# 中海油人工智能项目会议纪要

## 一、会议基本信息

* **时间**：2025 年 06 月 10 日
* **参与人员**：齐家平，李松伟，黄子豪，李建军，李雪莲，张子健，王洪雨
* **形式**：线下会议

## 二、项目需求与目标

### （一）核心需求

* **办公需求**：支持远程办公。
* **模型需求**：以小模型为主，计算效果较好时再考虑推广大模型。
* **数据输入：**规则库和指标体系。

### （二）关键目标

* 通过数据喂给模型进行训练，发现并挖掘规则。
* 利用邮件系统评估指标体系，通过具体案例抽取规则和指标进行推理判断。
* 先基于小模型训练，后期扩展优化至大模型

## 三、讨论要点及结论

### （一）资源分配与模型应用

|  |  |  |  |
| --- | --- | --- | --- |
| 讨论要点
 | 我方观点
 | 客户反馈
 | 结论
 |
| CPU 拆分问题
 | 需借助大模型推理数据，判断 96 核 CPU 拆分为 2 台或 3 台虚拟机
 | 客户已收集拆分的虚拟机数量及资源明细，用户侧将确定架构及详细配置
 | 客户提供资源数据，我方结合小模型分析，后期拓展至大模型
 |
| 模型选择与分工
 | 小模型用于计算总量，大模型用于规则和计算
 | 客户要求用算法规则套入，注重泛用性，认为大模型会增加对接和开发工作量
 | 先使用小模型快速落地成果，后续优化并考虑大模型
 |
| 规则与训练数据
 | 小模型训练需结合人为输出的规则和约束，保证全面准确
 | 客户总结的规则用于小模型训练，可补充提供华为 C86 等数据（优先考虑 C86）
 | 小模型训练数据 + 大模型训练，前期以小模型为主
 |

### （二）业务系统与数据处理

* **业务系统情况**：现有 19 个业务系统，后续将增至 40 个。部分业务数据可先开展小模型训练，需申请服务器并进行数据脱敏。
* **资源分配难点**：一个业务系统如何分配虚机数量是难点，客户表示能出总量，后续考虑具体分配或者人工分配。
* **数据提供**：客户可提供 10 台服务器的 CPU、内存、最大最小使用率、平均使用率等基础数据，PASS 数据后续补充。

### （三）工作阶段规划

* **一阶段**：梳理数据集，调整指标项，完成代码模型。
* **二阶段**：模型训练，调试。
* 三阶段：成果反馈，微调等。

## 四、痛点与解决方案

|  |  |
| --- | --- |
| 痛点
 | 解决方案
 |
| 业务沟通时间长，资源信息不清晰
 | 通过 AI 处理用户输入的最小表单（客户调研表），替代人工反复沟通计算
 |
| 数据量少（仅 40 条）
 | 客户补充提供华为数据（C86、ARM、X86，优先 C86）
 |
| 整理全面规则费劲
 | 采用启发式算法，小模型为主，结合大模型能力
 |

## 五、后续安排

* **方案输出**：本周五（6 月 13 日）线上沟通并输出方案。
* **沟通频率**：按周沟通。
* **任务跟进**：明确文档整理、排期及人员分工。

## 六、其他事项

* 客户可提供大模型，我方无需专属资源。
* 指标体系通过邮件系统验证。