2025/12/20

视觉原生教育助教系统

Vision-Native Pedagogical Agent

汇报人：rancan

01

趋势与选题

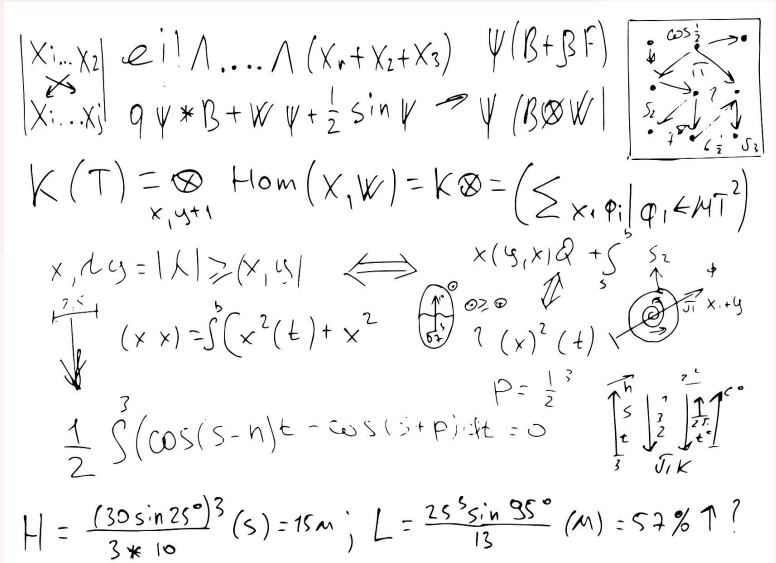
1. 场景背景：虚拟人“只讲不答”

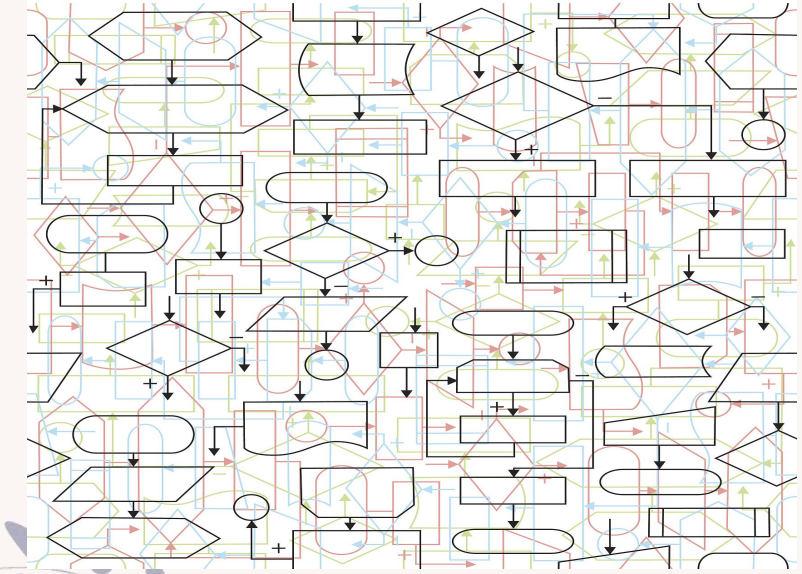
现状：课件→讲解视频 的生产流程（离线、需要人工兜底）

缺口：传统教育=讲(Lecture)+问(Question)+答(Answer)

当前数字人多停留在“讲”，缺少课后实时问答能力

2. 核心挑战：视觉富文档的解释生成瓶颈





场景

视觉富文档（Visually-Rich Documents）

特征：结构化知识密集（复杂公式、板书推导、逻辑流程图）。

痛点：传统 VLM 难以处理符号密集（Symbol-Dense）且 逻辑链长（Long Logical Chains）的内容。

挑战

开放式幻觉（Open-ended Hallucination）：

模型在缺乏约束时倾向于“看图编故事”，利用内部训练数据覆盖视觉证据，从而出现一本正经地胡说八道，公式下标错误，推导逻辑断裂。

不可审计（Lack of Auditability）：

生成的解释无法精确回溯到像素级的视觉证据（如“第几步推导”、“哪个变量定义”）。

https://arxiv.org/abs/2510.24476

3. 目标：从“看图说话”到“可验证证据编排

•**核心理念：**不追求通用的“全知”助教，而是构建一套跨模态精准对齐与可审计的解释协议。

多模态输入

课件页面
学生问题
对话历史

PGP智能体核心

视觉检索
结构化解析
对抗性自检

可信教学解释

回答文本
证据指针
不确定性声明

传统 VQA

看图说话
痛点：开放式生成，易产生幻觉，无法溯源

方案

可验证证据编排
核心：跨模态精准对齐 + 像素级可审计解释

02

模块设计

1. PGP Agentic Workflow： 闭环式证据编排架构

意图分流
目标定义

Light Router

生成类型化结构笔记，
构建闭世界地基

Visual Perception

基于硬约束，
执行证据绑定

Pedagogical Generator

执行软性合规检查与一致性校验，驱动增益循环

Pedagogical Evaluator

Light Router

意图分流

Path A：纯概念题 → Text RAG（降噪）。

Path B：视觉推导题 → Visual RAG（聚焦）。

目标定义

提取 Query 约束：Target = {Symbols} ∪ {KeyPhrases}

作用：为后续的“增益重读”画出靶心。

https://github.com/Alibaba-NLP/ViDoRAG

阿里 EMNLP

Visual Perception

闭世界证据库

模式约束解析

输入： 非结构化原始像素

输出： 三个正交证据表

- Symbols: 变量与物理含义
- Steps: 细粒度逻辑原子
- Definitions: 属性约束

设计意图： 确立“词汇表”，强制 Generator 只能在此集合内“填空”

拓扑分块与索引

语义分块

将连续 N (3-5) 个 Step 封装为一个语义块。建立 Sequence_ID 索引

设计意图： 为 Module 4 的“软性合规检查”提供物理抓手（检测引用是否跨越了 Block 边界）

增益驱动求精

触发条件： 当 Module 4 反馈“证据不足”时启动。

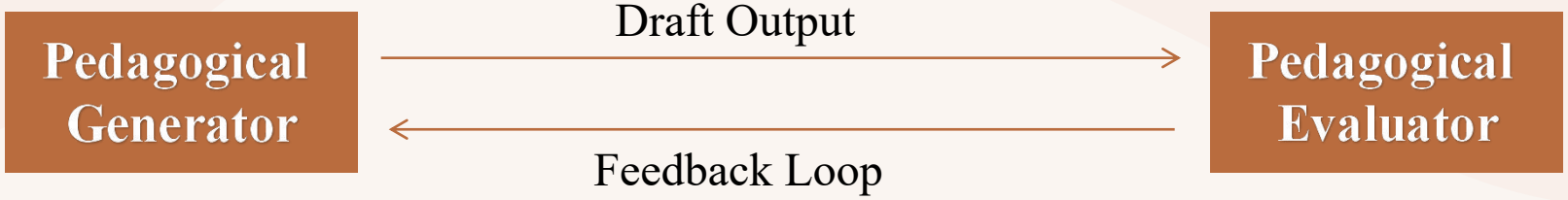
过滤门控： 仅当新提取信息 ∈ Target 时更新笔记。

设计意图：

目标导向：防止盲目重读引入噪声

收敛性：确保闭世界证据库的扩张是受控的

3. 生成后端：编排与审计



生成动作

基于 Query 和 Visual Notes 构建一个“教学骨架”

在填充骨架内容时，施加严格的“硬约束”

防御性的自检步骤，用于处理逻辑断点

两大硬约束

No-New-Symbol：禁止凭空创造符号

Typed-Link：推导句必须引用 step_id

核心动作

软性合规

检测： 检测引用链的拓扑跳跃（如 Step 1 → 9）

策略： 置信度衰减（Decay）而非硬性阻断

内部一致性

检测： 逻辑极性冲突（如文字说“增大” vs 符号 ↓）

策略： 拒答或 触发 增益重读

https://arxiv.org/abs/2512.03501
COMPUTE 2025 Best Practices Track

系统可靠性

目标

验证 Generator 是否绝对忠实于 Visual Notes，而非利用模型内部知识（Parametric Knowledge）作答。

指标

$$NDS = \frac{\text{Count(Flipped Logic in Answer)}}{\text{Count(Injected Flips in Notes)}}$$

- **判据**：$NDS \approx 1.0$ 表示系统是完全可控的（Notes 错，答案必错）；$NDS \ll 1.0$ 表示存在严重的幻觉泄露（即模型无视证据，利用内部常识"纠正"了答案，这在 RAG 协议中是不可接受的）。

系统鲁棒性

目标

验证 Evaluator是否能利用协议内部逻辑拦截错误证据。

指标

$$CDR = \frac{\text{Count(Triggered Rejections/Retries)}}{\text{Count(Injected Conflicts)}}$$

- **判据**：CDR 越高，说明系统越能充当"看门人"，有效拦截脏数据进入教学环节。

03

创新点

方法论创新

内容：提出 PGP 协议，将不可控的 VLM 生成重构为闭世界证据编排

关键词：No-New-Symbol, Typed Notes

架构创新

内容：基于推导局部性原理，利用滑动窗口实现软性合规检查。

关键词：Soft Compliance, Topology-Aware。

评估创新

解耦对抗性评测

内容：将安全性解耦为可控性与鲁棒性双维通过交换测试独立验证。

THE END

谢谢