AiDD AI+ 软件研发数字峰会
AI+ software Development Digital summit

AI驱动软件研发
全面进入数字化时代
中国·深圳 11.24-25

AI+
software
Development
Digital
summit

TableGPT-大模型的漫长落地之路初探

赵俊博 浙江大学

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

KEYLINK ing

K+峰会

K+峰会 深圳站

K+ 全球软件研发行业创新峰会

会议时间：2024.05.24-25

K+峰会 上海站

K+ 全球软件研发行业创新峰会

会议时间：2024.09.20-21

AiDD 峰会

AiDD峰会 深圳站

AI+ 软件研发数字峰会

会议时间：2023.11.24-25

AiDD峰会 北京站

AI+ 软件研发数字峰会

会议时间：2024.07.19-20

AiDD峰会 深圳站

AI+ 软件研发数字峰会

会议时间：2024.11.15-16

演讲嘉宾

赵俊博（Jake）

浙江大学百人计划研究员

浙江大学计算机学院百人计划研究员、博士生导师

浙江大学图灵班项目主任

浙江大学计算机创新技术研究院 人工智能前沿中心主任

前连续创业者，前 Facebook，前 NVIDIA

纽约大学博士，师从图灵奖得主 Yann LeCun

福布斯30Under30科技赛道封面人物，阿里云MVP，百度青年 AI 学者，首届 WAIC 青年科学家

TableGPT： 将表格、自然语言和命令统一为一个 GPT

预训练大语言模型（LLM）的高速发展革新了人机交互、信息获取的方式，但是 LLM 在面对精准定量问题中展现出来的“刚性”和准确性仍然差强人意。

在本次报告中，我们介绍TableGPT，一个融合自然语言交互、对结构化数据的向量化理解和交互链式指令集的完备系统，在落地场合中通过与用户进行自然语言交互实现对表格文件或者数据库表的增删改查和其他复杂操作，兼容自动化大小模型交互，并支持可视化图表生成和简单的报告撰写。

团队介绍

科研支持——M3实验室

实验室主要研究方向包括Data-centric AI、AI+X交叉、预训练大模型与AIGC等课题。

实验团队在NIPS、ICLR、ICML等顶会每年发表论文10余篇，多次获得best paper，多次在Nature等著名期刊发表论文。

工业落地支持——浙江大学计算机创新技术研究院

浙江大学
计算机创新技术研究院
Institute of Computing Innovation, Zhejiang University

研究院旨在打造数字经济人才聚集地，做培育科技企业的“创新加速器”，将在资本、技术、人才、场地等各维度赋能企业，加速企业孵化，打造创新与创业相融合的新样板窗口。

硬件支持——中昊芯英（杭州）有限公司

中昊芯英构建了“自研训练芯片+超算集群+AIGC 预训练大模型”的产业价值链，打造完整的软硬件一体化方案，为全球客户提供具备生产变革能力的人工智能创新技术方案，加速人工智能的工程落地与产业化进程

目录
CONTENTS

PART 01
研究背景

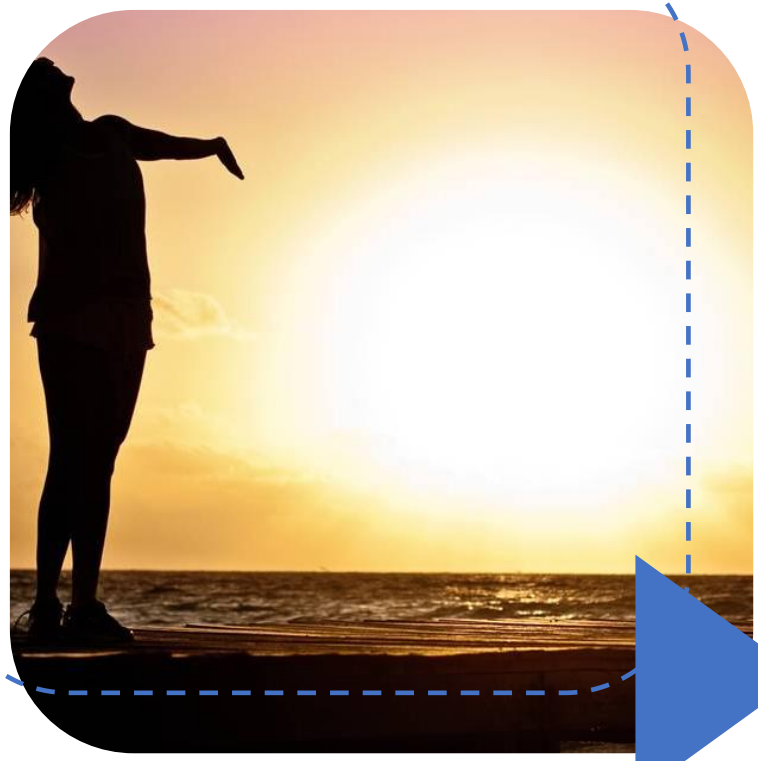
- 给一个灵魂拷问
- 针对LLM现状的思考：柔性和刚性
- 为什么我们要做一个刚性的LLM
- 我们要做一个什么样的刚性LLM

给一个灵魂拷问：

你会为一个闲聊的玩具买单吗？

可能，我们至死是少年...
所以还是会的

但是...



针对LLM现状的思考：
柔性和刚性

针对LLM现状的思考：柔性和刚性

什么是LLM的柔性和刚性？

柔性LLM

ChatGPT系列等

错了就错了，无所谓！
目的是提高人机交互体验

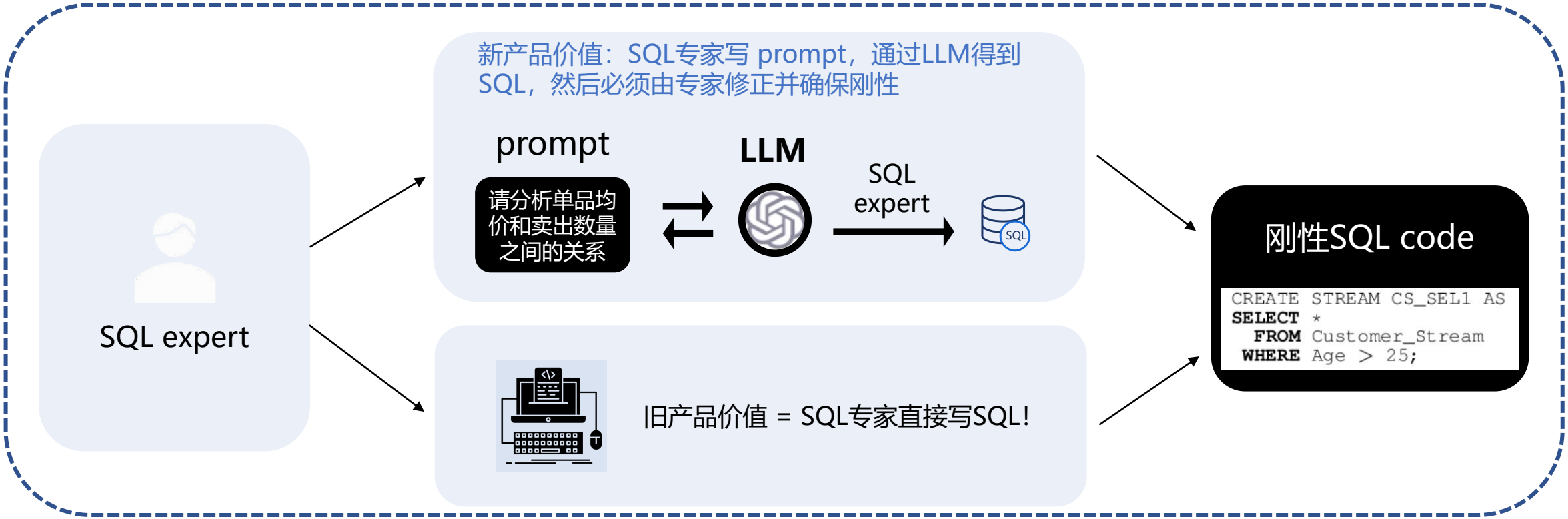
刚性LLM

TableGPT

严格、严谨
不能出错

针对LLM现状的思考：柔性和刚性

举一个NL2SQL的例子来分析一个柔性LLM的产品价值



对于一个柔性 LLM 的产品价值分析，产品边际价值 = 新产品价值 – 旧产品价值 – 切换成本 （《产品方法论》）

真正最大化产品边际价值的方式就是 把这个SQL专家 换成不懂SQL的人，同时确保刚性

而这，就是TableGPT做的事

为什么我们要做一个**刚性的LLM?**

因为，**文理需要兼修...**

为什么我们要做一个刚性的LLM

文科生通常更注重人文关怀和感性理解，而理科生则更注重逻辑推理和理性分析。

我们要做一个
什么样的刚性LLM

我们要做一个什么样的刚性 LLM

对话能力是刚性LLM的基础

在对话基础上，具备普适性业务数据处理能力，而表格是我们日常实用的数据格式，并且数据库是各处都有的基础设置

所以我们要做一个能用自然语言对话、能处理表格并具备刚性的LLM

我们要做一个什么样的刚性LLM

能做什么？

一款可以读懂表格的 LLM，

可以根据表格内容聊天以完成工作的伙伴

谁能用？

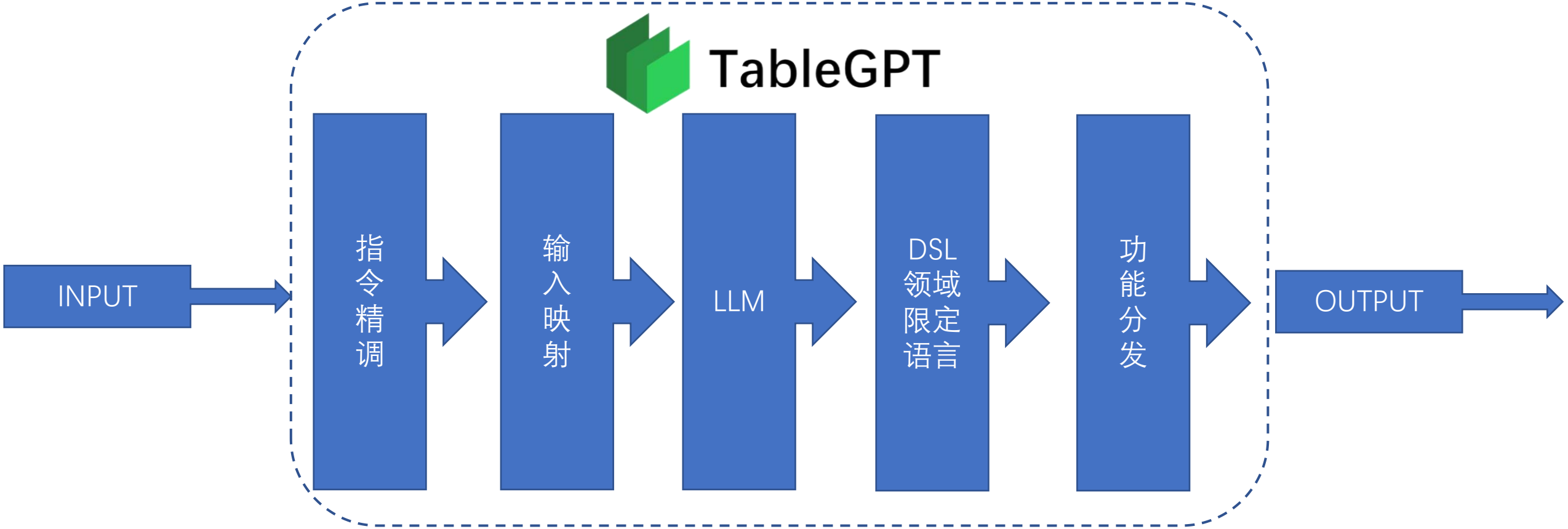
Everyone！

企业管理者、数据分析师、学生……

零门槛让你玩转表格 & 数据库类产品

我们要做一个什么样的刚性LLM

实现一个刚性LLM的初步思路



为什么是TableGPT?

业界也涌现了一些为table设计的LLM生态产品

如GPT plugins-Show me diagrams
其能够完成面向数据到图表的可视生成

能够实现的

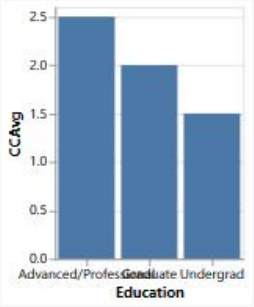
无法实现包含数理或逻辑关系的预测
不能根据领域知识私有数据生成文字描述

无法实现的

REQUEST TO SHOW ME DIAGRAMS


```
{
  "openApiSchemaVersion": "1.0",
  "diagramLanguage": "vegalite",
  "diagramType": "bar-chart",
  "diagram": "{\n  \"$schema\": \"https://vega.github.io/schema/vega-lite/v5.json\",\n
\"description\": \"Average monthly credit card flow for different education
levels.\",\n  \"data\": {\n    \"values\": [\n      {\"Education\": \"Undergrad\",
\"CCAvg\": 1.5},\n      {\"Education\": \"Graduate\", \"CCAvg\": 2.0},\n
{\"Education\": \"Advanced/Professional\", \"CCAvg\": 2.5}\n    ]\n  },\n  \"mark\":
\"bar\",\n  \"encoding\": {\n    \"x\": {\"field\": \"Education\", \"type\":
\"nominal\", \"axis\": {\"labelAngle\": 0}},\n    \"y\": {\"field\": \"CCAvg\",
\"type\": \"quantitative\"}\n  },\n  \"width\": { \"step\": 50 }\n}"
}
```


Here is a bar chart showing the average monthly credit card flow for different education levels:



此类产品多是对LLM的table适配
训练未真正融入表格特征
难以对表格深度的理解和分析

PART 02
TableGPT功能展示

TabelGPT功能展示

表格自然操作

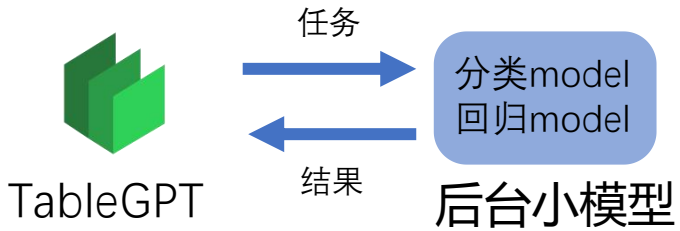
1. 自然语言查询 / 修改 / 增加 / 删除
2. 自然语义理解GroupBy等分组聚合条件
3. 指令链——分步拆解用户意图并规划指令实现链条

数据可视化&生成分析

1. 智能理解表格含义自主选择可视化形式（柱状图、折线图、箱图等）和内容（横纵轴数据）
2. 根据可视化内容或表格内容总结数值特点，生成可结合领域知识的分析报告

自动化建模预测

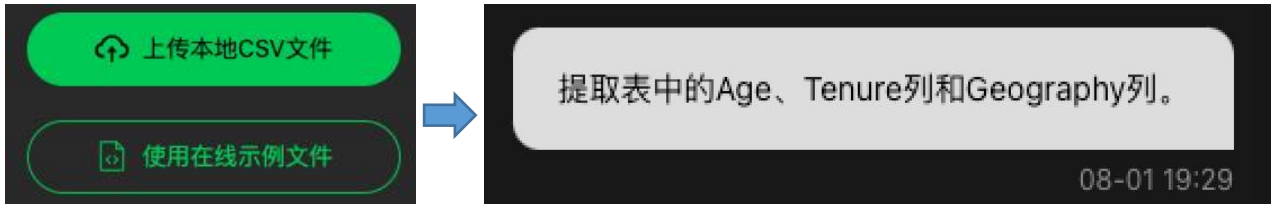
1. 结合表格特征自动化配置建模预测任务，助力数据分析



大小模型协作，自动化适配不同的表格建模任务

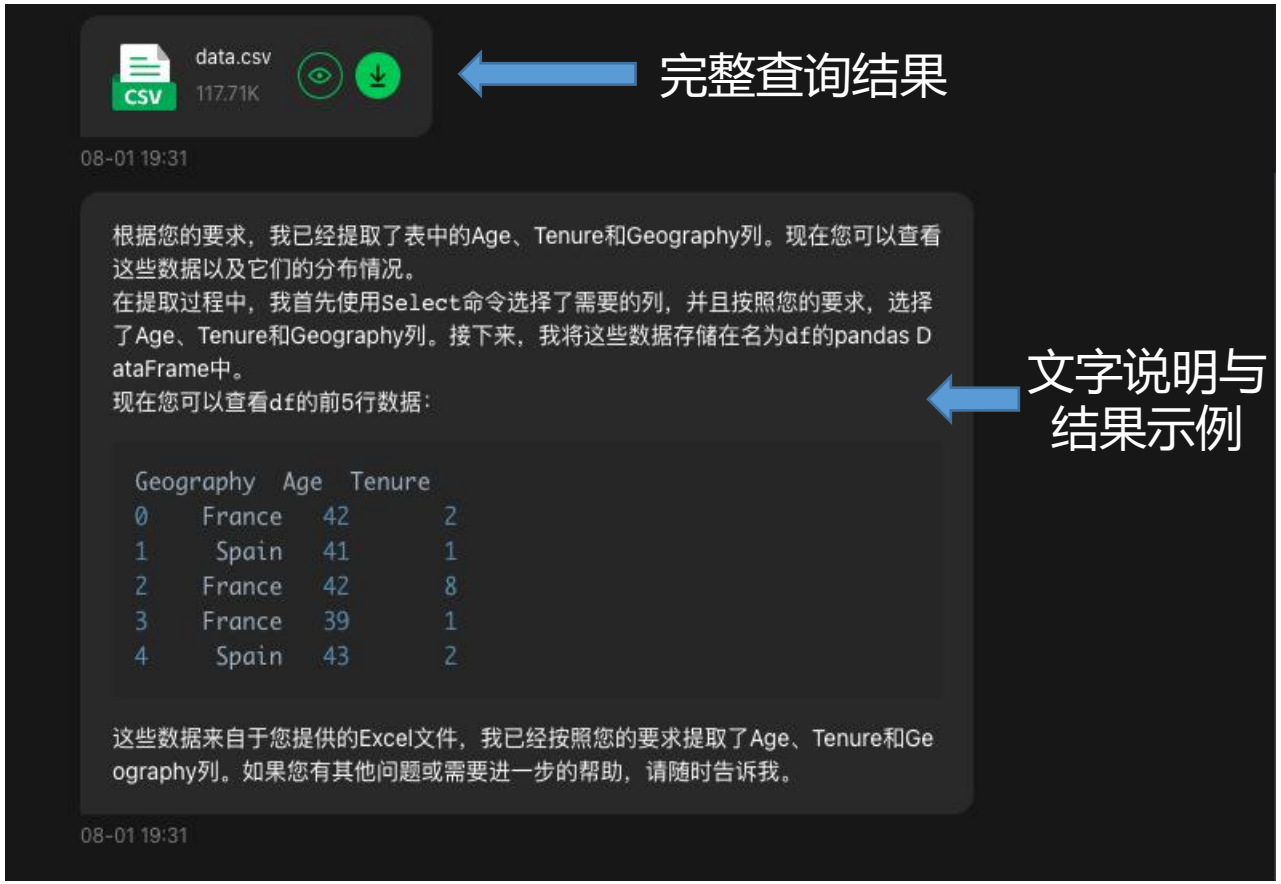
表格自然操作 → 表格可视分析 → 表格自动建模

用户上传数据



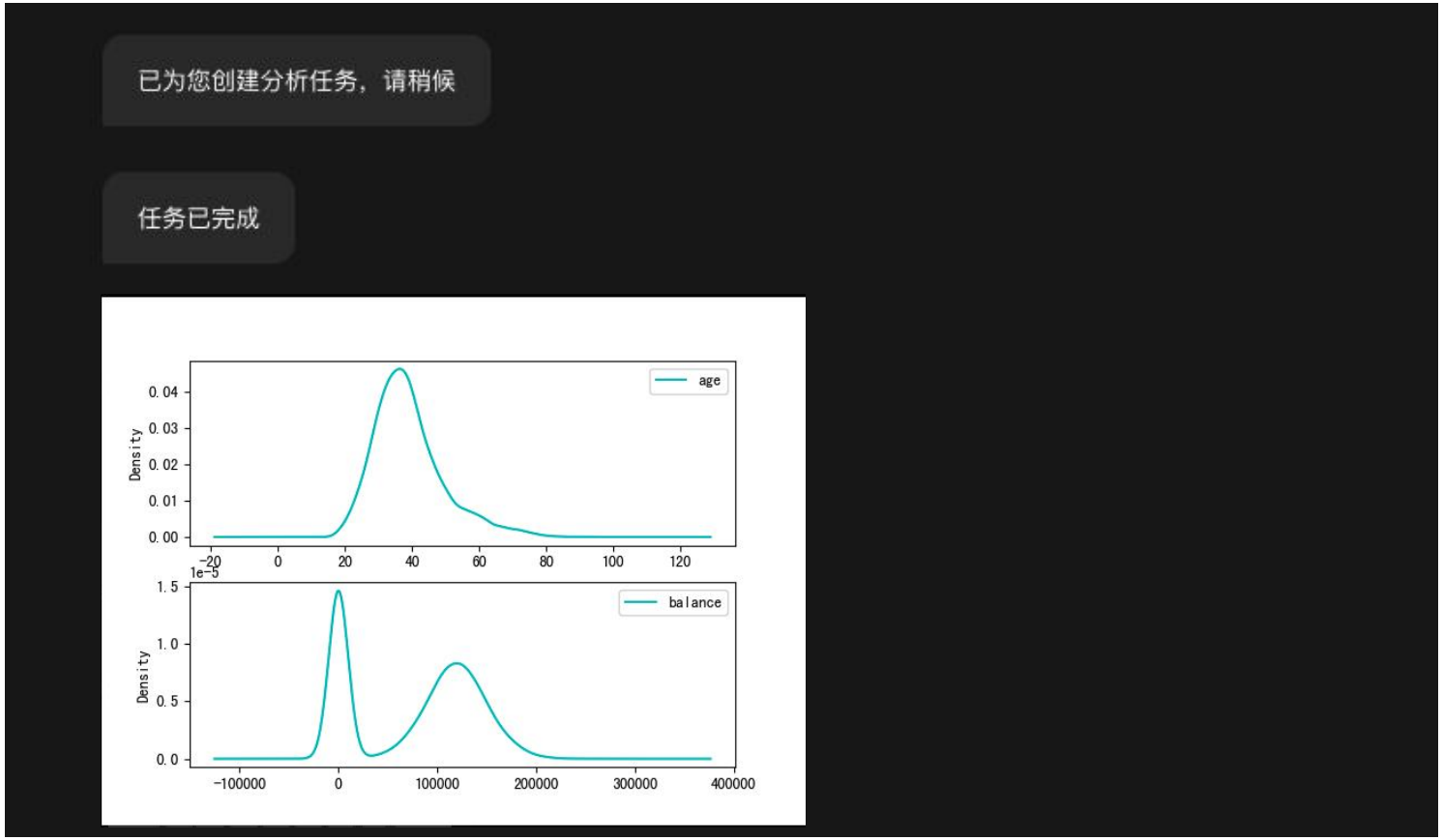
输入指令

TableGPT反馈操作结果：



用户输入数据可视化指令

TableGPT绘制可视化图表



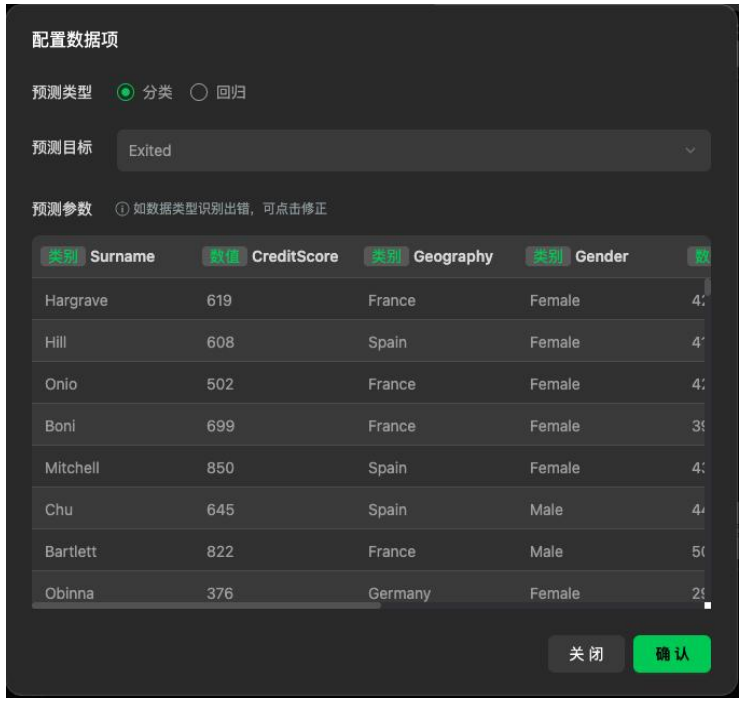
表格自然操作 → 表格可视分析 → **表格自动建模**

用户输入建模指令

用户配置预测目标列

配置

自动训练小模型



表格自然操作 → 表格可视分析 → **表格自动建模**

TableGPT返回模型链接

用户输入待预测文件

TableGPT返回预测结果

TableGPT——一款具备结构化数据分析、作图 、处理、建模等功能的刚性预训练大模型

与国内外同类系统对比

功能	ChatExcel	SheetCopilot	Data-Copilot	TableGPT（我们的工作）
自然语言交互	✓	✓	✓	✓
数据可视化	×	✓	✓	✓
分析及报告	×	×	✓	✓
数据预测	×	×	✓	✓
指令链	×	×	✓	✓
基础模型	未知	API	API	自主可控
指令歧义性检查	×	×	×	✓
私有化部署	×	×	×	✓

PART 03
关键技术

- 如何完成一个落地的LLM
- TableGPT落地之路

如何完成一个

落地的LLM

我们先来看LLM的属性有哪些

柔性
语言能力、创造力

刚性
复杂工作上的准确性

可解释性
工作过程可信与可视程度

可交互性
用户对于工作过程的参与程度

应该如何完成一个落地的LLM?

- 四个评估维度无法同时达到
- 基于场景有所取舍

ChatBot、characterAI:
文本理解与对话能力
- 柔性
- 可交互性

Copilot:
代码生成与解读能力
- 刚性
- 可解释性

TableGPT落地之路

- 刚性
- 可解释性
- 可交互性

LLM的落地之路——以TableGPT为例：

柔性到刚性 -> 从文科生培养到理科生！

柔性LLM

ChatGPT系列等

错了就错了，无所谓！
目的是提高人机交互体验

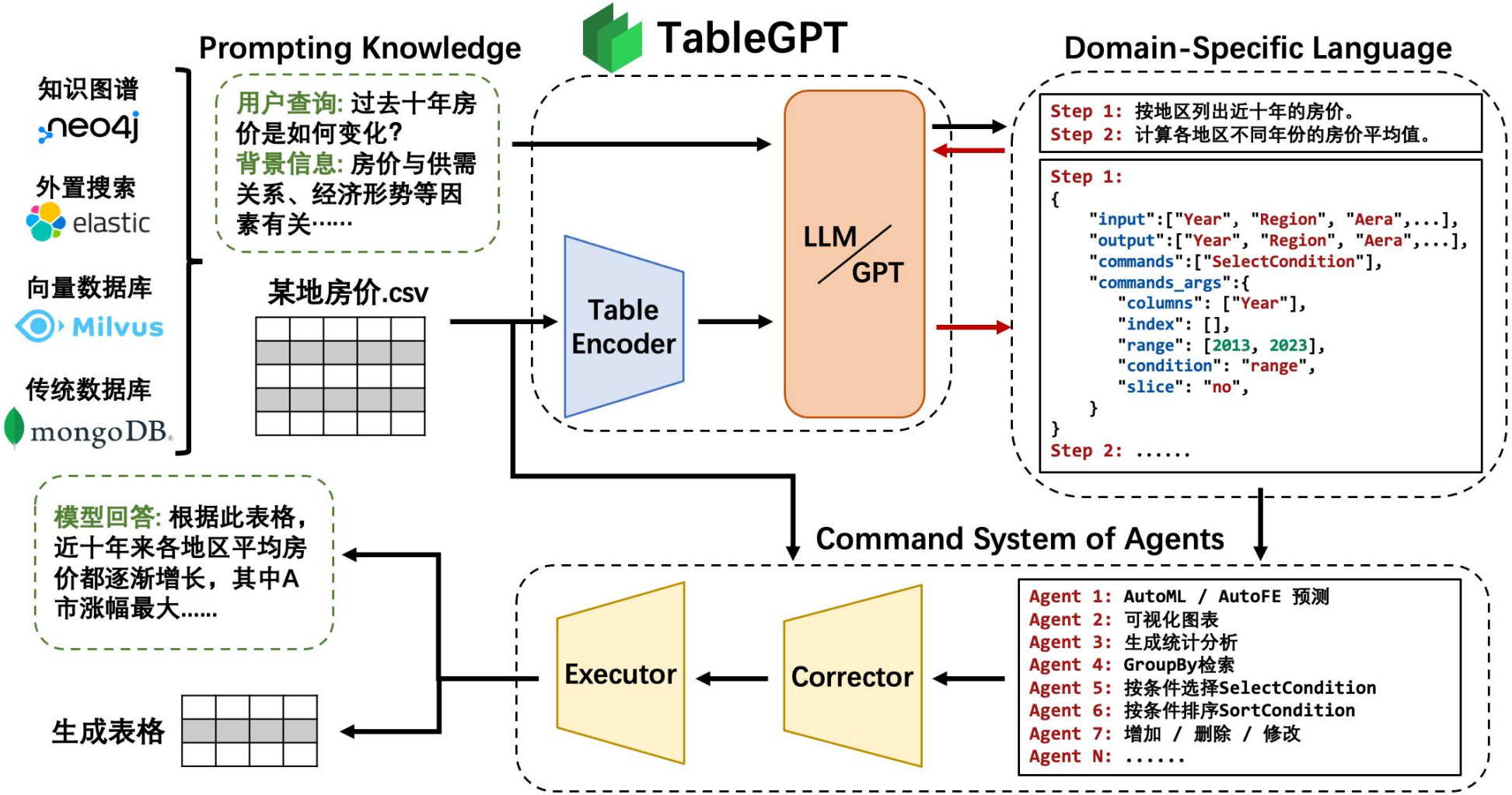
刚性LLM

TableGPT

严格、严谨
不能出错

TableGPT技术：顶层结构图

-> 本质为 LLM-based AGENT!



TableGPT技术：Domain-Specific Language（DSL）

我们设计了一种高效的连结语言——**DSL，作为语言模型与表格操作的刚性媒介**

过去十年房价是如何变化的?

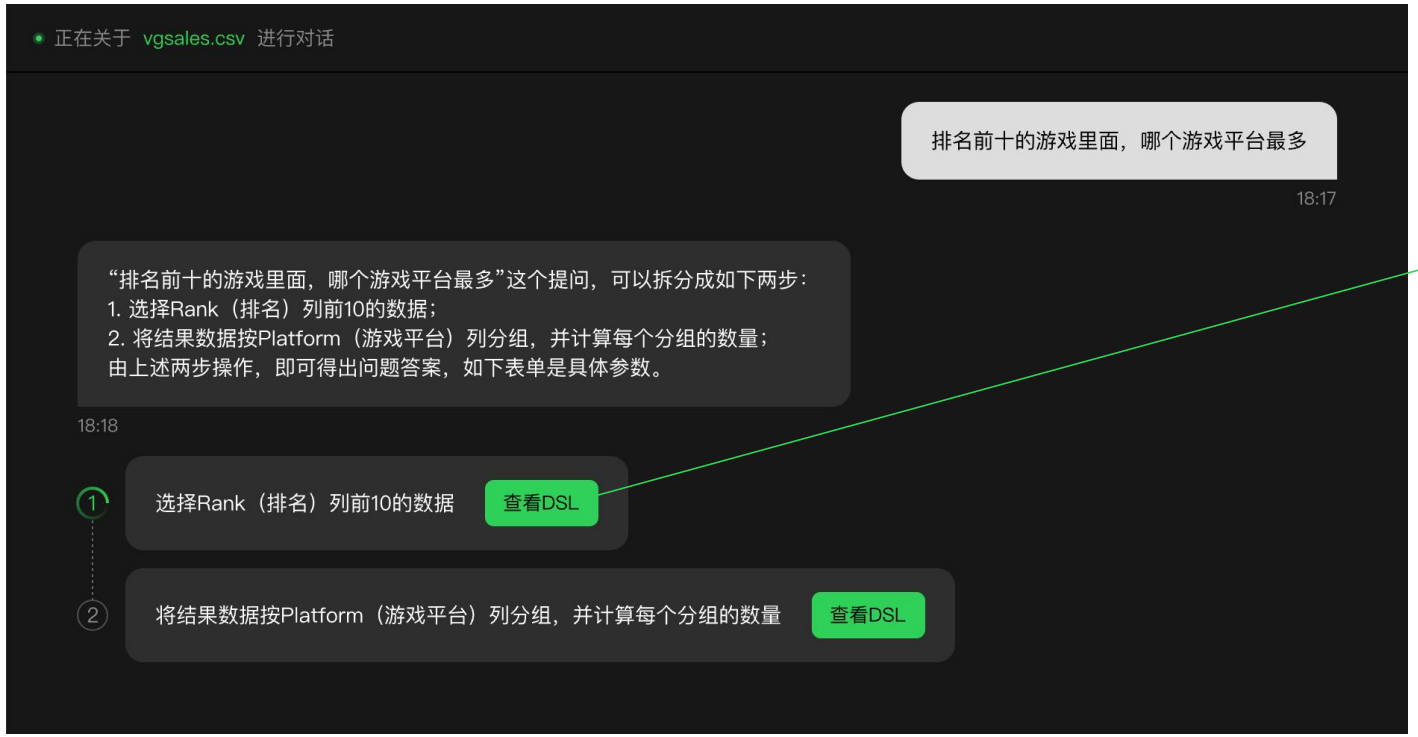
Step 1: 按地区列出近十年的房价。
Step 2: 计算各地区不同年份的房价平均值。

Step 1:
{
输入表格 "input":["Year", "Region", "Rera",...],
输出表格 "output":["Year", "Region", "Rera",...],
组件类型 "commands":["SelectCondition"],
组件参数 "commands_args":{
"columns": ["Year"],
"range": [2013, 2023],
"condition": "range",
}
}
Step 2: ......

- 作为CoT基础
- 参数自动校验
- 可移植性与高通用性

Domain-Specific Language（DSL）

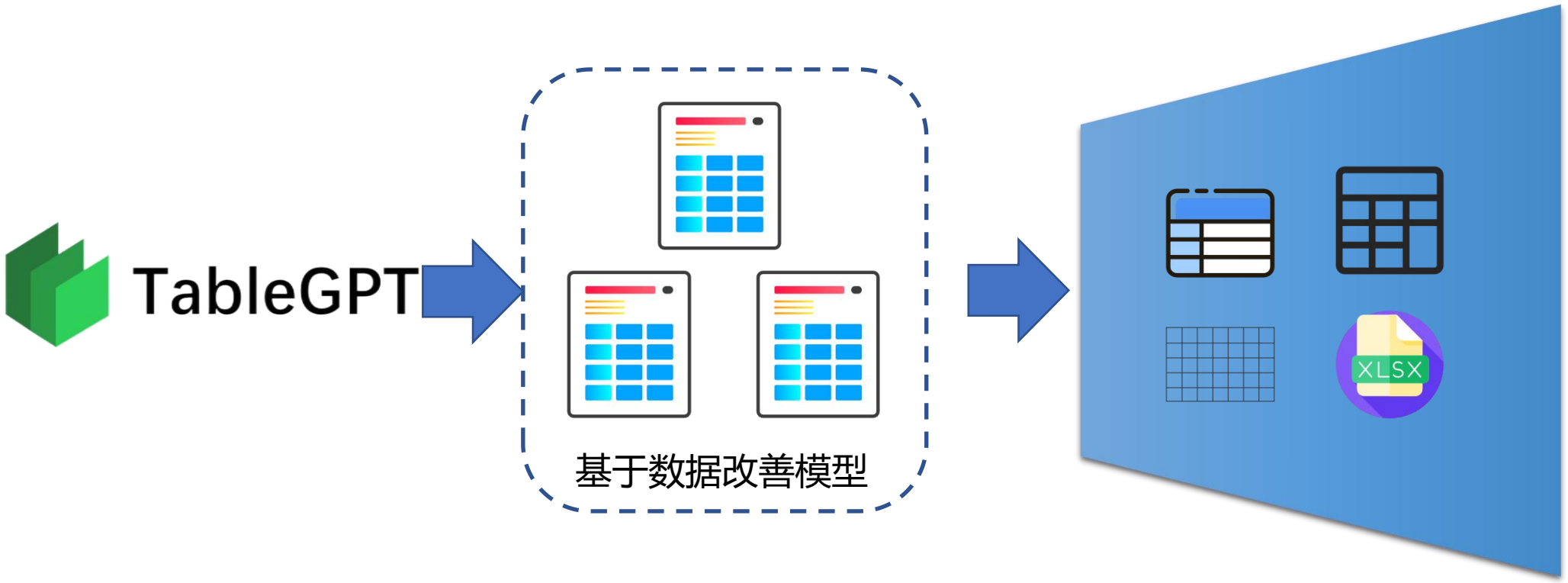
- 可信性。DSL与自然语言的零成本平转将处理过程可视化，提高过程透明度，保障结果的可解释性
- 可交互性。DSL高度结构化的组织形式，便于人类对处理过程进行点对点调控，保障结果的可控性



TableGPT落地之路

- 刚性
  - 低数据资源微调
  - 领域知识微调
  - 多模态对齐
- 可解释性
- 可交互性

如何实现TableGPT的刚性



- 基于领域与任务数据进行微调（Supervised Fine-Tuning）
- 基于领域数据样例召回进行上下文学习（In-Context Learning）

如何实现TableGPT的刚性

基于领域与任务数据进行微调（Supervised Fine-Tuning） 中存在的挑战：

1. **数据收集成本高、标注成本高 -> 低资源指令微调**

输入

用户查询: 过去十年房价是如何变化?

背景信息: 房价与供需关系、经济形势等因素有关……

输出

Domain-Specific Language

Step 1: 按地区列出近十年的房价。

```
Step 1:
{
    "input":["Year", "Region", "Aera",...],
    "output":["Year", "Region", "Aera",...],
    "commands":["SelectCondition"],
    "commands_args":{
       "columns": ["Year"],
       "range": [2013, 2023],
       "condition": "range",
       "slice": "no",
    }
}
```

Step 2: 计算各地区不同年份的房价平均值。

```
Step 2:
{
    "input":["Year", "Region", "Rera",...],
    "output":["Year", "Region", "Price"],
    "commands":["GroupBy"],
    "commands_args":{
       "by": ["Region", "Year"],
       "aggregate_args": {"Price": ["mean"]}
    }
}
```

2. 不同领域之间数据跨度大
3. 模型面向的数据模态多样化

TableGPT刚性微调-低资源指令微调 (LTD Instruction Tuning)

- 基于LLM的Instruction-tuning在逐渐成为定制化模型训练的主流方法，但该方法成本较高：
  - Instruction-tuning需求**Token数量更多**
    以Flan-T5的数据集为例，增加Instruction后token数量变为原来的1.6倍
  - 对LLM的fine-tuning**成本高**
    chatGPT上个版本的模型fine-tuning价格是chatGPT的15倍，未来对GPT-4的微调接口将更贵

Model	Usage
gpt-3.5-turbo	$0.002 / 1K tokens

Model	Training
Davinci	$0.0300 / 1K tokens

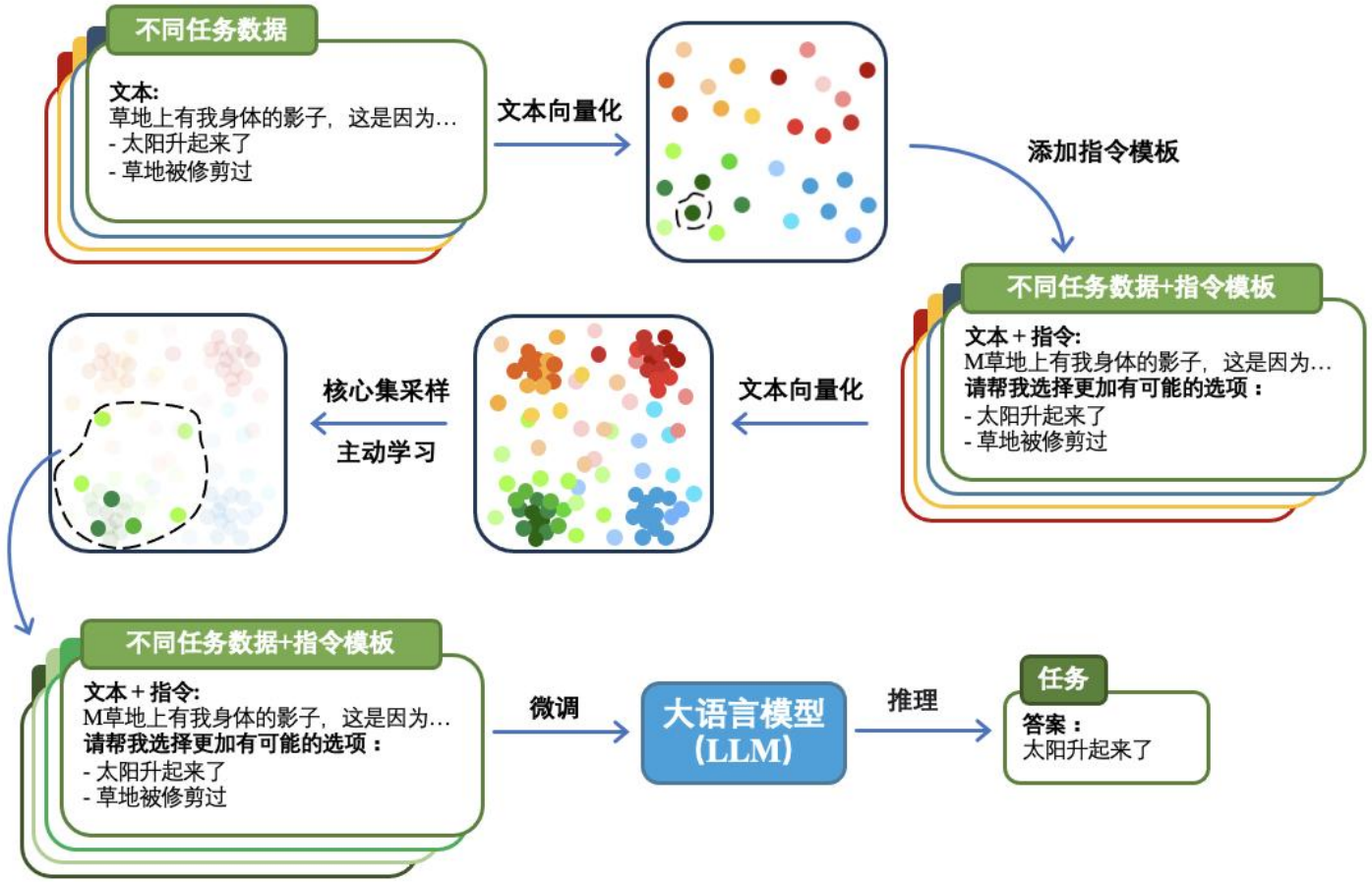
以P3训练数据（0.3B tokens，1.2G）为例，训练的价格约6万人民币
训练一个模型的成本价达到6万/次调参

应对方案：可训练资源限制（Low Training Data）下的Instruction-tuning

Maybe Only 0.5% Data is Needed: A Preliminary Exploration of Low Training Data Instruction Tuning. Arxiv 2023 (under reviewing)

TableGPT刚性微调-低资源指令微调 (LTD Instruction Tuning)

- 基于**无标注数据**进行高质量数据筛选
- 基于**指令**对数据规模进行**扩增**
- 基于**主动学习**或**核心集策略**进行数据采样



Maybe Only 0.5% Data is Needed: A Preliminary Exploration of Low Training Data Instruction Tuning. Arxiv 2023 (under reviewing)

TableGPT刚性微调-低资源指令微调 (LTD Instruction Tuning)

- 在NLI任务上使用原数据集**0.5%**的数据训练出**性能更好**的模型

model	RTE	CB	ANLI R1	ANLI R2	ANLI R3	Avg
P3	**76.17%**	**75.00%**	**44.00%**	**35.70%**	**39.42%**	**54.06%**
Fixed Instruction (10%)	71.11%	66.07%	43.60%	38.90%	42.17%	52.37%
NLI-related (5%)	79.06%	82.14%	60.40%	46.50%	46.67%	62.95%
NLI coreset (0.5%)	**74.73%**	**73.21%**	**49.60%**	**41.90%**	**43.75%**	**56.64%**
Vanilla Model (0%)	54.51%	41.07%	33.40%	33.40%	33.58%	39.19%

- 在P3整个数据集上使用原数据集**0.3%**的数据训练出**性能更好**的模型

Methods	Number of Training Tokens	NLI					SC			CR		WSD	AVG
		ANLI R1	ANLI R2	ANLI R3	CB	RTE	COPA	HellaSwag	StoryCloze	WSC	Winogrande	WiC	
T5 + LM	pre-trained	32.89	33.76	33.82	34.34	53.03	54.88	48.16	27.00	54.09	50.65	50.03	42.97
T0-3B	100% (250B)	33.84	33.11	33.33	45.36	64.55	72.40	27.29	84.03	65.10	50.97	50.69	50.97
ours	0.3% (60M)	34.04	33.73	35.56	72.22	56.77	72.94	29.86	87.44	57.17	54.43	50.08	52.41

Maybe Only 0.5% Data is Needed: A Preliminary Exploration of Low Training Data Instruction Tuning. Arxiv 2023 (under reviewing)

如何实现TableGPT的刚性

基于领域与任务数据进行微调（Supervised Fine-Tuning） 中存在的挑战：

1. 数据收集成本高
2. **不同领域之间数据跨度大** -> 快速领域微调

不同领域高频词有明显差异

金融领域高频词：
均线、资金流入、委比、振幅、换手率、成交量、成交额、股价、分时指标、强势股、涨停股、流通股本、总市值、流通市值、流通比例、股东户数、户均持股数、增减持、分红、上市天数、销售毛利率...

交通领域高频词：
正点率、准点率、总驶里程、运营里程、客流量、班次、激增 、满载率、首末班、均匀化、迫降、失速、地勤、系统卡阻、曲柄、整流罩、流通、调度...

3. 模型面向的数据模态多样化

TableGPT刚性微调-快速领域微调 （FAvDP）

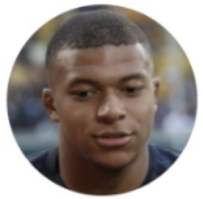
FAvDP技术

1. 识别出新语料的实体并连接到知识库得到<实体，知识>对

2. 将<实体，知识>对转化为提示文本，基于词序列、位置序列和分段序列构建损失函数进行训练

原文本: 基利安·姆巴佩成为自 1966 年以来第一位在世界杯决赛中上演帽子戏法的球员。

实体识别:基利安·姆巴佩成为自1966年以来第一位在世界杯决赛中上演帽子戏法的球员。



基利安·姆巴佩(Q21621995) 实例：
人类
国籍：法国
职业：足球运动员
简介：法国足球运动员

帽子戏法 (Q123086)实例：
足球术语
描述：在一场比赛中完成三次进球

Prompt处理结果

基利安·姆巴普（是法国足球运动员）成为自 1966 年以来第一位在世界杯决赛中上演帽子戏法（是在一场比赛中完成三次进球）的球员。

Fast Adaptation via Prompted Data: An Efficient Cross-Domain Fine-tuning Method for Large Language Models. EMNLP 2023 (under reviewing)

TableGPT刚性微调-快速领域微调 （FAvDP）

性能优势

在7B模型（LLaMA）
上实现3%性能提升

Methods	Settings	Injection		Datasets	
		Knowledge	Label	MedQA-USMLE	MedMCQA
ChatGPT (OpenAI, 2022)	Zero-shot	-	-	57.0	44.7
OPT-6.7b (Zhang et al., 2022)		-	-	27.34	28.64
Galactica-6.7b (Taylor et al., 2022)		-	-	30.16	30.48
LLaMA-7b (Touvron et al., 2023)		-	-	27.1	24.3
Standardized SFT	PEFT*	-	-	27.34	32.37
Adaptation (Gururangan et al., 2020)		✗	✗	27.73	35.81
FAvPD on LLaMA-7b (Our Method)		✓	✗	31.26	40.59
FAvPD on LLaMA-7b (Our Method)		✗	✓	29.85	39.44
FAvPD on LLaMA-7b (Our Method)		✓	✓	**32.68**	**42.58**

Table 3: The Performance of FAvPD on Question Answering Task. Accuracy Score Reported. PEFT* indicates we apply LoRA (Hu et al., 2022a) tuning (100M trainable parameters) to LLaMA-7b Architecture.

训练速度优势

经过领域注入模型会
提前15~20%左右的
时间到达性能最高点

运算成本优势

注入知识量需求大大减少
算力需求大大减少

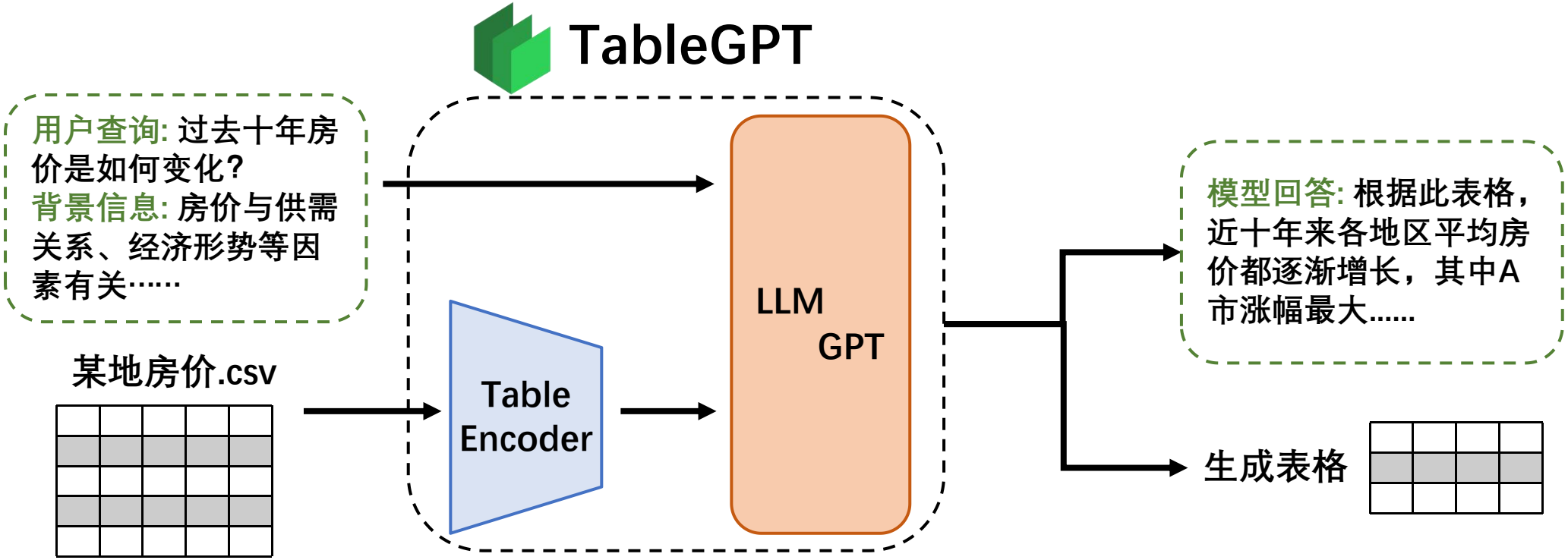
Model	Statistics of Corpus and Knowledge	Computational Consumption
ERNIE (Zhang et al., 2019)	4500M subwords, 140M entities	8 NVIDIA 2080Ti GPUs for 24 hours
CoLAKE (Sun et al., 2020)	26M examples, 3M entities	8 32G NVIDIA V100 GPUs for 38 hours
K-Adapter (Wang et al., 2021a)	5.5M sentences, 1M examples	4 16G NVIDIA V100 GPUs for 4 days
LUKE (Yamada et al., 2020)	3.5B words, 11M entities	16 NVIDIA Tesla V100 GPUs for 30 days
FAvPD(Ours)	10K sentences, 26K entities, 22K labels	1 NVIDIA 2080Ti GPU within 1 hour

Fast Adaptation via Prompted Data: An Efficient Cross-Domain Fine-tuning Method for Large Language Models. EMNLP 2023 (under reviewing)

如何实现TableGPT的刚性

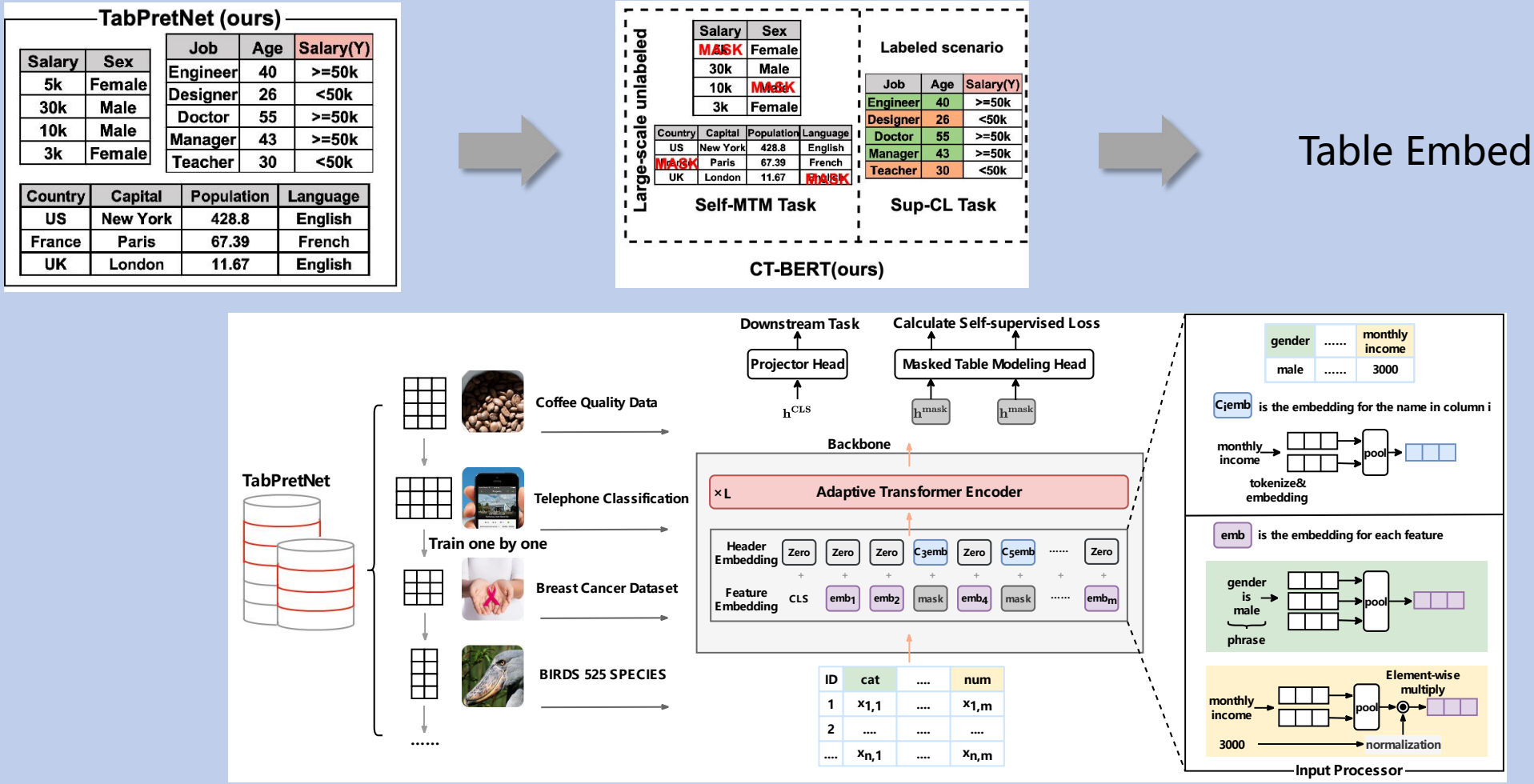
基于领域与任务数据进行微调（Supervised Fine-Tuning） 中存在的挑战：

1. 数据收集成本高
2. 不同领域之间数据跨度大
3. **模型面向的数据模态多样化 -> text-table模态对齐**



TableGPT刚性实现：text-table模态对齐

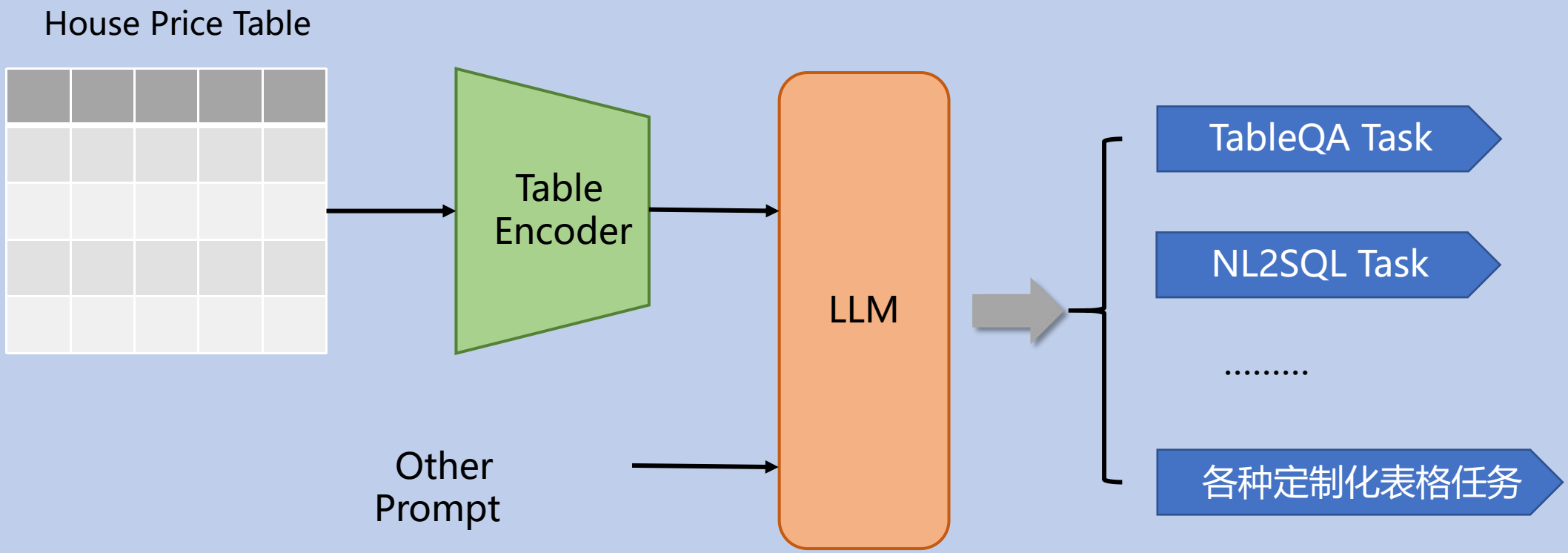
对表格编码后的向量信息可以用来做什么？　　1、解决常见的表格预测任务



CT-BERT: Learning Better Tabular Representations Through Cross-Table Pre-training. VLDB 2023 (under reviewing)

TableGPT刚性实现：text-table模态对齐

对表格编码后的向量信息可以用来做什么？ 2、提取表格的全局信息，帮助LLM更好的理解表格



TableGPT刚性修改：低代码可交互AutoML平台

对TableGPT输出指令参数进行快捷修改

数据

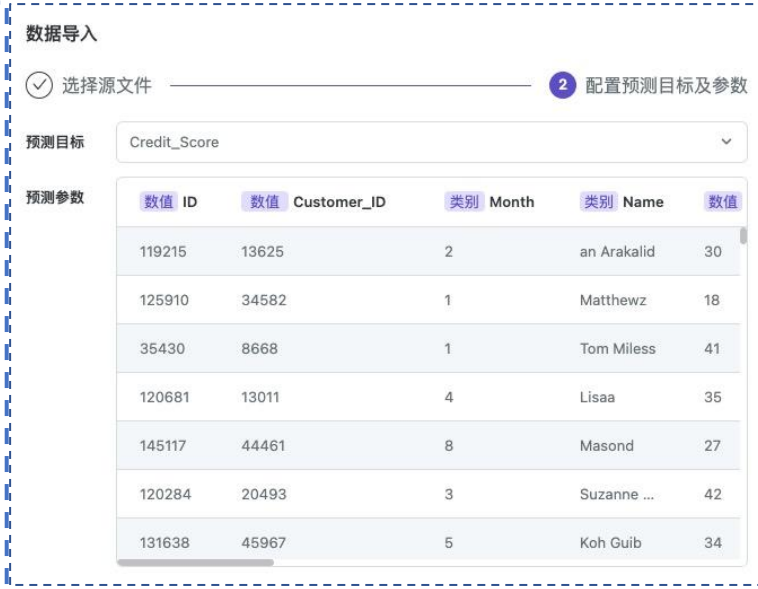
属性	用户1	用户2	...
姓名	张三	李四	...
贷款金额	6000	缺失	...
年收入	100000	200000	...
风险评估	良好	不良	...

支持

AutoML交互平台

自动建模

RandomForest



流程搭建完毕，自动开始优化训练

准确率：83.7%

TableGPT落地之路

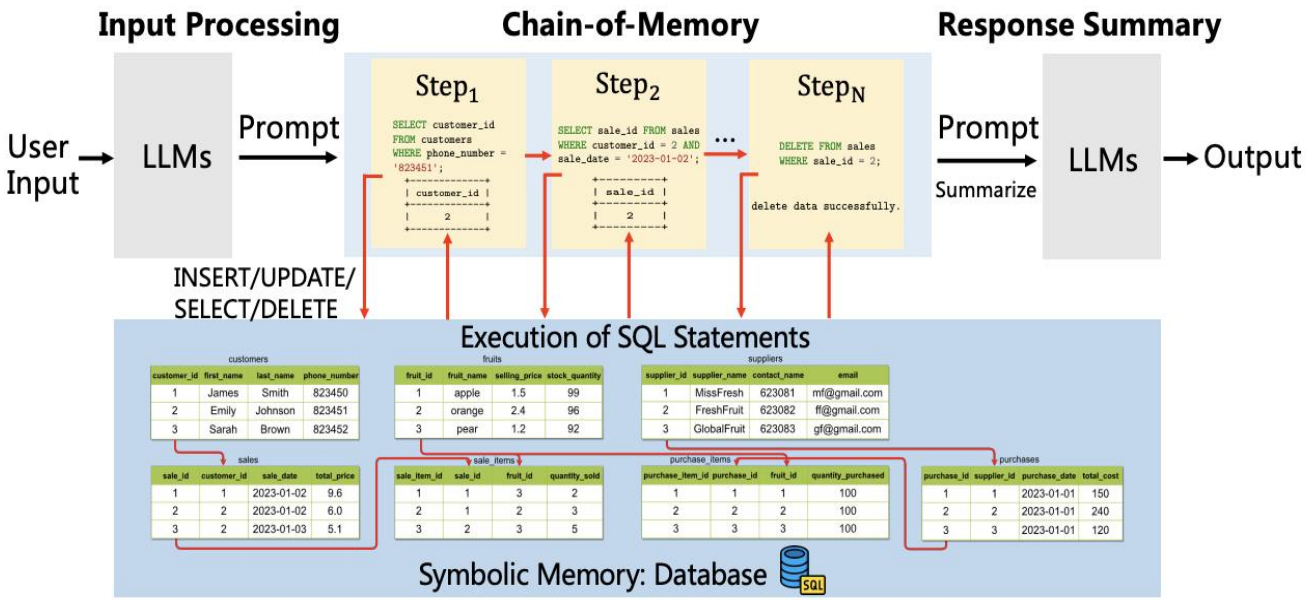
- 刚性
- **可解释性**
  - **DSL思维链**
- 可交互性

TableGPT可解释性：Chain-of-DSL（DSL范式下的CoT）

- 将复杂问题拆分为多个中间步骤，每个中间步骤涉及一个或多个 DSL

Model	Easy	Hard	All	Accuracy
ChatGPT	10/15	1/35	11/50	22%
ChatDB (ours)	13/15	28/35	41/50	82%

- 提升复杂推理能力
- 高度可解释性的询问链



ChatDB: Augmenting LLMs with Databases as Their Symbolic Memory. Arxiv 2023

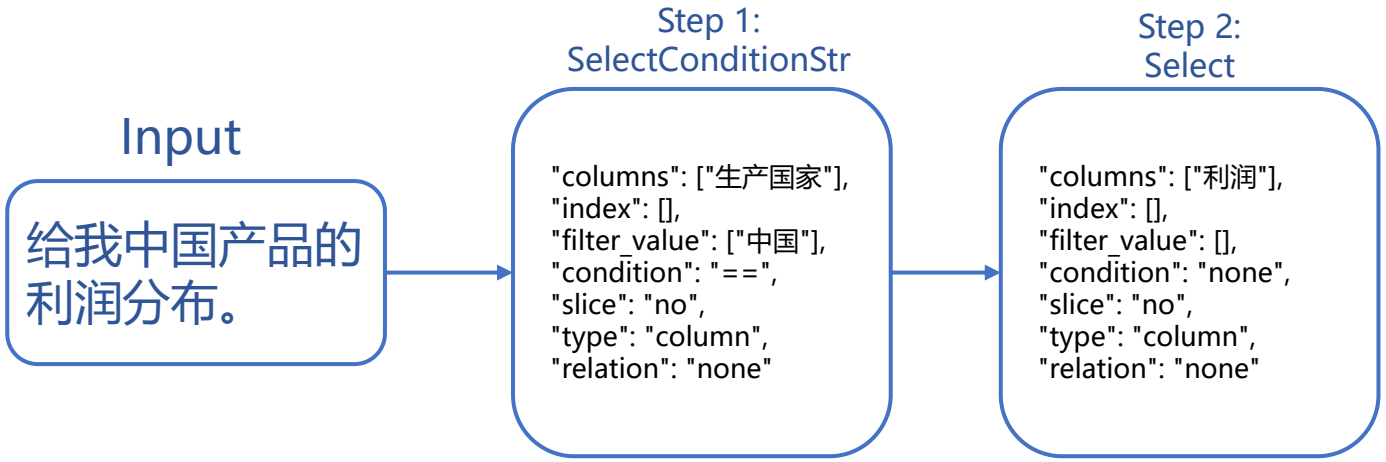
TableGPT可解释性：Chain-of-DSL（DSL范式下的CoT）

Table Format

'numeric_cols': ['产品编号', '成本', '利润', '市场份额', '广告费用', '研发投入', '产品重量', '产品尺寸', '产品时长']

'category_cols': ['产品类型', '品牌', '包装类型', '生产国家', '销售区域', '目标市场', '销售渠道', '主要竞争对手', '消费者群体', '产品特点']

- **将复杂问题转化为多个中间步骤，每个中间步骤涉及一个或多个可解释的DSL，降低复杂度**
- **有效提升TableGPT对复杂、模糊指令的推理能力**



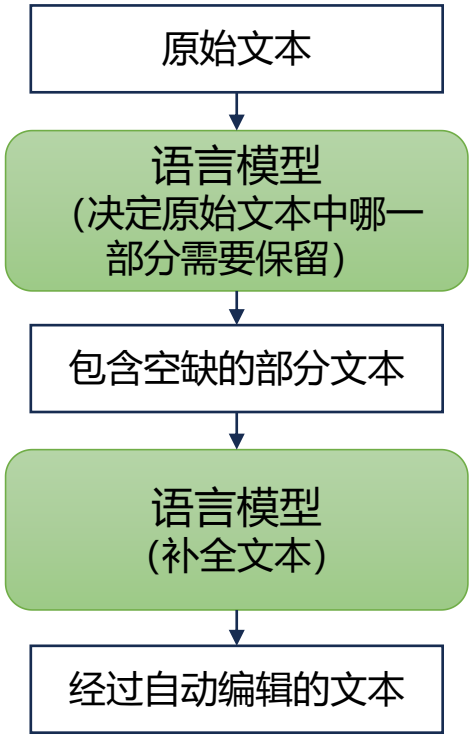
TableGPT落地之路

- 刚性
- 可解释性
- **可交互性**
  - **自动化prompt**
  - **可编辑prompt**

TableGPT可交互性：可交互的prompt编辑工程

一个常见的场景：当前文本和目标高度近似...但仍有优化空间！

-> 一个细粒度可控的自动化文本优化器可以高效修改和完善文本！



更好的生成效果

当原始文本和目标文本有大量重合时，基于编辑的范式可以大量保留原始文本，降低生成难度。

细粒度可控

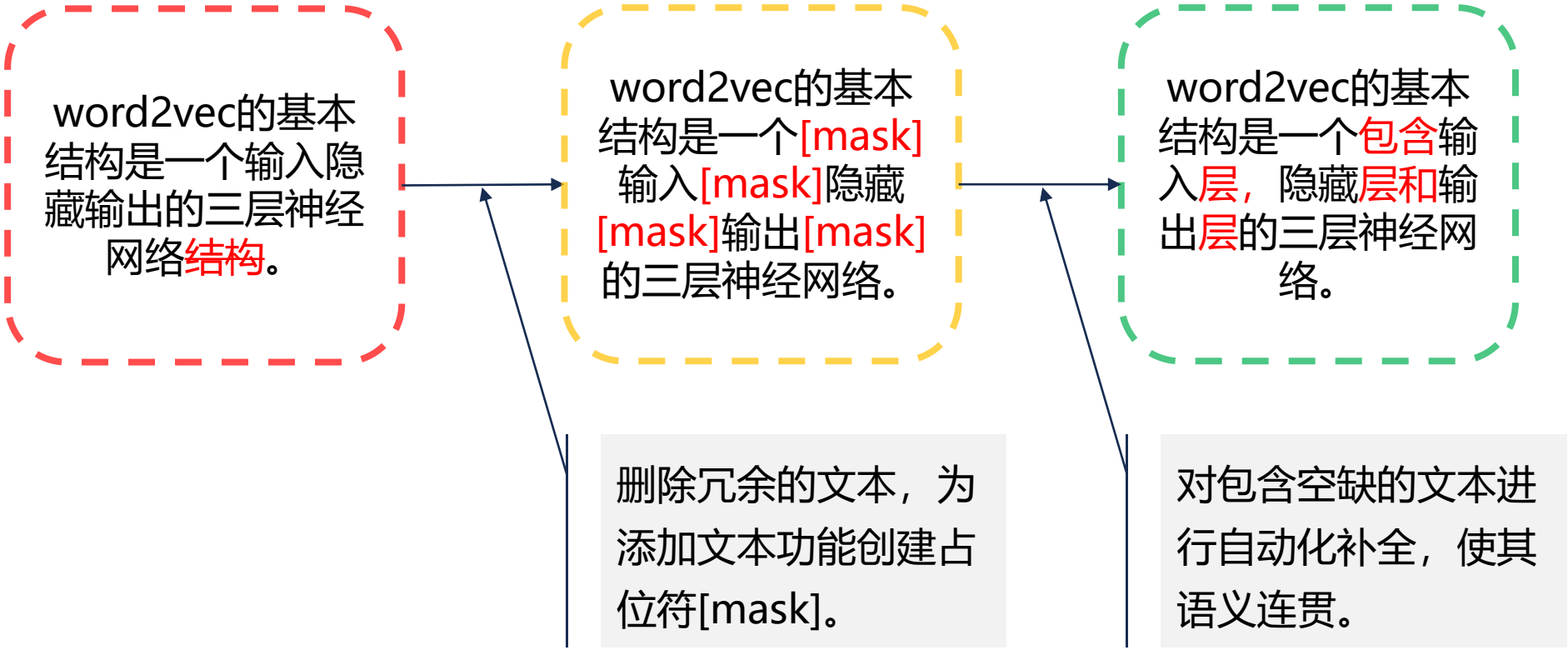
用户在编辑过程中的每一步都可以给出细粒度的监督信号。比如指定原始文本的哪一部分不应修改。

并行生成潜力

编辑动作的相互依赖较小，可以通过非自回归语言模型进行并行生成，节省推理时间。

TableGPT可交互性：可交互的prompt编辑工程

比如，编辑范式可被用于输入文本的纠错和优化



这一范式将用于TableGPT的后续优化，包括模板优化，中间结果优化，生成结果的细节处理等。

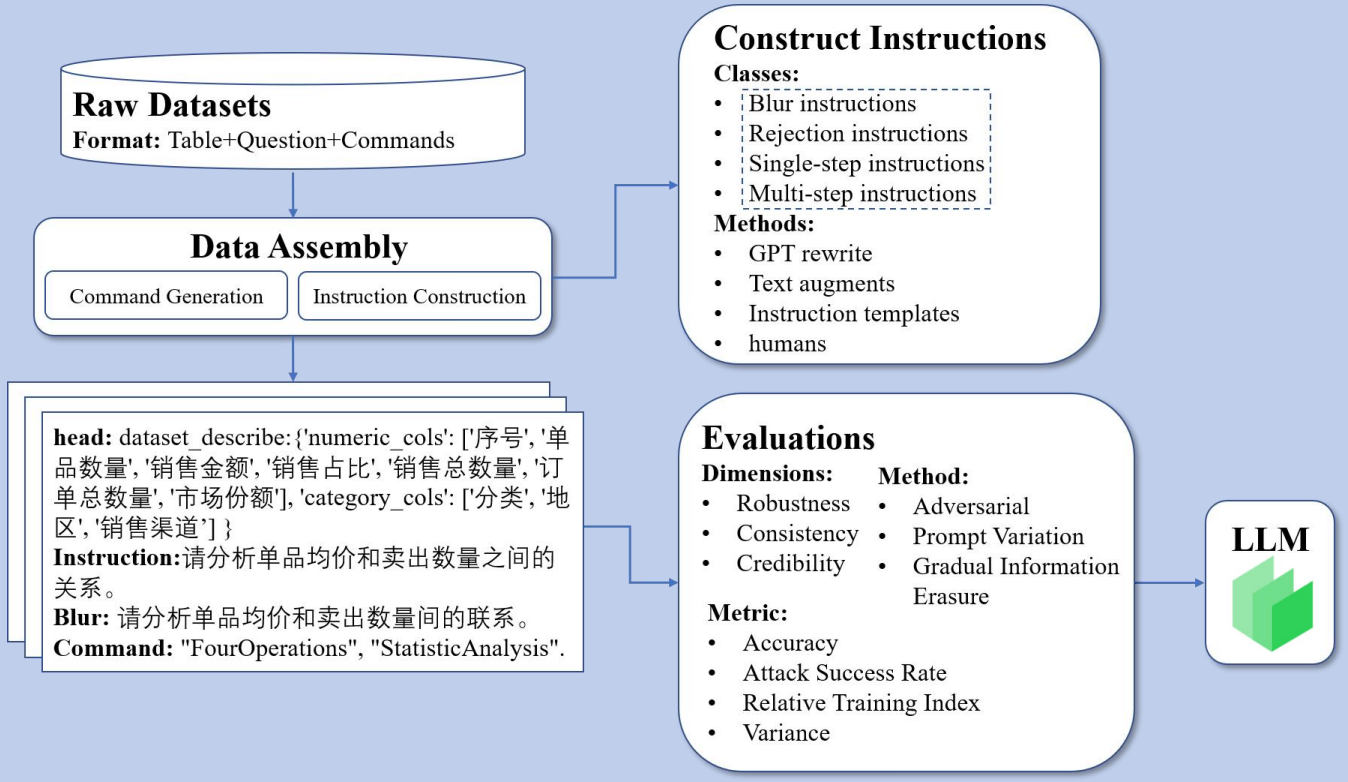
TableGPT -specific评估体系

细粒度的评估

- 我们设计4种功能指令形式，以半自动化的框架驱动指令构造
- 基于以上指令，我们对TableGPT进行多维度的评估，选择真正“好”的模型

低成本的功能验证

- 训练同时进行异步评测（loss无法体现模型优劣），用尽可能少的评估数据选择最有效的模型



Assessing Hidden Risks of LLMs: An Empirical Study on Robustness, Consistency, and Credibility

TableGPT -specific评估体系

可信评估数据筛选

设计RTI-Index，筛选可信数据，去除被“记忆”的评估样本

Algorithm 1 Calculate RTI $\mathbf{R}_D$ of dataset D

Input:
$D = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$: dataset
$g(x, \rho)$: auto-attacker on $\mathbf{x}$ with probability ρ
$f(x, \theta)$: model output on $\mathbf{x}$ with parameter θ
Output:
$\mathbf{R}_D$: RTI score of dataset D

for $\mathbf{x}_i$ in D **do**
  $\rho = 0.1$
  $\mathbf{x}'_i = g(\mathbf{x}_i, \rho)$
  while $f(\mathbf{x}_i, \theta) = f(\mathbf{x}'_i, \theta)$ **do**
    $\rho = \rho + 0.1$
    $\mathbf{x}'_i = g(\mathbf{x}_i, \rho)$
  end while
  $r(\mathbf{x}_i) = \rho$
end for
$\mathbf{R}_D = \mathbb{E}(R)$ where $R \sim r(\mathbf{x}), \mathbf{x} \in D$
return $\mathbf{R}_D$;

对抗鲁棒性

通过word、character、visual三种level的对抗攻击评价鲁棒性

- Word：delete、insert、replace
- Character：delete、insert、replace
- Visual：视觉相似字符替换

模糊指令一致性

构造同目标、表达模式不同形式的输入instruction，增强TableGPT含义处理能力

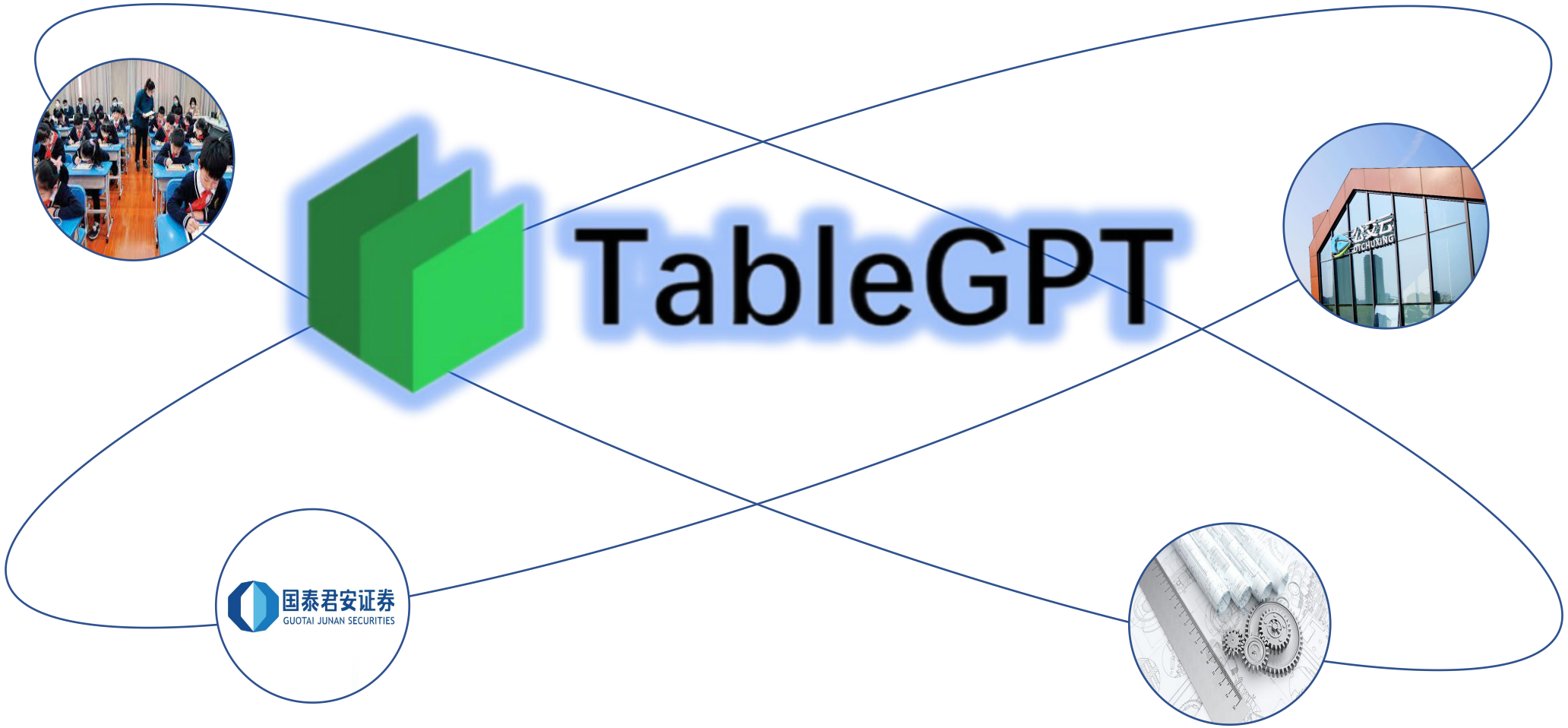
各家发行商各类型游戏都卖的怎么样?
各家发行商最擅长的游戏类型销量是多少?

构造方式：
- 手动构造
GPT生成同义指令
- 嵌入instruction模版

Assessing Hidden Risks of LLMs: An Empirical Study on Robustness, Consistency, and Credibility

PART 04
TableGPT落地案例

TableGPT落地案例



目前TableGPT已经在金融、交通、工业等多个领域实现应用落地

案例1-金融领域

用户

国泰君安证券 GUOTAI JUNAN SECURITIES

业务痛点

智能选股	智能选基金	信贷风控	产品精准推荐
传统的选股方法容易忽略市场情感和短期新闻影响，需要更智能的方式来辅助决策。	客户选择基金时需要大量的信息筛选和比较，传统方法难以从海量信息中准确提取有效信息，需要更智能的推荐方式。	传统信贷风控往往基于静态数据和模型，难以灵活应对不同客户情况，需要更智能的方式来分析客户风险。	传统的产品推荐往往基于静态的客户信息，无法捕捉客户的动态变化和情感需求，需要更智能的方式来进行个性化推荐。

案例1-金融领域解决方案

为保证在金融领域的选股、选基、信贷风控等精准定量场景下准确输出，TableGPT内置了多种功能组件，并通过工程化校验保证系统本身的“刚性”和可控性。

智能选股

输入问题：给我XX股的近期的换手率、振幅以及流通股本等关键指标的动态演变情况

TableGPT

根据数据，XX股近期换手率、振幅、流通股本从高到低数据如下：....

XX近期演变情况.csv

智能选基

输入问题：XX投资基金c类的净值管理多少只基金，有几只赚钱?

TableGPT

根据数据，XX投资基金c类的净值管理X只基金，有X只赚钱

XX投资基金c类数据.csv

案例2-公共交通领域

公交云 DTCHUXING 杭州市公共交通云科技有限公司

40+

累计服务40+公交及地铁城市

10W

日均服务公共交通100000辆+

2000W

日均服务乘客2000万+

对外的需求痛点

用户诉求千奇百怪，若以传统的机器人客服的实现方式，将用户的提问划分为各种意图并一一对应实现，将耗费大量的人力物力，且永远会发现有新的用户提问未满足

对内的需求痛点

数据分布在多张表中，每次想要提取数据分析，都需要需求方找到数据分析师，数据分析师再找到后端询问各字段藏在哪个表里，然后再写SQL将数据按照要求一一提取，过程冗长，成本极高

案例2-公共交通领域解决方案

将TableGPT连接至公交云数据库，自行理解提问者意图并将数据从各表中提取出，在查班次详情、查运营指标、半自动化调度等公交业务场景下对业务效率有显著提升。

查班次详情

输入问题：哪10条线路的首末班准点率最高，且从高到低排序

TableGPT

根据数据，1路-98%、2路-97%。。

线路准点率.csv

查运营指标

输入问题：找下X线路的近30天的所有营运指标

TableGPT

根据数据，已提供X线路的运营指标报表

X线路运营指标报表.csv

Going beyond Language – 3D AIGC

缺陷

目前的图像类/3D资产类的生成式的工作，在生成的内容的可控性上式欠缺的

- 测试标准
- 物理属性
- 拓扑优化

解决路径

线下可用3D资产

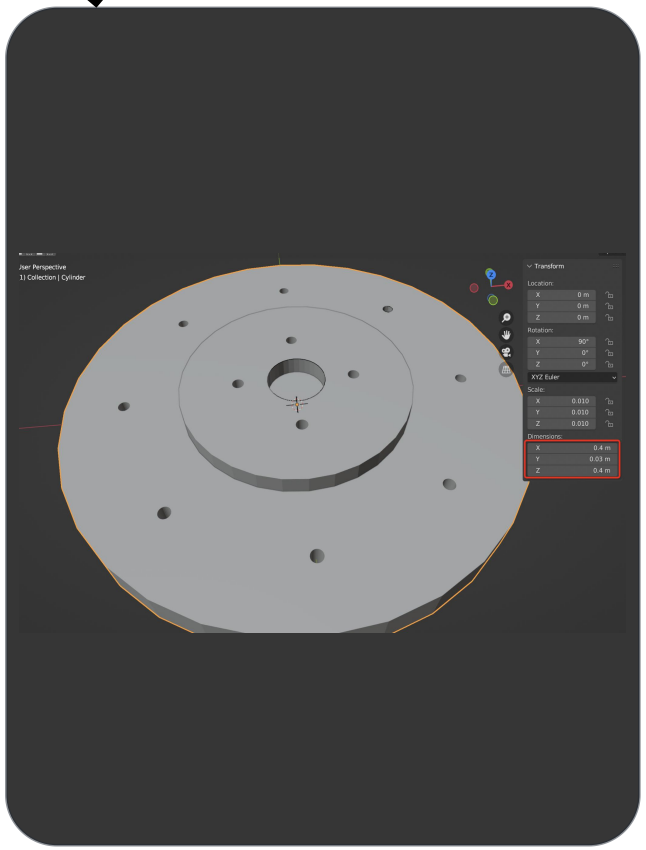
工业多模态数据 + **大语言模型** = **CAD自动化脚本**

Going beyond Language – 3D AIGC

生成一个刹车盘，刹车盘的具体参数和描述如下：
外圈参数和描述：直径40厘米，厚度3厘米，外圈上有8个直径为1厘米的圆形排气孔均匀排列；内圈参数和描述：直径20厘米，厚度6厘米，内圈有4个直径为1厘米的圆形排气孔均匀排列；中通参数和描述:圆柱形镂空，直径5厘米，需要符合欧洲检测标准

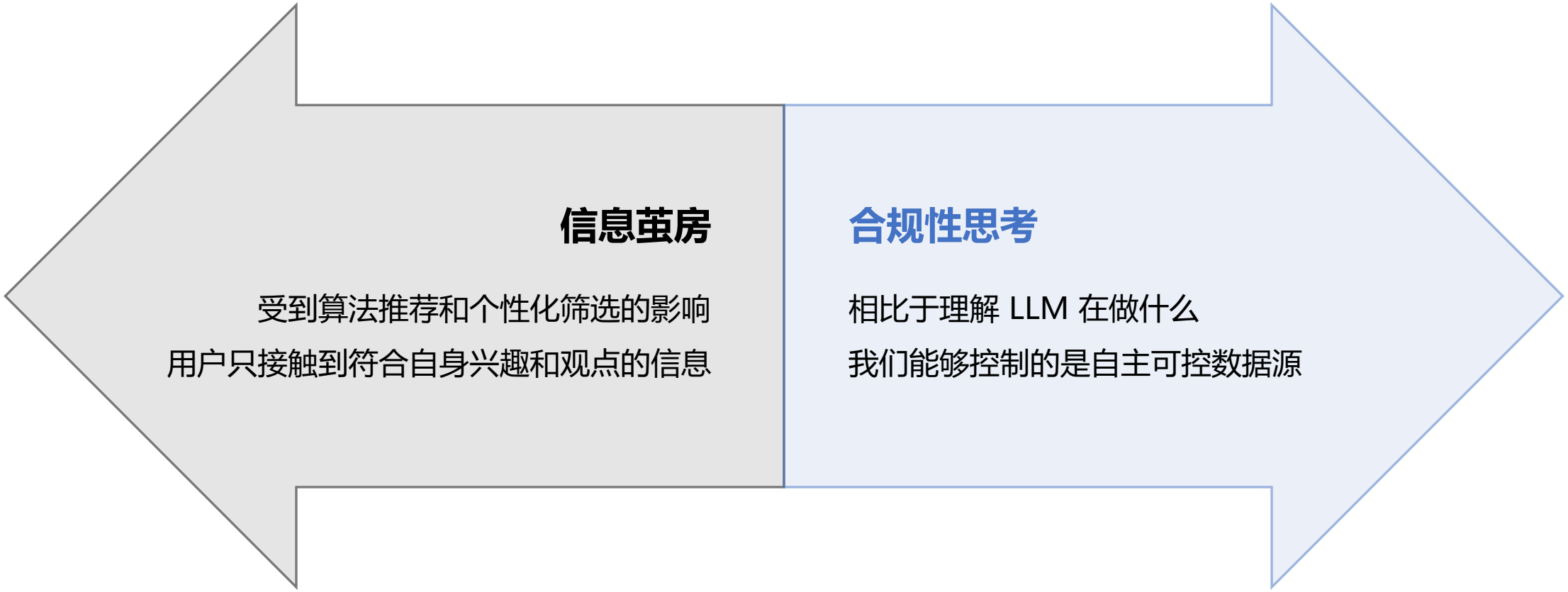
```
import bpy
import math

def create_brake_disk(export_path):
    # Create outer ring
    bpy.ops.mesh.primitive_cylinder_add(radius=20, depth=3, location=(0, 0, 0))
    outer_ring = bpy.context.active_object

    # Scale outer ring to specified diameter
    outer_ring.scale[0] = 2  # Diameter of 40cm
    outer_ring.scale[1] = 2  # Diameter of 40cm

    # Create air holes for outer ring
    air_holes_outer = []
    for i in range(16):
        angle = 2 * math.pi * i / 16
        x = 18 * math.cos(angle)
        y = 18 * math.sin(angle)
        bpy.ops.mesh.primitive_cylinder_add(radius=1, depth=3, location=(x, y, 0))
        air_hole = bpy.context.active_object
        air_holes_outer.append(air_hole)

    # Apply boolean modifiers to outer ring to create air holes
    for hole in air_holes_outer:
        mod = outer_ring.modifiers.new("Boolean", 'BOOLEAN')
        mod.operation = 'DIFFERENCE'
        mod.object = hole
        bpy.ops.object.modifier_apply({"object": outer_ring}, modifier=mod.name)

    # Create inner ring
    bpy.ops.mesh.primitive_cylinder_add(radius=10, depth=6, location=(0, 0, 0))
    inner_ring = bpy.context.active_object
```

.....



PART 05
总结与展望

从TableGPT看LLM未来：底座模型的重要性



从TableGPT看LLM未来

一个通用大模型到领域大模型的技术间隔，即LLM 的落地应用层面

- 这个事儿不容易，不是单纯拿一些领域数据精调就结束了
- 领域数据的清洗 和 面相特定任务的特定指令数据精调是个艺术
- 尊重领域数据的知识注入 -> 技术问题
- 尊重产品、工程和商业价值
- 重点要关注刚性的提升。
- Going beyond language！ -> 尊重业务本身的工作流

AiDD AI+ 软件研发数字峰会
AI+ software Development Digital summit

THANKS