

## 一、赛题描述

银行作为主要的基金代销渠道之一，其从业人员需要对各类基金公告进行整理与研究。招商银行作为基金代销头部机构之一，需要依托前沿技术推进基金代销业务的高质量发展。

近年来，人工智能技术在金融领域逐渐兴起，尤其是大语言模型（下称“大模型”）在投研领域的应用价值备受关注。基金公告包含了基金基础信息、运作情况和业绩表现等重要信息，是投研人员判断基金投资价值的关键依据，但是传统的基金公告分析主要依赖于人工解读，需要深厚的专业知识和高昂的时间成本。从业人员能够利用大模型强大的数据处理和学习能力，从海量知识库中提取关键信息，并进行深入分析和挖掘。因此，利用基金公告数据构建基于大模型的投研 Agent，对于辅助投研人员获取基金信息、提高投资研究效率以及促进金融科技创新具有重要意义。

本次赛题要求选手利用大模型构建投研 Agent，并使用该“智能体”完成基金公告数据解析理解、用户提问和分析需求等任务。参赛选手可以自行选择大模型，不限制使用公开的预训练模型，不限制使用公开数据进行微调，鼓励使用 SFT、RAG、LoRA 等大模型相关技术。

本次赛题从以下四个方面评测投研 Agent 的能力：

### （1）基础信息问答查询

选手需要利用大模型和基金公告数据，针对基础信息问答查询，生成答案，问题答案可以在公告数据中直接获取。

示例序号	示例问题	示例答案
示例一	"2023 年，中庚价值领航混合在报告期末的基金份额总额为多少？（保留 2 位有效数字，不加千位分隔符）"	"2023 年，中庚价值领航混合在报告期末的基金份额总额为 3902189118.33 份。"
示例二	"2023 年，中庚价值领航混合在报告期的基金份额净值增长率为多少？（以百分数表示，保留 2 位有效数字）"	"2023 年，中庚价值领航混合在报告期末的基金单位净值增长率为-3.84%。"

### （2）核心指标统计分析

选手需要利用大模型和基金公告数据，针对核心指标统计分析问题，生成答案，问题答案涉及指标数据的简单计算和关联对比。

示例序号	示例问题	示例答案
示例一	"2022 年，中庚价值领航混合在报告期末的可供分配利润比 2023 年高多少？（保留 2 位有效数字，不加千位分隔符）"	"2022 年，中庚价值领航混合在报告期末的可供分配利润比 2023 年高 1435179967.26 人民币元。"
示例二	"2022 年，中庚价值领航混合在报告期的份额净值增长率比 2023 年高多少？（以百分数表示，保留 2 位有效数字）"	"2022 年，中庚价值领航混合在报告期的份额净值增长率比 2023 年高 8.69%。"

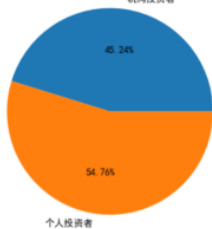
### （3）结构化信息抽取

选手需要利用大模型和基金公告数据，针对信息抽取需求，按指定 json 格式抽取结构化信息。

示例序号	示例问题	示例答案
示例一	"请以 json 格式抽取 2023 年报告期末，中庚价值领航混合中占基金资产净值比例前 5 大的股票名称以及相应比例，键名为股票名称，键值为净值比例，净值比例以百分数表示，保留 2 位有效数字。"	{"中国宏桥": "9.66%", "绿叶制药": "6.52%", "小鹏汽车－W": "5.11%", "赛腾股份": "5.48%", "中国海外发展": "4.45%"}

**(4) 基金分析报告生成**

选手需要利用大模型和基金公告数据，针对简单的基金分析需求，生成分析文本以及统计图。

示例序号	示例问题	示例答案				
示例一	"请分析 2023 年报告期末，中庚价值领航混合的基金份额持有人结构信息，并按份额比例绘制饼状统计图。"	"2023 年报告期末，中庚价值领航混合机构投资者持有份额占比 45.24%，……"				
		<div><p>持有人持有份额占比</p><table><tr><th>投资者类型</th><th>占比</th></tr><tr><td>机构投资者</td><td>45.24%</td></tr><tr><td>个人投资者</td><td>54.76%</td></tr></table></div>	投资者类型	占比	机构投资者	45.24%
投资者类型	占比					
机构投资者	45.24%					
个人投资者	54.76%					

本次赛题的评测方式为：给定基金公告数据和评测问题文件，要求构建投研 Agent 按指定格式生成答案。

**二、数据说明**

本次赛题提供的数据包括：基金公告数据 [announcements.zip](#)、评测问题文件 [question.json](#) 和示例文件 [example.zip](#)。（请点击文件名称进行下载）

**(1) 基金公告数据 announcements.zip**

文件包含本次赛题提供的基金公告文件，共 96 个 pdf 文件，涉及 16 支基金（每支基金 4 份季报、1 份中期报告、1 份年度报告），参赛选手可以对 pdf 文件进行解析并将文件内容作为构建投研 Agent 的数据基础。

**(2) 评测问题文件 question.json**

文件为评测投研 Agent 能力的问题集合，共 110 个问题，文件每一行为一个 json 串，键名包括" id "和" question "，示例如下：

<code>{"id": 0, "question": "2023 年，中庚价值领航混合在报告期末的基金份额总额为多少？（保留 2 位有效数字，不加千位分隔符）"}</code>
<code>{"id": 1, "question": "2022 年，中庚价值领航混合在报告期末的可供分配利润比 2023 年高多少？（保留 2 位有效数字，不加千位分隔符）"}</code>
<code>{"id": 2, "question": "请以 json 格式抽取 2023 年报告期末，中庚价值领航混合中占基金资产净值比例前 5 大的股票名称以及相应比例，键名为股票名称，键值为净值比例，净值比例以百分数表示，保留 2 位有效数字。"}</code>
<code>{"id": 3, "question": "请分析 2023 年报告期末，中庚价值领航混合的基金份额持有人结构信息，并按份额比例绘制饼状统计图。"}</code>
.....

### （3）示例文件 example.zip

文件包含 1 份示例基金公告 pdf 文件，1 份示例问题文件 question\_example.json，1 份示例答案文件 answer\_example.json。

## 三、提交说明

本次赛题选手需提交一份方案报告、一份视频文件、一份答案文件和一份源代码文件。选手需要在赛题截止时间前提交全部四份文件，不可重复提交。

### （1）方案报告（要求 PDF 文件，文件大小 20 页以内，不超过 10M）

方案报告为本次赛题的总结性报告，文件命名为"姓名+report.pdf"。参赛选手需要在方案报告中详细说明本次赛题的整体方案、关键技术和问答示例等内容，方案报告是方案评分的重要依据，必须包含但不限于以下内容：

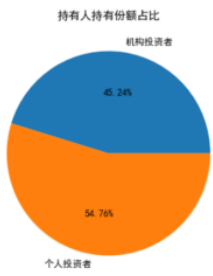
a、整体方案介绍：参赛选手需要说明所使用的大模型信息，并绘制流程图说明本次赛题的整体解决方案。

b、关键技术介绍：请重点介绍 SFT、RAG 等大模型相关技术方法（如有），并附上关键函数代码截图。

c、Prompt 工程介绍：请分别介绍四种类型任务的 Prompt 工程、优化思路等内容，并以表格形式举例说明投研 Agent 生成的答案，示例格式如下：

问题 id	问题	Prompt	答案
0	"2023 年，中庚价值领航混合在报告期末的基金份额总额为多少？（保留 2 位有效数字，不加千位分隔符）"	"假设你是一名资深基金经理，请回答中庚价值领航混合在报告期末的基金份额总额为多少？....."	2023 年，中庚价值领航混合在报告期末的基金份额总额为 3902189118.33 份。

d、基金分析报告生成答案：请以表格形式、按问题 id 顺序（报告生成题的问题 id 为 100 至 109）展示基金分析报告生成任务中投研 Agent 生成的图文答案，示例格式如下：

问题 id	问题	文本分析答案	统计图
3	"请分析 2023 年报告期末，中庚价值领航混合的基金份额持有人结构信息，并按份额比例绘制饼状统计图。"	"2023 年报告期末，中庚价值领航混合机构投资者持有份额占比 45.24%，……"	

**（2）视频文件（要求 MP4 文件，大小不超过 500M）**  
解说视频是对方案报告的口语介绍和录频展示，同样作为方案评分的重要依据，文件名称为"姓名+commentary.mp4"。解说视频中，参赛选手需要选择四种类型任务的题目各一道，以选择的题目为例简要介绍该类型任务的解决思路、Prompt 工程、关键函数等，而后运行代码，记录大模型生成答案的过程。

**（3）答案文件（要求 JSON 文件）**  
答案文件为投研 Agent 对于评测问题的回答内容，文件名称为"姓名+answer.json"，文件每一行为一个 json 串，键名包括"id"，"question"和"answer"，基金分析报告生成任务的答案请随方案报告提交，无需写在答案文件中，答案文件示例如下：

<pre>{ "id": 0, "question": "2023 年，中庚价值领航混合在报告期末的基金份额总额为多少？（保留 2 位有效数字，不加千位分隔符）", "answer": "2023 年，中庚价值领航混合在报告期末的基金份额总额为 3902189118.33 份。" }</pre>
<pre>{ "id": 1, "question": "2022 年，中庚价值领航混合在报告期末的可供分配利润比 2023 年高多少？（保留 2 位有效数字，不加千位分隔符）", "answer": "2022 年，中庚价值领航混合在报告期末的可供分配利润比 2023 年高 1435179967.26 人民币元。" }</pre>
<pre>{ "id": 2, "question": "请以 json 格式抽取 2023 年报告期末，中庚价值领航混合中占基金资产净值比例前 5 大的股票名称以及相应比例，键名为股票名称，键值为净值比例，净值比例以百分数表示，保留 2 位有效数字。", "answer": { "中国宏桥": "9.66%", "绿叶制药": "6.52%", "小鹏汽车－W": "5.11%", "赛腾股份": "5.48%", "中国海外发展": "4.45%" } }</pre>
.....

注：本次赛题不限制问题的 prompt 次数和内容，但"answer"必须为投研 Agent 生成的答案，严禁手工修改，我们将核实结果的可复现性。

**（4）源代码文件（要求 ZIP 文件，大小不超过 500M）**  
文件名称为"姓名+code.zip"，文件夹内包括但不限于可执行代码、执行结果、代码说明等文件，我们将依据源代码文件和项目方案报告核实结果的可复现性。

四、评审规则

本次赛题的最终成绩由方案得分和客观得分加权构成，加权比例为 7：3。

比赛成绩=0.7\*方案得分+0.3\*客观得分

(1) 方案得分

赛题结束后，我们将依据方案报告和解说视频，组织专家对参赛选手的解决方案进行评分，满分 100 分，根据选手对于问题和场景的深度理解、解决方案的创新性以及大模型相关技术的使用进行评估。

(2) 客观得分

得分由所有题目的得分累加，满分 120 分，换算至百分制后与进行加权得到比赛成绩。比赛结束后，依据下述规则对选手提交的答案进行评分：

a、基础信息问答查询：

共计 60 题，每题 1 分，答案信息错误的情况下不得分，得分=0；答案信息正确的情况下，答案信息和基础信息各占 0.5 分，得分=0.5 + 0.5 \* 基础信息正确数量占比。

示例序号	示例问题	示例答案	基础信息	答案信息
示例一	"2023 年，中庚价值领航混合在报告期末的基金份额总额为多少？（保留 2 位有效数字，不加千位分隔符）"	"2023 年，中庚价值领航混合在报告期末的基金份额总额为 3902189118.33 份。"	{"2023", "中庚价值领航混合", "份额总额"}	{"3902189118.33"}
示例回答内容				得分
"2023 年,中庚价值领航混合在报告期末的基金份额总额为 3902189118.33 份。"				1
"2023 年，中庚价值领航混合在报告期末的基金份额总额为 500000000000.00 份。"				0
"基金份额总额为 3902189118.33 份。"				0.67
"3902189118.33 份"				0.5
b、核心指标统计分析：				



共计 20 题，每题 1 分，答案信息错误的情况下不得分，得分=0；答案信息正确的情况下，答案信息和基础信息各占 0.5 分，得分=0.5 + 0.5 \* 基础信息正确数量占比。

示例序号	示例问题	示例答案	基础信息	答案信息
示例一	"2022 年，中庚价值领航混合在报告期末的可供分配利润比 2023 年高多少？（保留 2 位有效数字，不加千位分隔符）"	"2022 年，中庚价值领航混合在报告期末的可供分配利润比 2023 年高 1435179967.26 人民币元"	{"2022", "2023", "中庚价值领航混合", "可供分配利润"}	{"1435179967.26"}

示例回答内容	得分
"2022 年，中庚价值领航混合在报告期末的可供分配利润比 2023 年高 1435179967.26 人民币元"	1
"2022 年，中庚价值领航混合在报告期末的可供分配利润比 2023 年高 500000000000.00 人民币元"	0
"可供分配利润比 2023 年高 1435179967.26 人民币元"	0.75
"1435179967.26 人民币元"	0.5

c、结构化信息抽取：  
共计 20 题，每题 1 分，分值平均分配给每一个键值对。键名信息错误的情况下，该键值对不得分，键名信息正确、键值信息错误的情况下，该键值对得一半分，键名信息和键值信息均正确的情况下，该键值对得全部分。

示例回答内容	得分
{"中国宏桥": "9.66%", "绿叶制药": "6.52%", "小鹏汽车－W": " 5.11%", "赛腾股份": "5.48%", "中国海外发展": "4.45%"}	1
{"中国宏桥": "9.66%", "绿叶制药": "6.52%", "小鹏汽车－W": " 5.11%", "赛腾股份": "5.48%"}	0.8 缺少一个键值对
{"中国宏桥": "10%", "绿叶制药": "6.52%", "小鹏汽车－W": " 5.11%", "赛腾股份": "5.48%", "中国海外发展": "4.45%"}	0.9 "中国宏桥"键值错误
{"中国红桥": "9.6%", "绿叶制药": "6.52%", "小鹏汽车－W": " 5.11%", "赛腾股份": "5.48%", "中国海外发展": "4.45%"}	0.8 "中国宏桥"键名错误

d、基金分析报告生成：

共计 10 题，每题 2 分，依据基金公告分析题的图文结果进行主观评分，分析文本占 0.5 分，统计图占 1.5 分，统计图能够正确展示分析信息即可。

回答内容	得分
能够生成合理的分析文本和正确的统计图	2
能够生成合理的分析文本和有些许错误的统计图	1
能够生成合理的分析文本，未生成统计图	0.5
未能生成合理的分析文本，未生成统计图	0

### 五、关于计算资源的提示

本次 AI 赛道复赛命题组已充分考虑题目复杂度与计算资源获取难度。建议各位同学在赛题发布前，通过实验室、公有云等渠道获取计算资源（8 核、32G 内存、24G 显存左右），并先行调试环境。可以使用预训练后的大模型权重文件，可以通过模型压缩、参数半精度化等技术减小资源占用。预祝各位同学取得优异成绩。