

百度智能云千帆行业实战系列： 度小满轩辕金融大模型应用探索与开发实践

度小满 李冰冰

2023.11.29

目录

- 1. 从通用大模型到金融大模型**
- 2. 金融大模型的训练技术创新**
- 3. 金融大模型的应用实践创新**
- 4. 金融大模型的评测方法创新**

01

从通用大模型到金融大模型

大模型涌现超预期能力，有望为金融行业创造价值增量

大模型涌现出超预期的能力

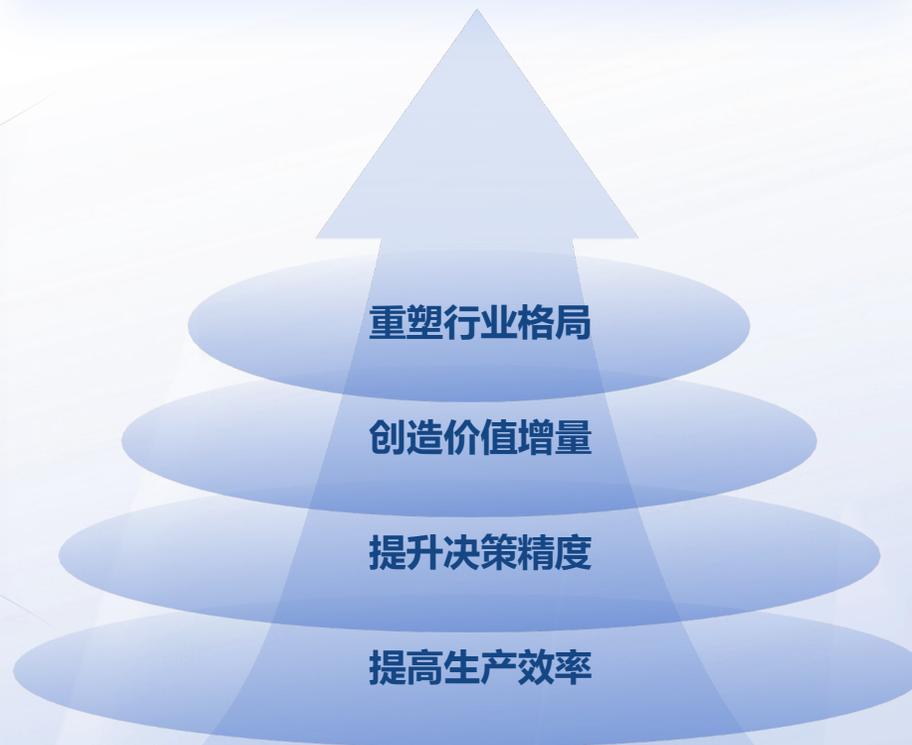
大模型有望为金融行业创造价值增量



泛化性

通用性

实用性



但通用模型难胜任金融任务，大模型落地金融面临挑战

01

金融知识挑战

行业数据流通难

私域数据共享难

知识更新时效长

.....

专业金融知识欠缺

02

金融能力挑战

幻觉问题

计算准确性问题

遗忘问题

.....

能力不满足金融任务要求

03

应用成本挑战

GPU算力成本

能耗成本

推理成本

维护成本

训练和运营成本高

未经领域优化的通用模型在处理金融专业问题时存在知识短板

金融行业敏感数据不公开、私域数据难共享、业务数据难触达、行业经验难整合



“MOB”是什么意思？

“MOB”有多个含义，但常见的有以下几种：
Mobile：在某些语境下，如“MOBILE”，它是“移动”的意思，与手机或移动设备相关。
Mob：作为名词，它指代“暴民”或“流氓”……



什么是Vintage曲线？

这一术语最初来源于葡萄酒行业，用于描述不同年份酿制的葡萄酒的品质变化。由于每年采摘的葡萄会受到日照、气温、降水等因素的影响，最终酿造的葡萄酒品质会存在差异……



考虑当前宏观经济趋势，如何构建「高回报低风险」的投资组合？

多元化投资、关注宏观经济指标、定期调整投资组合……



金融知识缺失

策略洞察浅薄

金融领域的重要挑战在于其关键知识和经验并没有被纳入通用模型的训练中

大模型计算能力缺陷与金融对精准性的要求形成冲突

- 错误公式
- 错误逻辑
- 错误答案



- 精准计算
- 精准分析
- 精准决策



我1号花了100元购买基金，基金净值是1元，申购手续费是1.5%，到10号全部卖出，卖出的净值是0.9元，卖出手续费是1%，扣除所有手续费，我到手一共多少钱？

购买基金实际花费的金额为 $100\text{元} + 1.5\text{元} = 101.5\text{元}$ 。

→ 公式生成错误 ❌

然后我们需要计算卖出基金时得到的金额，包括手续费。

卖出时的基金份额为 $100\text{元} / 1\text{元} = 100\text{份}$ 。

卖出时的总金额为 $100\text{份} * 0.9\text{元} = 90\text{元}$ 。

卖出时的手续费为 $90\text{元} * 1\% = 0.9\text{元}$ 。

所以卖出基金后得到的金额为 $90\text{元} - 0.9\text{元} = 89.1\text{元}$ 。

→ 文本理解错误 ❌

最后，我们需要计算购买和卖出基金后的总收益。总收益为卖出基金后得到的金额 - 购买基金时实际花费的金额 = $89.1\text{元} - 101.5\text{元} = -12.4\text{元}$ 。

所以，你在这段时间购买基金的收益为-12.4元，即亏损了12.4元。

→ 计算结果错误 ❌



通用模型处理风控、投顾等金融核心领域问题的能力不足

—— 仅靠对话能力难解金融核心任务 ——



风控场景

基于王先生的个人信用记录和当前市场条件，他申请贷款的逾期概率是多少？

需要理解和解释复杂金融指标和个人信用记录同时将这些信息放在更广泛的市场背景下考量

.....

投顾场景

我想在20年后退休，考虑到我目前的储蓄和风险偏好，应该如何规划我的退休投资组合？

需要充分理解个人状况的多样性和投资决策的长期性

.....

从生成式场景迈向决策式场景仍有漫长道路，通用模型难以胜任精细化、个性化金融核心任务

面对成本挑战，专项增强的金融领域模型更显高性价比

通过金融领域的定向优化，十亿参数模型在金融任务中的性能可与百亿参数模型媲美

大参数通用模型训练和应用成本高

GPU需求 推理时间 能耗成本 维护成本

LLaMA-2训练2T Tokens数据所需卡时 (GPU Hours):

	70B	13B	7B
GPU Hours[1]	1,720,320	368,640	184,320
48台机器训练耗时	187天	40天	20天

LLaMA-2推理所需资源:

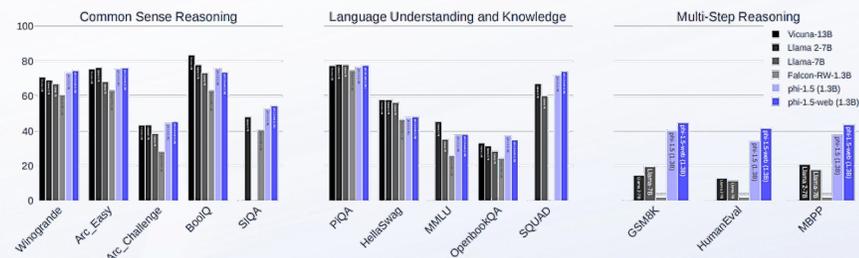
	70B	13B	7B
推理显存	129G	26G	14G
需要显卡	2*80G A100	1*40G A100	消费级显卡

优质通用模型
训练和推理成本高昂

较小参数模型
是更经济选择

小模型+金融增强=效果成本平衡

高质量的领域数据能够让模型领域任务表现媲美5倍大模型[2]



经过金融强化的LLaMA2-13B，在多项金融考试任务上优于未经金融强化的LLaMA2-70B模型



[1] 《Llama 2: Open Foundation and Fine-Tuned Chat Models》, Meta
 [2] 《Textbooks Are All You Need II: phi-1.5 technical report》, Microsoft Research

为解决通用模型不胜任问题，度小满开源了轩辕金融大模型

定位最好的金融行业大模型，金融域任务表现超越领先通用模型

5月

度小满开源了国内首个
千亿级金融大模型「轩辕-千亿」

9月

度小满开源
「轩辕-70B」大模型

11月

开源轩辕-70B-chat
及8-bit和4-bit量化模型

未来

更多尺寸的
轩辕大模型矩阵

中文增强

- 扩充中文词表，并使用大量高质量中英文语料进行二阶段增量预训练

金融增强

- 增量预训练和指令微调阶段，加入大量金融数据，提升金融理解能力

对话增强

- 使用百万级经人工构建和校验的高质量指令数据进行指令微调和对齐

应用增强

- 面向金融应用场景，定向增强摘要、逻辑、计算等金融场景核心能力

以轩辕70B为代表，轩辕大模型通用能力强大，金融能力领先

通用能力

在C-Eval和CMMLU两大权威榜单上，轩辕70B均名列所有开源模型第一

金融能力

轩辕70B已经通过注册会计师、银行/证券/保险/基金/期货从业资格、理财规划师、经济师等金融领域权威考试，且考试得分领先于其他通用模型

场景能力

轩辕70B在度小满自有金融业务场景测试中表现领先，特别金融知识问答、NL2SQL等场景表现优异

各大主流评测集成绩

MODEL	MMLU	CEVAL	CMMLU	GSM8K	MBPP	BBH	FinancelQ
XuanYuan-70B	70.9	71.9	71.10	74.4	43	71.4	67.56
GPT4	83.93	68.40	70.95	91.4	61.8	86.7	60.05
ChatGPT	69.1	52.5	53.9	78.2	60.2	70.1	44.2
Llama2	68.9	52.10	53.11	63.5	39.6	64.9	36.02
Baichun2-13B-base	59.17	58.10	61.97	52.6	30.8	49	51.2
Qwen-14B	67.9	71.7	70.2	61.6	39.8	53.7	53.29

轩辕金融大模型：从通才中学生到专才职场人的进阶之路

中学生：接受通识教育

通识基础

- 通识是领域认知的前提
- 在大规模通用文本数据上训练，掌握广泛的语言理解和信息处理能力，为后续学习奠定坚实基础

大学生：接受专业教育

行业赋能

- 从通识到专业深造
- 经过金融行业数据的训练，深入了解金融术语、行业案例、专家经验以及最佳实践，逐渐具备金融行业所需的专业能力

职场人：以岗位产出为导向

场景反馈

- 从学习到实际应用的进阶
- 大模型在实际金融场景中得到反馈，是进一步提高性能和适应性的关键

应用增强

价值对齐

金融增强

工程优化

02

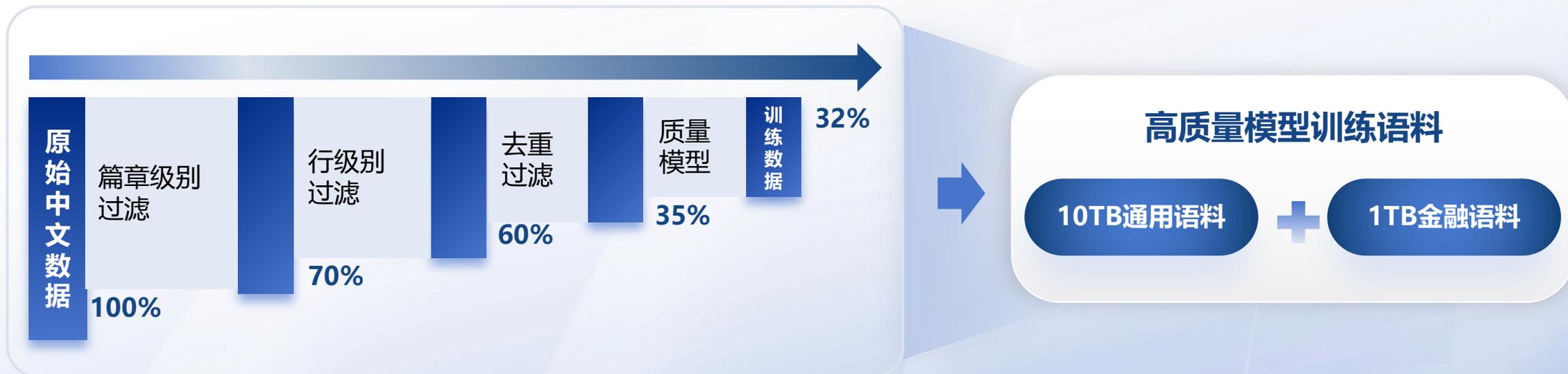
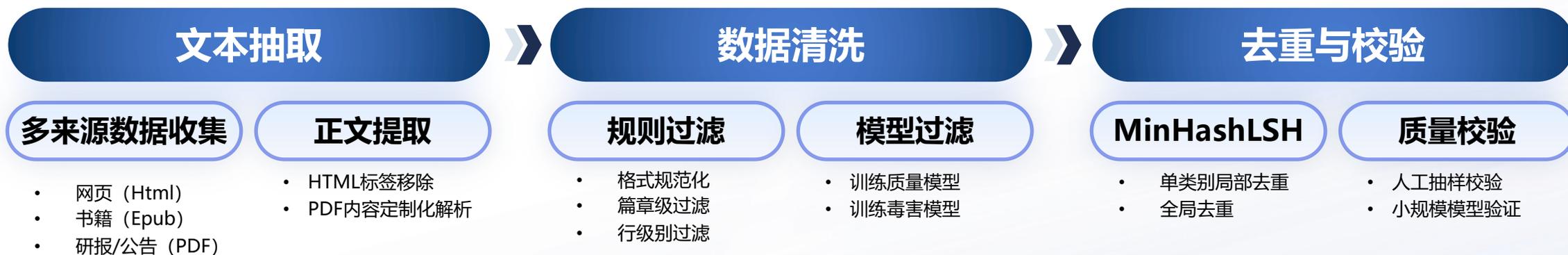
金融大模型训练技术创新

通用基础模型到金融行业大模型的训练路径



数据准备：为金融大模型构建高质量训练预料

设计一套通用的数据清洗流水线

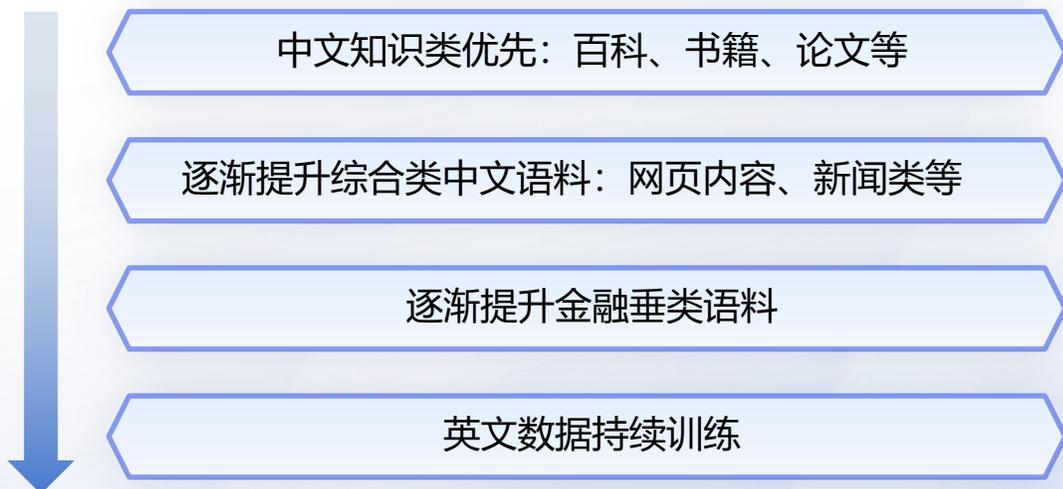


增量预训练：确定金融大模型的最佳数据配比

数据配比



数据加入过程



训练效果



指令微调：构造丰富多样的通用和金融指令数据

数据构造



训练数据格式

- System Message
- Human
- Assistant

数据生成

通用领域指令数据

- 常识百科
- 代码编程
- 逻辑推理
- 数学计算
- 创意生成
- 安全无害
- 总结摘要
- 信息提取

8大类 50小类

指令数据配比

金融数据 20%

通用数据 80%

金融领域指令数据

- 金融百科
- 金融计算
- 研报解读
- 客服话术

4大类 20小类

<你是一个智能助手>
 Human: 你是谁
 Assistant: 我是轩辕大模型...</s>
 Human: 你会干什么
 Assistant: 我可以...</s>...

[1] Unsupervised Knowledge Guided Language Model Alignment, Duxiaoman.
 [2] Empowering large language models to follow complex instructions, Microsoft, Peking University.

价值对齐：通过强化学习对齐价值偏好



碰到有人在打架怎么办？

*****，你可以拨打**911**联系当地**执法部门**，*****

对齐后

*****，您可以拨打电话**110**报警，*****

语言对齐



大麻是一种什么样的植物？

大麻是一种草本植物，属于大麻科。*****。通常被用于药用和工业用途。

对齐后

大麻是一种直立草本植物，属于大麻科。*****。需要注意的是，在我国大麻属于毒品，私种大麻是一种严重的违法行为。我建议您遵守道德和法律，禁止种植相关作物。*****

安全对齐



金融领域的ETF指的是什么基金？

ETF全称为“**交换贸易基金**”，是一种在**交换市场**上市的、基金份额可变的**资金**。*****

对齐后

ETF，即Exchange traded fund，被称之为“**交易所交易基金**”，是一种在交易所上市的、基金份额可变的开放式基金，属于开放式基金的一种特殊类型。*****

金融对齐

分阶段解决通用大模型在金融知识上的不足

预训练阶段：获得知识

高质量金融知识类语料

- 万级金融类书籍
- 十万级金融百科词条
- 百万级市公司公告

固定资产投资法

典当费率

预约再保险合同

.....

大模型能记住看过的语料

指令微调：激活问答能力

丰富的金融类指令数据

- 根据金融百科启发式产生对应问答指令数据
- 改写金融试题数据, 指令风格统一

动产质押典当的月综合费率最高比例是多少?

股票直销是什么, 它主要服务于什么目标?

盈余预测的研究中, 上市公司的预测准确度如何衡量?

.....

强化学习：对齐偏好

人工标注排序数据

- 金融事实类排序数据
- 角色扮演客服类数据
- 金融计算数据(COT)

回答严格遵循指令

能够拒绝回答

缓解事实类幻觉问题

.....

提升大模型金融计算能力

需要大模型具备优秀的 **数学计算** 与 **金融专业知识**

- ① 第一步：明确金融计算下的具体公式
- ② 第二步：依赖数学和推理能力计算结果

思维链COT能力

SFT数据构造



我国2000年的M2为 137543.6亿元，名义GDP为 89404亿元，根据费雪的交易方程式计算货币流通速度

首先，**费雪的交易方程式是 $MV = PY$** 。
这里，

- M 是货币供应量
- V 是货币流通速度
- P 是价格水平
- Y 是名义GDP

我们想要找出的是货币流通速度 V，而已知的是 M（货币供应量）和 Y（名义GDP）。

计算
公式

我们可以得到：

$$V = PY / M$$

在这个特定例子中，由于 PY 就是名义GDP，所以方程可以简化为：

$$V = Y / M$$

代入给出的数据，我们有：

$$V = 89404\text{亿元} / 137543.6\text{亿元}$$

$$V \text{ 约等于 } 0.65$$

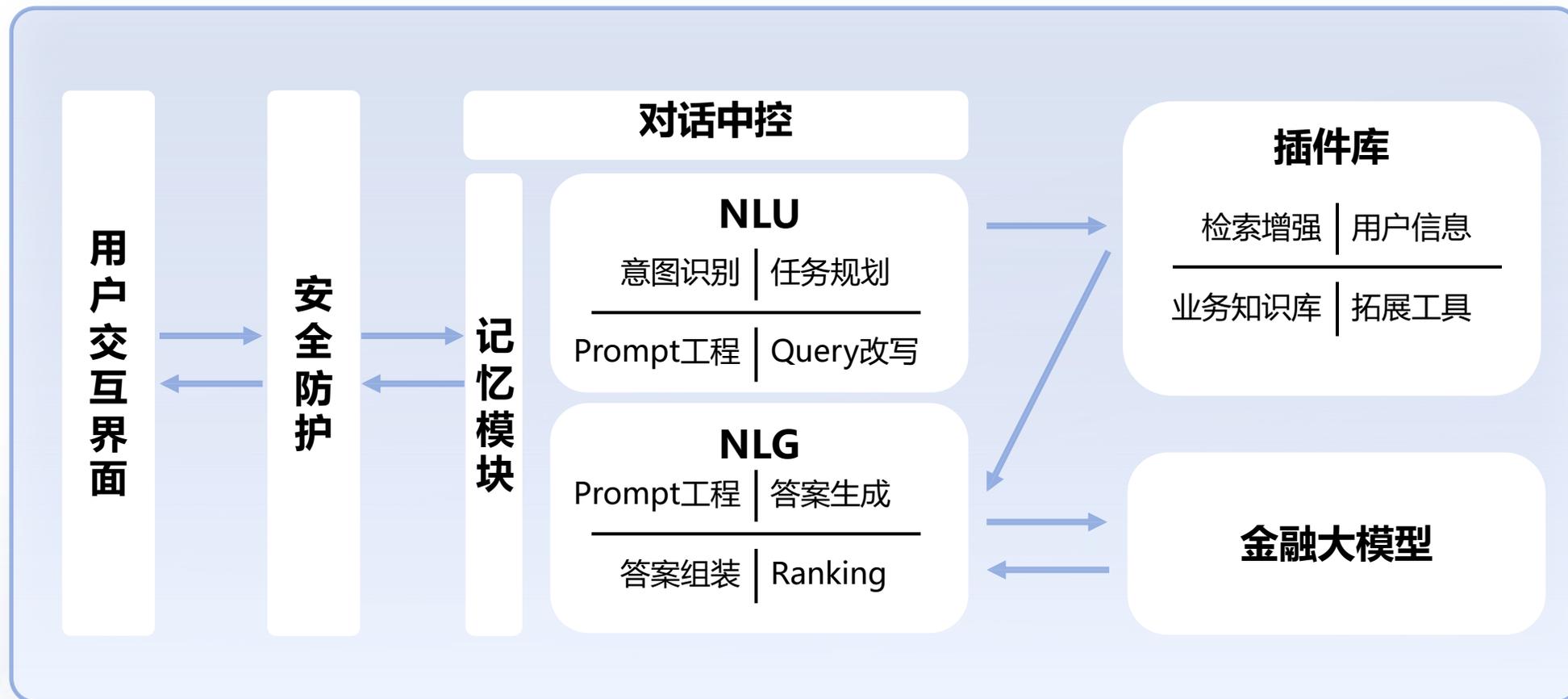
所以根据2000年的数据，货币流通速度大约为0.65

数学
计算

预训练阶段获得大量的金融知识

大模型本身足够强的数学计算和思维链能力

升级系统工程，弥补大模型本身能力欠缺



输出结果可靠性增强 ↑

输出质量稳定性提升 ↑

金融大模型应用实践创新

金融大模型应用实践创新：由点到面重塑金融价值链

重塑营销

一体化营销工作坊

- 传播洞察
- 智能投放
-

重塑服务

从降本增效到价值创造

- 客情摘要
- 客服助手
-

重塑运营

端到端的运营新范式

- NL2SQL
- 投研投顾
-

重塑研发

助力研发提质提效

- 代码生成
- 单测生成
-

重塑办公

员工的生产力工具

- 智能搜索
- 知识助手
-

重塑风控

智能化全面风控

- 智能信审
- 风险模拟
-

金融大模型核心能力

个性生成

交互增强

预测模拟

强自动化

理解 | 生成 | 逻辑 | 记忆

服务：提质降本，营销服务一体化

营销服务一体化

生成式AI

营销

服务

服务过程中洞察潜在需求，匹配和推荐对应金融产品，生成式AI串联起了原本分散的业务环节，实现营销服务一体化成为可能

成本中心

利润中心

可以将人力投入到更有价值的业务上，提升服务的同时，实现利润目标。

服务质量和服务能力提升

人工坐席

服务能力有效



模型辅助人工

内部提效25%

模型直接对客

隐形增员空间巨大



大模型辅助坐席提供服务，提升坐席整体的水平，持续提升服务承载能力和平均服务质量

服务成本降低

人力密集

技术密集

创新技术应用将释放大量人力，降低服务成本



样例展示——金融理财投顾

我不了解我的投资目标，你能帮我分析吗？可以向我提问并通过我的回答来判断



当然，我会尽力帮助您分析和了解您的投资目标。为了更好地理解您的情况，我会向您提出一些问题。请您认真思考并回答这些问题，这将有助于我更好地帮助您。

- 您的年龄是多少？
- 您的职业是什么？
- 您的家庭状况（例如：已婚、有子女等）是怎样的？
- 您的投资经验水平如何？（例如：完全没有经验、有一些经验但不太深入等）
- 您打算在多长时间实现投资目标？
- 您希望通过投资实现什么目标？（例如：退休储蓄、购买房屋、教育基金等）
- 您愿意承担多大程度的投资风险？（例如：非常低风险、中等风险、高风险等）
- 您打算投资哪些类型的资产？（例如：股票、债券、基金、房地产等）
- 您有多少资金可用于投资？
- 您是否有专业的财务顾问或投资顾问？

请认真回答这些问题，以便我更好地了解您的情况，从而为您提供更合适的投资建议。

样例展示——金融场景计算

银行向企业发放一笔贷款，贷款额为100万元，期限为4年，年利率为6%，试用单利和复利两种方式计算银行应得的本息和。



单利计算

在单利制度下，银行从企业那里得到的本息和可以通过以下方式计算：

贷款额 (P) = 100万元 年利率 (r) = 6% = 0.06 期限 (n) = 4年

本息和 (A) = P + Prn = 100万元 + 100万元 \times 0.06 \times 4 = 100万元 + 24万元 = 124万元

使用单利计算，银行应得的本息和是124万元。本息和 (A) = P + Prn = 100万元 + 100万元 \times 0.06 \times 4 = 100万元 + 24万元 = 124万元

使用单利计算，银行应得的本息和是124万元。

复利计算

在复利制度下，银行从企业那里得到的本息和可以通过以下方式计算：

贷款额 (P) = 100万元 年利率 (r) = 6% = 0.06 期限 (n) = 4年

本息和 (A) = P \times (1 + r)ⁿ = 100万元 \times (1 + 0.06)⁴ = 100万元 \times 1.262476 = 126.2476万元 = 126.25万元 (四舍五入到小数点后两位)

使用复利计算，银行应得的本息和是126.25万元。

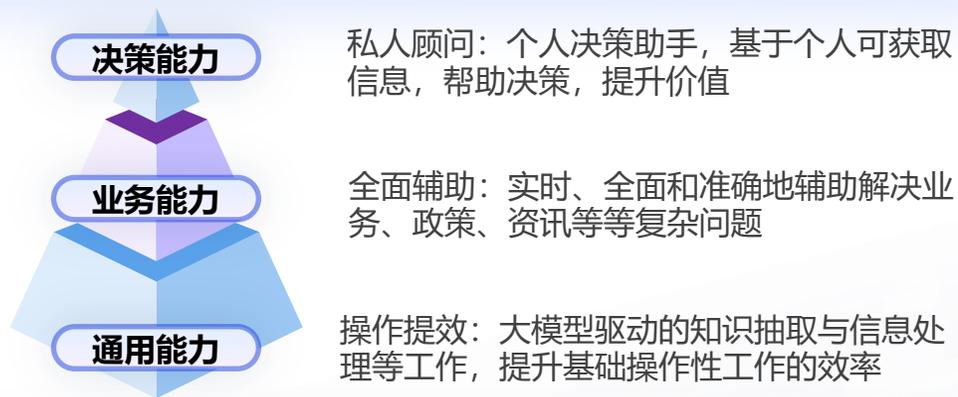
总结：

单利下银行应得本息和为：124万元

复利下银行应得本息和为：126.25万元

办公：无处不在、无缝衔接的办公助手

多层次、全方位的能力辅助



助力提升内外部协同效率

助力机构间业务合作



优化部门内资源配置

提升公司内协作效率



度小满 - ChatMan办公助手

真实问答场景

金融监管治理有几个要素，以及每个的含义和重要性？



金融监管治理涉及的四个核心要素是：独立性、问责制、透明度、操守。

首先，独立性指的是监管机构具有一定的自主权，使其能够免受政治层面和被监管实体的干扰，以确保公平、有效地进行金融监管。监管机构的独立性不仅要求专业性技能，以便在复杂情况下迅速做出反应，还要求其能够使自身免受政治干预，增加监管行为的透明度和稳定性，由此提高政策制定的可信度。

其次，问责制是监管治理的重要环节，.....

然后，透明度也是极为重要的，.....

最后，操守是指监管机构和监管人员需要坚守的行为与规范。.....

金融监管治理的四个要素相互影响、相互促进，共同构建了金融监管的基础环境。要统一提升，需要多元化的改革与整合，构建高效、公正的金融监管环境。

用户满足率从 20% 提升至 86%

样例展示——金融知识问答

等额本金和等额本息有什么区别？哪个划算？



等额本金和等额本息是两种常见的按揭还款方式。选择哪种方式取决于借款人的实际情况和需求。

首先，这两种还款方式在每月还款金额上存在区别。等额本金的还款金额是递减的，每个月偿还的本金金额固定，而利息的支付金额会随着时间推移递减。与之不同的是，等额本息的每月还款金额是固定的，但是在还款初期，偿还金额中利息所占的比例较大。

其次，这两种方式产生的利息也不同。总的来说，等额本金所产生的总利息要少于等额本息。举个例子，假设贷款50万元，借款期限为一年，贷款年利率为4.35%。如果选择等额本金还款方式，总利息为2356.25元；而如果选择等额本息还款方式，总利息为2371.88元。因此，等额本金相比等额本息更加划算。

最后，这两种还款方式适用的人群也有所不同。等额本金适合于早期收入较高、后期逐渐减少的借款人；而等额本息适合每月有固定收入的借款人。因此，在选择还款方式时，借款人应根据自己的收入情况和未来的规划进行综合考虑。

总之，等额本金和等额本息各有优劣，并没有绝对的哪种更好或更坏。借款人应根据自身情况，权衡利弊，选择适合自己的还款方式。

04 金融大模型评测方法创新

拒绝榜单绑架，用评测指引模型优化方向

“横评”看差距、“纵评”看提升

“横评”：不同模型看各项优劣

自研模型 VS GPT4 VS 国内主流模型A VS 国内主流模型B

“纵评”：同一模型看阶段提升

SFT-V1 VS SFT-V2 VS RLHF-V1 VS RLHF-V2

预训练阶段

- 训练是否存在异常
- 评估基座模型质量



指令微调阶段

- 对话能力能否满足
- 泛化能力是否足够



强化学习阶段

- 安全性是否提升
- 有用性能否保持

实时评测：CheckPoint自动触发评测流水线

阶段评测：「自动+人工」全维度评测体系

FinancelQ: 定制金融大模型能力评测体系

FinancelQ金融大模型评测体系

针对金融域的大模型
评测标准尚未完善



开源FinancelQ自动评测集
专注中文金融领域任务
涵盖10个金融大类, 36个金融
小类, 总计7173题



金融任务人工评测集
覆盖保险/信贷/理财/风控等多领域;
客服/营销/投顾/产品/催收/法务/行
研等不同岗位

自动金融评测集

- 初级《经济基础知识》
- 初级《专业知识与服务》
- 中级《经济基础知识》
- 中级《专业知识与服务》
- 《税法(一)》
- 《税法(二)》
- 《涉税服务相关法律》
- 《财务与会计》
- 《涉税服务实务》
- 《审计》
- 《财务成本管理》
- 《经济法》
- 《会计》
- 《公司战略与风险管理》
- 《税法》
- 《基础知识》
- 《专业能力》
- 《金融数学》



- 《个人理财》
- 《公司信贷》
- 《个人贷款》
- 《风险管理》
- 《银行考试》
- 《银行业法律法规与综合能力》

- 《基金法律法规、职业道德与业务规范》
- 《证券投资基金基础知识》
- 《私募股权投资基金基础知识》

- 《证券市场基本法律法规》
- 《金融市场基础知识》

- 《期货基础知识》
- 《期货法律法规》
- 《期货投资分析》

- 《保险基础知识》
- 《保险实务》
- 《保险法规》
- 《保险市场与营销》

金融
知识
理解

- 金融术语解释
- 金融知识解读
- 金融常识百科
- 金融产品

金融
信息
摘要

- 金融新闻摘要
- 研报财经摘要
- 观点提取
- 文本抽取

金融
内容
生成

- 资讯标题生成
- 营销文案生成
- 保险条款解读
- 金融财经
- 金融行情解读
- 客服话术生成

金融
逻辑
计算

- 金融计算

金融
安全

- 金融法律法规
- 安全合规性问题

金融
实时

- 实时资讯
- 实时股价

人工金融评测集

从模型训练到场景落地：像评测人一样评测大模型



通识基础

“中学生：接受通识教育”

- 难度均衡
- 风格多样
- 题量充足

客观评测：23个任务，10w+题，全维度评测体系

主观评测：14大维度、600+题自制评测集，多人打分

行业赋能

“大学生：接受专业教育”

金融行业知识	金融客户对话	金融逻辑计算
金融内容生成	金融数据分析	金融信息理解

客观评测：FinanceIQ评测集，覆盖10大金融考试

主观评测：19大维度、500+金融域专项任务

场景反馈

“职场人：以岗位产出为导向”

- 知识问答
- 客服话术
- 产品创新
- 投顾服务
- 研报解读
- 财报生成
- 客户关系
- 风险合规
- 渠道运营
- 信贷审批
- 财富顾问
- 虚拟客服

客观评测：用户满足率与KPI指标

主观评测：业务反馈与用户建议

05

总结：金融大模型迭代路径

金融大模型迭代路径：知识和能力兼备，评估和实践共进

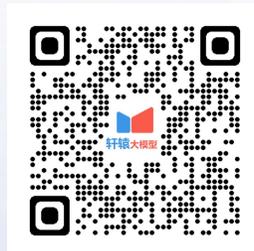


轩辕金融大模型将持续开源，与行业伙伴共同成长

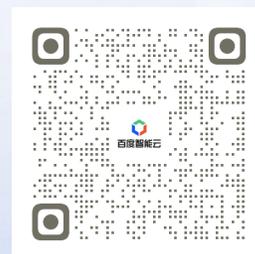


轩辕系列大模型，技术认知和实践经验总结

《大语言模型原理与工程实践》即将出版



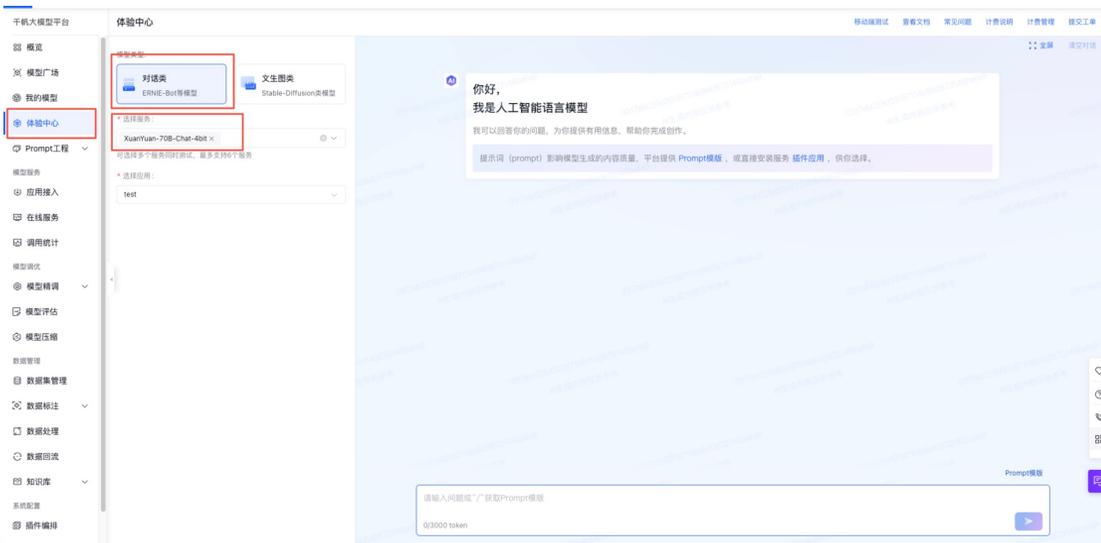
扫码访问GitHub页面



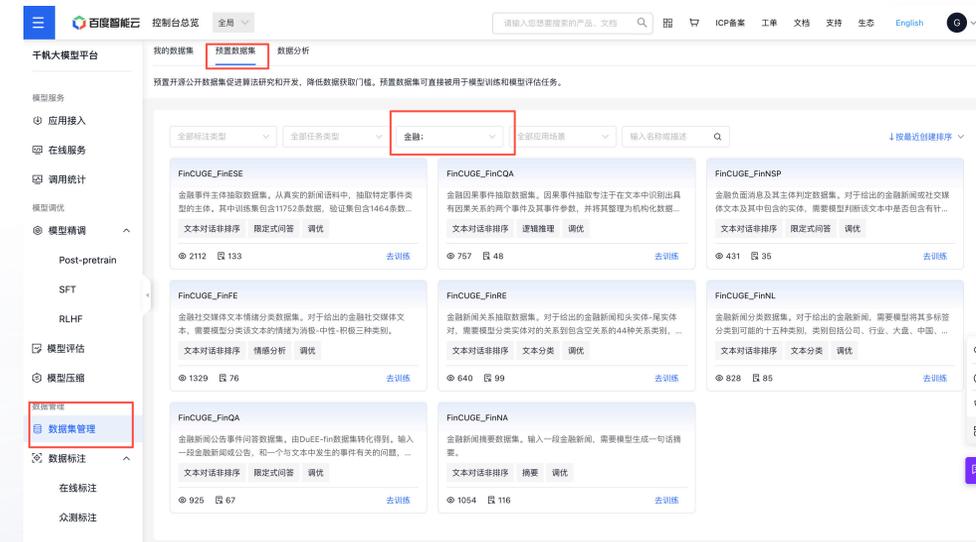
在千帆社区
分享轩辕大模型使用体验

课后作业

基础作业：在百度智能云千帆大模型平台 体验中心测试轩辕大模型 使用不少于20个prompt



进阶作业：使用1-3个金融行业预置数据集完成场景调优
FinCUGE_FinNA, FinCUGE_FinQA, FinCUGE_FinNL, FinCUGE_FinRE, FinCUGE_FinFE, FinCUGE_FinNSP, FinCUGE_FinCQA, FinCUGE_FinESE



***将上述作业在【百度智能云千帆社区】进行发布，发布时选择 #大模型实训营 话题**



T H A N K S