

AI 个性化学习系统升级方案

3号员工 · 调研报告 | 对标：Khanmigo、Duolingo Max、松鼠 AI |
2026-07-22

一、现在怎么做的

1.1 学生建模

当前学生数据存在 `students/{name}.json` 中，结构如下：

- `subjects`：每科独立记录 `scores`、`weak_points`、`mastered`、`status` (`strong/weak/improving` 三档)
- `tests`：历次考试记录，逐题存 `correctFlag`、选项、答题
- `topic_stats`：知识点级别累计统计 (`correct/total` + 时间序列 `history`)
- `mistake_book`：错题列表 (200 条上限)
- `plan`：周计划
- `ai_analysis`：AI 自动生成的诊断文本

1.2 能力诊断

`weakness_analysis.py` 中 `diagnose_weakness()` 做时序分析：

- 读取每个知识点 `history` 中的正确率序列
- 分类：顽固 (连续 3 次低于 60%)、波动 (时高时低)、进步中、下滑
- 排序输出 + 附最近错题

1.3 出题引擎

`gen_new_tests.py`：

- 读取学生 `weak_points` → DeepSeek API 出题
- `exam_config.py` 静态知识图谱 (前置依赖链)，出题时补前置
- 双层验证：程序验算 + DeepSeek 独立推导

- 动态题数分配：根据近期得分率自动调整各科题数

1.4 间隔重复

vocab_db.py 实现 SM-2 变种：5 级记忆等级（1天/3天/7天/30天/90天），按到期时间排序取词。

1.5 AI 伙伴

tutor_routes.py：Socratic 教学、错题回顾、拍照解题（Qwen VL）、专项突破练习、家长看板。

二、业界怎么做的

2.1 松鼠 AI

"纳米级知识点拆分 + 五重错因分析"（知识漏洞、思维路径、概念混淆、粗心、猜测）。200 亿+学习行为数据支撑，MCM 模型逆向追踪学生思维路径。通用大模型占比不到 5%，核心是自研垂直教育模型。

2.2 Khanmigo (2025-2026)

给 AI 提供结构化学习历史摘要，使"下一题正确率"提升 3.4%，复习前置薄弱技能再提升 2.7%，合计 6.1%。

2.3 Duolingo Max

追踪 150+ 学习参数（犹豫位置、连续错、记忆衰减曲线），以 80% 正确率为目标动态调难度。

2.4 知识追踪前沿

RL-DKT 框架（Nature 2025）：DKT（LSTM/GRU）追踪知识状态时序演化，RL Agent 实时选择下一题。比传统 BKT：任务时间缩短 12.5%、退出率降低 50%、预测准确率提升 7.6%。

DSMKT（2025）：双序列建模（自注意力+GRU）+ 知识蒸馏。

2.5 间隔重复升级

Duolingo：除间隔天数外，加入犹豫度、首次正确率 vs 复习正确率、同义干扰效果，200+ 特征输入 ML 模型预测最佳复习时机。

2.6 多模态

Khanmigo 分析手写推导过程，松鼠 AI 草稿纸智能解析 + 情绪识别。

三、差距在哪

维度	我们现在	业界水平	差距
学生画像	3 个状态标签	6 维动态画像	大
知识粒度	粗粒度（如"分数进阶"）	纳米级（如"异分母分数加法"）	大
知识追踪	累计正确率统计	BKT/DKT/RL-DKT 深度学习	大
错因分析	对/错二值判断	5 重错因分类	大
学习路径	固定（前置链到薄弱点）	RL 动态优化	大
间隔重复	5 级固定间隔	200+ 特征 ML 预测	中
遗忘曲线	没有显式建模	Ebbinghaus 衰减融入模型	大
多模态	仅拍照识别	手写+语音+情绪	中
教学闭环	出题到做题到批改到结束	测到学到练到测到分析到调整	中

四、我们怎么改

改进 1：知识粒度细化 + 错因分类 [P0]

改法：每个知识点拆成 3-5 个纳米级，错题加错因维度（概念混淆/粗心/方法不会/审题错误/猜测）

工时 1.5h | 价值：诊断精度从"身体不舒服"提升到"胃疼"

改进 **2** : 遗忘曲线建模 + 智能复习排程 **[P0]**

改法 : Ebbinghaus 衰减 $\text{retention} = e^{-(t/s)}$, 低于阈值自动入复习队列

工时 1.5h | 价值 : 解决学过就忘, 长期记忆率提升 40-60%

改进 **3** : 周/月学习报告可视化 **[P0]**

改法 : 每周自动生成报告 : 知识点正确率曲线 + 薄弱诊断 + DeepSeek 自然语言解读

工时 2h | 价值 : 数据变洞察, 家长一眼看懂

改进 **4** : 探索-利用动态出题策略 **[P1]**

改法 : 20%巩固 + 50%突破 + 30%拓展, 连续 3 次全对标记已掌握

工时 2h | 价值 : 避免题海战术

改进 **5** : 学习行为日志 **[P1]**

改法 : 追踪对话轮数、主动提问比例、犹豫时间、学习时段偏好

工时 1.5h | 价值 : 区分能力问题 vs 习惯问题

改进 **6** : 对话分析发现薄弱点 **[P1]**

改法 : 每日扫描聊天记录, DeepSeek 提取求助高频知识点, 交叉验证考试数据

工时 1.5h | 价值 : 比考试更早发现困难

改进 **7** : 纳米知识点拆分 (**Yoyo** 数学试点) **[P1]**

改法 : Yoyo 数学 12 个知识点拆成 40-50 个, 一个学期后评估

工时 4h | 价值 : 诊断精度质变

改进 **8** : **A/B** 策略标签 **[P2]**

改法 : 每次出题记录策略版本 (v1/v2/v3), 对比正确率

工时 0.5h | 价值 : 知道改进是否有效

五、投入产出比一览

#	改进项	工时	改动量	优先级
1	错因分类	1.5h	小	P0
2	遗忘曲线排程	1.5h	小	P0
3	学习报告升级	2h	中	P0
4	动态出题策略	2h	中	P1
5	学习行为日志	1.5h	小	P1
6	对话分析薄弱	1.5h	小	P1
7	纳米知识点拆分	4h	中	P1
8	A/B 策略标签	0.5h	极小	P2

总计 14 小时，不增加 API 费用，全部开发时间成本。

六、不做事

- RL 完整框架（2 个学生数据不够跑效果）
- 多模态情绪识别（摄像头硬件成本）
- 自研垂直教育模型（需要百亿级数据，不是这个阶段的事）

七、核心策略

把 DeepSeek 用好，用 prompt engineering 替代模型训练，用 JSON 替代数据库。把大模型用好比训练小模型更划算，这是最适合当前阶段的路线。

建议从 P0 三项开始：改进 3（学习报告）、改进 2（遗忘曲线）、改进 1（错因分类），总计 5 小时。