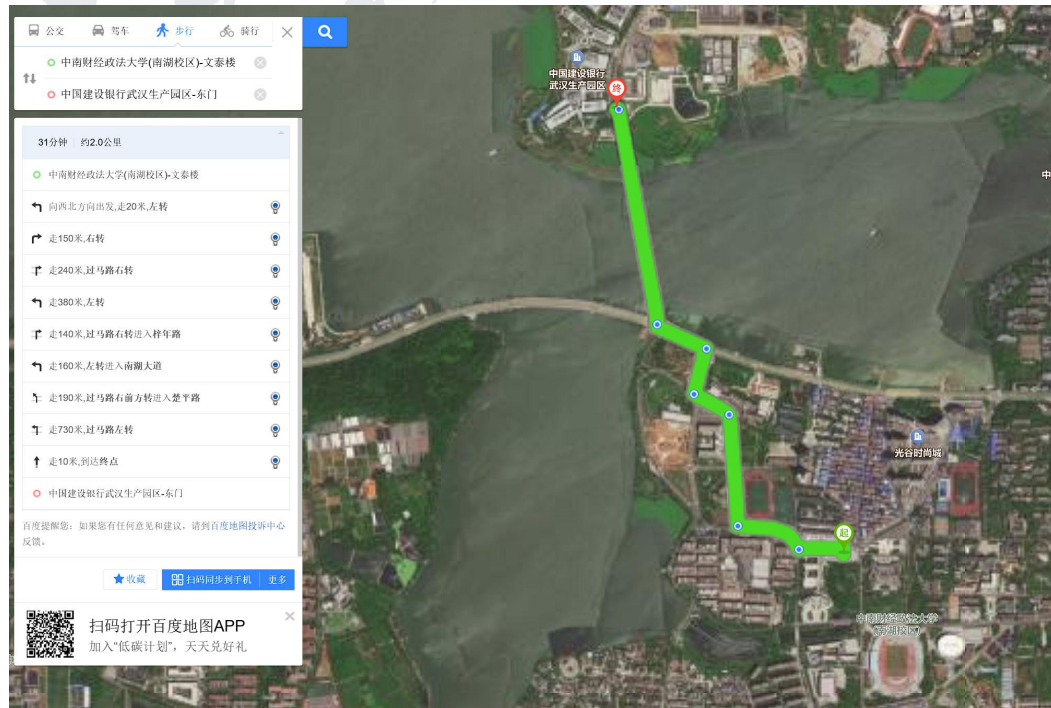
Contents：

1. Understanding the Era's Characteristics and the Wealth Code
Digital, Intelligent, and Green Transformation

2. New Paradigm, New Tools, and New Framework
I. New Paradigm: Digital Technologies & Unstructured Data
II. New Tool: Fine-Tuned Large Financial Models & AI Agents
III. New Framework: Green Finance and Sustainabable Development

Part II

New tools: Fine-tuning & RAG of large models and AI agents

陈思翀

博 文 明 理 · 厚 德 济 世

Banks's AI Revolution

Financial Large Models (FLMs) Initiative

Key Enablers

Cross-functional team leveraging strengths in:

- **Talent**
  Domain experts & AI specialists

- **Computing Power**
  Cloud infrastructure & GPU clusters

- **Algorithms**
  Cutting-edge AI research

- **Data Assets**
  Secure financial datasets

- **Business Scenarios**
 Real-world banking applications

Strategic Goals

Cross-functional synergy driving AI transformation

1. **Improve customer experience**
   Personalized services & 24/7 AI support

2. **Empower employees**
   AI-assisted decision making tools

3. **Reduce workload**
   Automation of routine operations

博 文 明 理 · 厚 德 济 世

FLMs Development Strategy

From Open-Source LLMs to Financial Intelligence
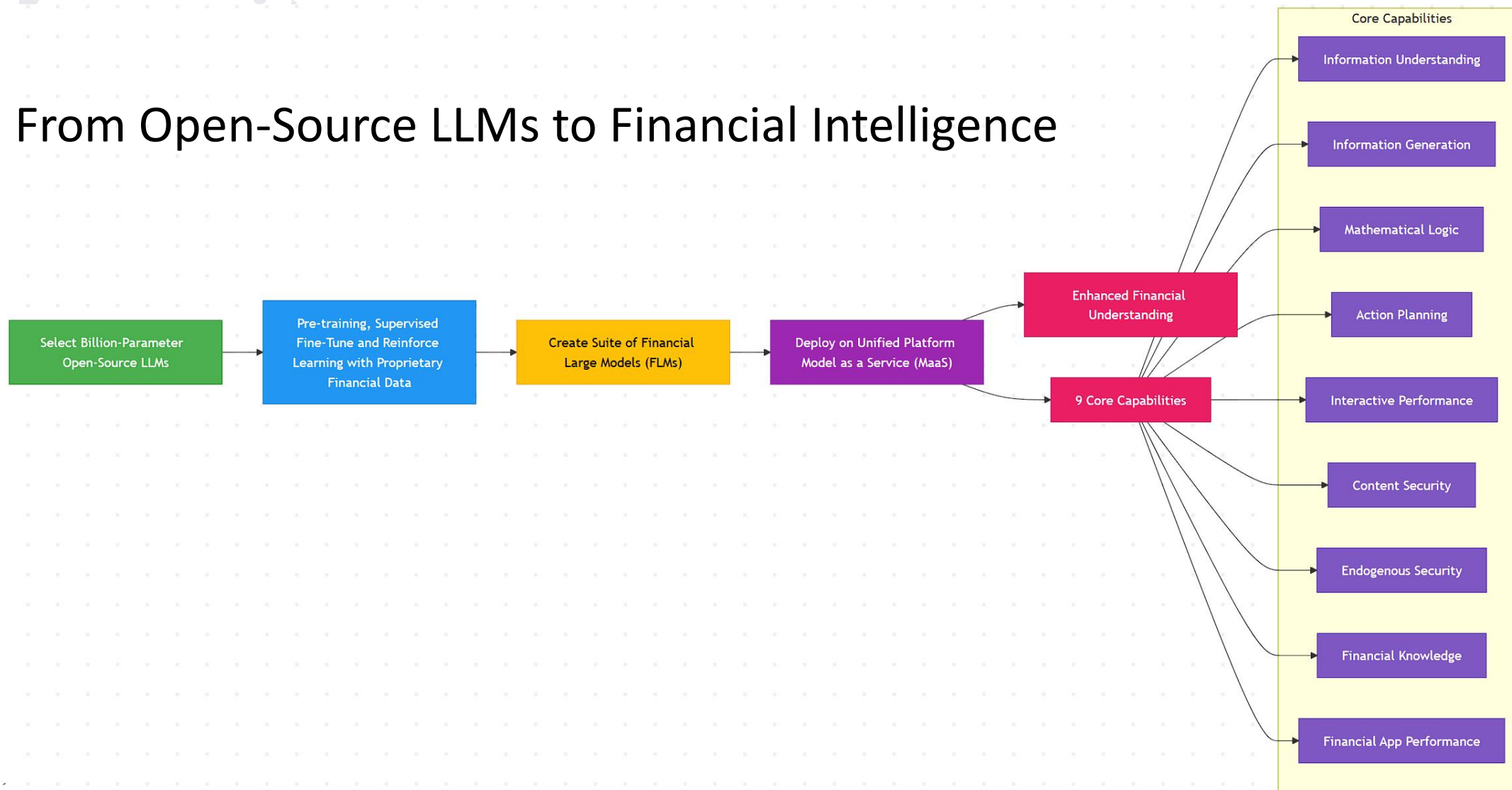
Spicific high-impact use cases

Accelerating business transformation with FLMs

For example, China Construction Bank by the end of 2024:
✅ 16 major FLM updates deployed ✅ 193 business scenarios enabled ✅ 50% workforce adoption

Core Business Impact

💳 **Credit Approval**

- Financial analysis time reduction: **Hours → Minutes**
- Risk assessment accuracy ↑35%
- Full-process automation

🤖 **Account Management**

- Campaign success rate **+22%** with AI-driven targeting
- Client retention **+18%** via personalized engagement
- Real-time portfolio optimization

Enabling Technologies

ChatCCB   Enterprise Chatbot   AI Toolbox   Applet Editor

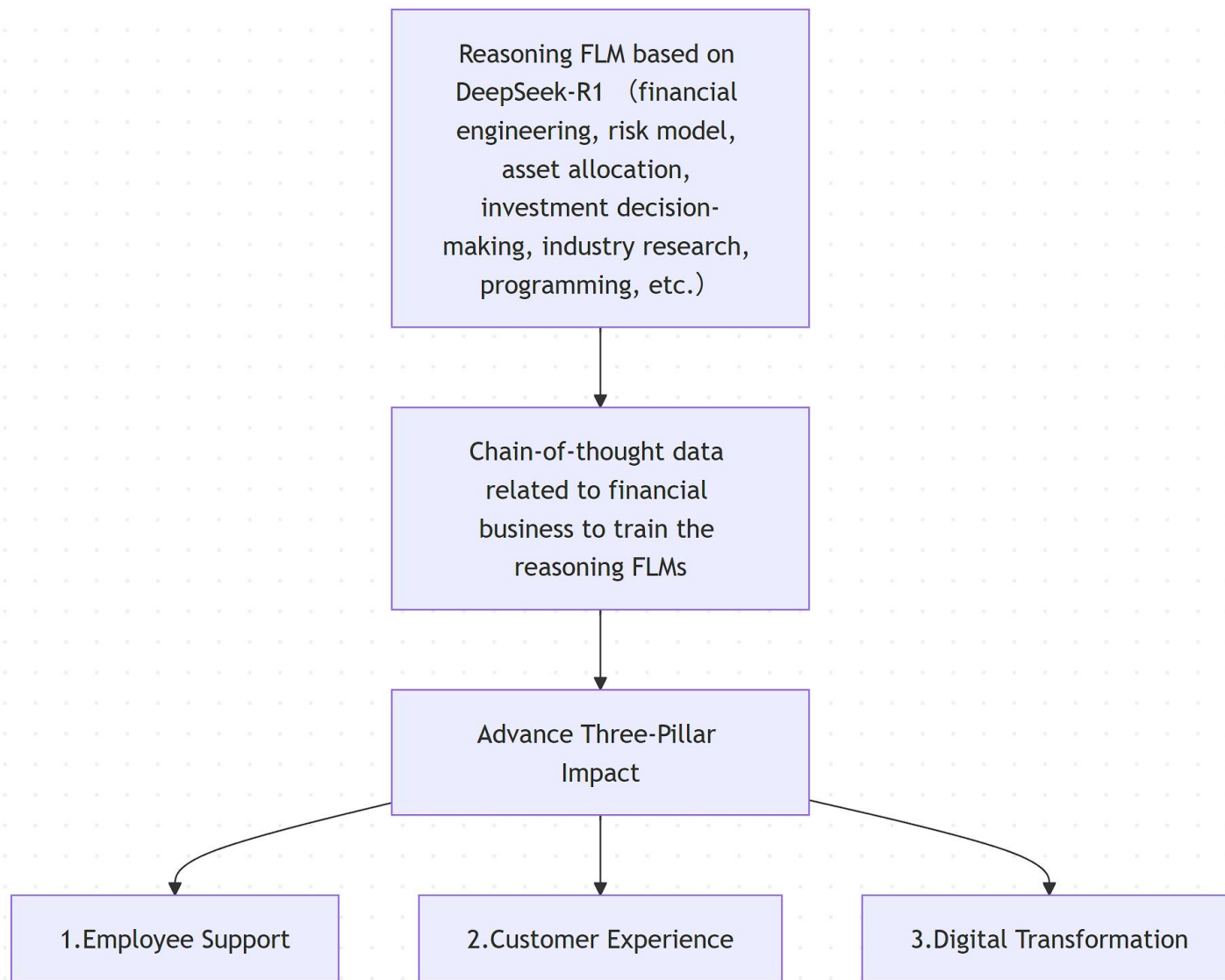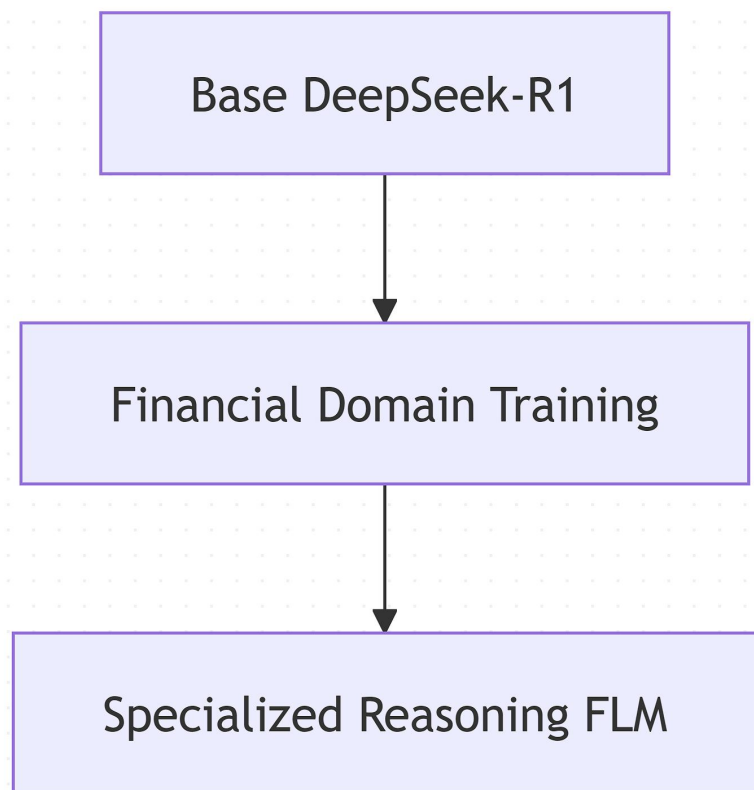Code Interpreter   Data Analyst   Vector KB   Knowledge Engine

Operational Efficiency

- 80% reduction in manual data processing
- Cross-departmental knowledge sharing
- Compliance automation

博 文 明 理 · 厚 德 济 世

Domain-Specific Reasoning Engine (Custom Fine-Tuning)

DeepSeek-R1 Specialization

```
┌─────────────────────────┐
│    Base DeepSeek-R1     │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│ Financial Domain Training│
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│ Specialized Reasoning FLM│
└─────────────────────────┘
```

```
┌──────────────────────────────┐
│  Reasoning FLM based on       │
│  DeepSeek-R1 （financial      │
│  engineering, risk model,     │
│  asset allocation,            │
│  investment decision-         │
│  making, industry research,   │
│  programming, etc.）          │
└──────────────────────────────┘
            │
            ▼
┌──────────────────────────────┐
│  Chain-of-thought data        │
│  related to financial         │
│  business to train the        │
│  reasoning FLMs               │
└──────────────────────────────┘
            │
            ▼
┌──────────────────────────────┐
│  Advance Three-Pillar         │
│  Impact                       │
└──────────────────────────────┘
      │          │          │
      ▼          ▼          ▼
┌──────────┐ ┌──────────┐ ┌──────────┐
│1.Employee│ │2.Customer│ │3.Digital │
│ Support  │ │Experience│ │Transform.│
└──────────┘ └──────────┘ └──────────┘
```

博 文 明 理 · 厚 德 济 世

Weakness of large models and the RAG solutions

1. Solve the inherent defects of large models

(1) Hallucinations : Large models (LLM) generate text based on probability, which may output content that seems reasonable but lacks factual basis. RAG enhances the accuracy of generated content by retrieving external knowledge bases (such as vector databases and knowledge graphs) to supplement real-time data.

(2) Lack of knowledge timeliness : The training data of large models has time limit (for example, ChatGPT-3.5 only covers data up to 2021). RAG solves the problem of model knowledge lag by dynamically updating the external knowledge base and integrating the latest industry trends, policies and regulations.

(3) Insufficient coverage of fields : General large models have limited coverage of professional knowledge in vertical fields (such as finance and law), and RAG can be attached to the enterprise private knowledge base to improve the accuracy of response in specific scenarios.

2. Reduce resource input and security risks

(1) Reduce fine-tuning cost : Traditional fine-tuning requires a large amount of annotated data and computing power, while RAG only needs to update the knowledge base to adapt to new requirements, especially suitable for scenarios with frequent data changes (such as financial market).

(2) Data security guarantee : RAG allows enterprises to store sensitive data (such as customer financial information) locally to avoid direct invocation of public models that lead to data leakage.

3. Improve interpretability and controllability

(1) The answers generated by RAG can be traced back to specific retrieval documents, which is convenient for verifying the credibility of the source and meets the stringent requirements of compliance and transparency in the financial industry.

Compare the applications of with/without RAG responses

Question (Q): What are the current requirements of the Securities Regulatory Commission (SEC) on the deadline for disclosure of major matters by listed companies?

- I. If without RAG:

    General description, such as "within four trading days after a major event".

- 2. If with RAG:

    Citing specific rules;

    Differentiating between different types of submission deadlines;

    Mention the possible consequences;

    Emphasis can be placed on electronic submission.

Conclusion: Activating the RAG process helps ensure that models produce more accurate, factually consistent, and correctly referenced responses, especially in critical financial regulatory scenarios.

Without RAG	With RAG
Listed companies are required to submit Form 8-K reports in a timely manner after the occurrence of major emergencies, usually within four trading days. This ensures that the market keeps abreast of key developments and protects investors right to know.	In accordance with SEC regulations, regular disclosures must be filed within four trading days of the occurrence of a triggering event in Form 8-K. in that case : • Regulation FD disclosure (Section 7.01) Mandatory Same-day filing ; • Earnings revision matters (Article 4.02) The time limit starts from the date when the resolution is passed by the board of directors. Failure to report within the time limit will result in regulatory accountability and loss of simple registration qualification for forms. Listed companies shall establish a complete disclosure and internal control system to ensure the timeliness of declaration, and EDGAR system electronic submission is the mandatory channel. Note: The core provisions are based on SEC final rule 33-8400 (17 CFR Parts 228,229,230,239,240 and 249).

陈思翀

博文明理 · 厚德济世

RAG process

1. Indexing

The **embedding** model quantizes the document library for storage.

2. Search

The **embedding** model quickly recalls similar documents (rough screening).

The **reranker** model gives a detailed score and ranking of the results (fine ranking).

3. Generation

The highly correlated documents after sorting are input into the **generative large language model** to generate answers.

Embedding ensures retrieval speed, and Reranker improves the quality of results. The combination of the two takes into account both efficiency and accuracy.

陈思翀

博 文 明 理 · 厚 德 济 世

Embedding is responsible for the initial screening of candidate texts, and Reranker is responsible for the optimization of the order. The two support the retrieval quality of RAG system.

Take Alibabas Qwen3 as an example: Qwen3-Embedding completes the initial screening of documents, and Qwen3-Reranker optimizes the sorting, which is used in a multilingual RAG system.

	Qwen3-Embedding-8B	Qwen3-Embedding-4B	Qwen3-Embedding-0.6B	Gemini Embedding	Cohere-embed-multilingual-v3.0	text-embedding-3-large	multilingual-e5-large-instruct	gte-Qwen2-7B-instruct
MMTEB Mean-Task	70.58	69.45	64.33	68.37	61.12	58.93	63.22	62.51
MTEB (en v2) Mean-Task	75.22	74.60	70.70	73.30	66.01	66.43	65.53	70.72
MTEB-Code	80.68	80.06	75.41	74.66	51.94	58.95	65.00	56.41

Qwen3-Embedding系列模型

Model Type	Models	Size	Layers	Sequence Length	Embedding Dimension	MRL Support	Instruct Aware
Text Embedding	Qwen3-Embedding-0.6B	0.6B	28	32K	1024	Yes	Yes
	Qwen3-Embedding-4B	4B	36	32K	2560	Yes	Yes
	Qwen3-Embedding-8B	8B	36	32K	4096	Yes	Yes
Text Reranking	Qwen3-Reranker-0.6B	0.6B	28	32K	——	——	Yes
	Qwen3-Reranker-4B	4B	36	32K	——	——	Yes
	Qwen3-Reranker-8B	8B	36	32K	——	——	Yes

Model	Param	MTEB-R	CMTEB-R	MMTEB-R	MLDR	MTEB-Code	FollowIR
Qwen3-Embedding-0.6B	0.6B	61.82	71.02	64.64	50.26	75.41	5.09
Jina-multilingual-reranker-v2-base	0.3B	58.22	63.37	63.73	39.66	58.98	-0.68
gte-multilingual-reranker-base	0.3B	59.51	74.08	59.44	66.33	54.18	-1.64
BGE-reranker-v2-m3	0.6B	57.03	72.16	58.36	59.51	41.38	-0.01
Qwen3-Reranker-0.6B	0.6B	65.80	71.31	66.36	67.28	73.42	5.41
Qwen3-Reranker-4B	4B	69.76	75.94	72.74	69.97	81.20	14.84
Qwen3-Reranker-8B	8B	69.02	77.45	72.94	70.19	81.22	8.05

From Traditional Al Assistants to Al Agents

Al Agent is a kind of intelligent entity with the ability to perceive the environment, make decisions and perform tasks autonomously. It realizes the closed loop from "thinking" to 'acting' by using the big model as the "brain" and interacting with the environment through the call of tools. from "think" to "act" through the use of tools and environment interaction.

AI agents typically contain the following modules

Perception	Knowledge	Decision	Execution	Learning

陈思翀

Cultivate civilization and reason, cultivate virtue and benefit the world

The difference between an AI agent and a traditional AI assistant

Dimensions	Traditional AI assistants (like ChatGPT)	AI agents
Competence scope	Information/advice only	Planning + execution of complete tasks
Interactive mode	The text dialogue only	Multimodal interaction + tool operation
Autonomy	Depends on user instructions	Proactively perceive the environment and make decisions
Cases	Driving-asistance	VLA simulates the cognitive, decision-making and execution process of a human driver

陈思翀
博文明理·厚德济世