AI MONTHLY REVIEW / OCTOBER 2026

技术迭代 × 业务需求 × AI探索

10月AI技术迭代 AI教学视频 生产洞察

国内外AI继续在多模态生成、任务执行和声音控制上迭代。围绕悦学资源处的教学视频需求，本月探索如何选择合适的制作路线，并将样片制作沉淀为可重复调用的Skill。

汇报部门 智能研发部 **报告期** 2026.10 **资料核验** 2026.11.03

01 **技术迭代：模型与制作工具共同推进**
观察13项国内外更新及其对内容生产的启发。

02 **项目切入：教学视频资源的制作需求**
把新能力放到具体课程资源与业务要求中检验。

03 **路线选择：画面生成与程序化合成**
区分视频生成模型和HyperFrames、Remotion的工作方式。

04 **探索结果：完成Skill搭建与交付**
形成可观看的样片、动作预览及可复用技能包。

国内外AI技术迭代

精选13项专题更新：国际8项、国内5项，同一系列合并。按首个可核验节点排序，分别记录发布、试用、预览与下线状态。

国际 · Google　10.01 / 10.13

Gemini 4 Argon：多步骤复杂任务继续推进

10月1日先面向可信网络防御者分阶段发布；13日向开发者、企业及符合条件的订阅用户开放，强调长时推理与复杂工程任务。

教育应用推演

内容生产可进一步拆成资料整理、脚本、素材与检查等连续步骤，并为每步保留可审阅结果。

分阶段发布 → 开放使用　[A1]

国际 · Anthropic　10.07

Claude Haiku 5.5：高频任务的成本选择

面向摘要、分类、重复操作等高频任务。官方称平均运行成本较 Haiku 4.5 低约75%，并同步下调 Sonnet 5.5 的缓存读取价格。

教育应用推演

可按任务难度分配模型：常规整理优先控制成本，关键内容校验单独设置质量标准。

已发布　[A2]

国际 · OpenAI　10.08 / 10.26

GPT-6.1：速度档位与视觉任务分工

8日为 GPT-6.1 Sol 增加 Ultrafast 档位；26日发布 GPT-6.1 Terra，面向视觉推理、计算机操作和文字任务。

教育应用推演

生产流程可分别评估脚本整理、界面操作与画面检查的模型需求，关注整项任务的完成质量。

API 更新 / 新模型　[A3]

国际 · Meta　10.12

Muse Wonder：智能体产品继续扩展

发布 Muse Wonder 与 Wonder Full，强调个人和专业智能体任务。前者通过 Meta Model API、Muse Code 提供，后者面向 Muse Code 订阅用户。

教育应用推演

工具选择逐渐围绕完整工作任务展开，教育内容团队也需要评估跨素材、跨文件操作是否可靠。

API / 产品访问　[A4]

国内 · 阿里 / QwenCloud　10.12-10.22

Qwen Audio 3.1：配音控制继续细化

12日列出 TTS Flash、21日列出 TTS Plus、22日列出 TTS Next。系列覆盖低延迟语音、高质量配音与统一音频生成等不同定位。

教育应用推演

教学配音可围绕语速、情绪、音色和时间控制选型，并对数字、术语与否定词单独复听。

系列更新　[B1]

任务效率与生成前审阅

模型在任务执行、声音表达和内容审阅上持续推进，为教育内容生产提供更多组合方式。

国际 · OpenAI 10.16

GPT-Live：声音可由文字描述创建

官方日志新增 prompt-to-voice：通过文本描述创建自定义声音，再复用于 GPT-Live 会话；语音生成仍按会话规则计费。

教育应用推演

虚拟讲解角色可以把声音设定作为可复用资产，与脚本、形象和动作保持一致。

API 可用 [A5]

国内 · DeepSeek 10.17

V4.1-Flash-1017：任务效率与搜索接入

V4.1 Flash 升级为1017版本，网页、App同步更新，API调用方式与价格保持不变；同日开放自研搜索API。

教育应用推演

资料检索、脚本整理与自动化处理的选项增加，可用固定样本复测准确性、耗时和总成本。

模型迭代 / API 开放 [B2]

国内 · MiniMax 10.19

M3.1 Flash：智能体任务的效率升级

面向智能体、编程和计算机操作。官方称在线推理接近提速3倍、使用成本降低70%，已通过 API、M Plan 和 MiniMax Code 提供。

教育应用推演

可探索文件整理与批处理编排，把完成一条合格视频所需的重试和复核计入成本。

API 可用 / 权重待发布 [B3]

国际 · Google 10.19

Gemini Notebook：把审阅放在渲染之前

为 Slide Decks 与 Video Overviews 增加共创流程，可选择 Review，在生成前调整大纲或视频制作说明，再开始渲染。

教育应用推演

教研审阅可以前移到选题、脚本和分镜阶段，尽早修改知识顺序及讲解重点。

逐步推出 [A6]

对教育内容团队的启发：把资料、声音、画面和审阅组织为可执行步骤，再比较完整任务的质量与成本。

生产服务与交互能力

生成质量、服务生命周期与交互能力共同变化，技术选型需要回到具体的业务目标和交付要求。

国际 · Google 10.22

Veo 3.1 Preview：生产链要管理模型退役

Gemini API 中 Veo 3.1 的三个 Preview 模型 ID 于22日下线，官方提供 Gemini Omni Flash 或企业平台 GA 模型的迁移方向。

教育应用推演

可复用流程应记录模型版本、输入和中间产物，便于在服务变化时复现与迁移。

指定预览入口下线 [A7]

国内 · 字节 / BytePlus 10.27

Seedance 2.5 4K：草稿先行的制作流程

BytePlus 的4K说明介绍480p Draft预览、最终4K输出及源素材细节保留，并披露精选影视制作和后期客户的试用。

教育应用推演

教学素材可先用低成本草稿确认构图与节奏，再进入高质量输出阶段，减少无效重做。

精选影视客户试用 [B4]

国内 · 阿里 / QwenCloud 10.27

Qwen Audio Realtime Flash：双工语音继续扩展

官方日志列出 qwen-audio-3.1-realtime-flash，支持实时双工语音、工具调用、搜索与声音克隆。

教育应用推演

视频内容未来可接入暂停提问、口头回答等互动环节；需要另外设计问题和学习反馈。

模型上线 [B5]

国际 · Google DeepMind 10.27

Genie 4：从观看内容走向交互模拟

发布面向可信测试者的研究预览，展示带原生音频的实时世界模拟，以及通过自然语言改变场景的能力。

教育应用推演

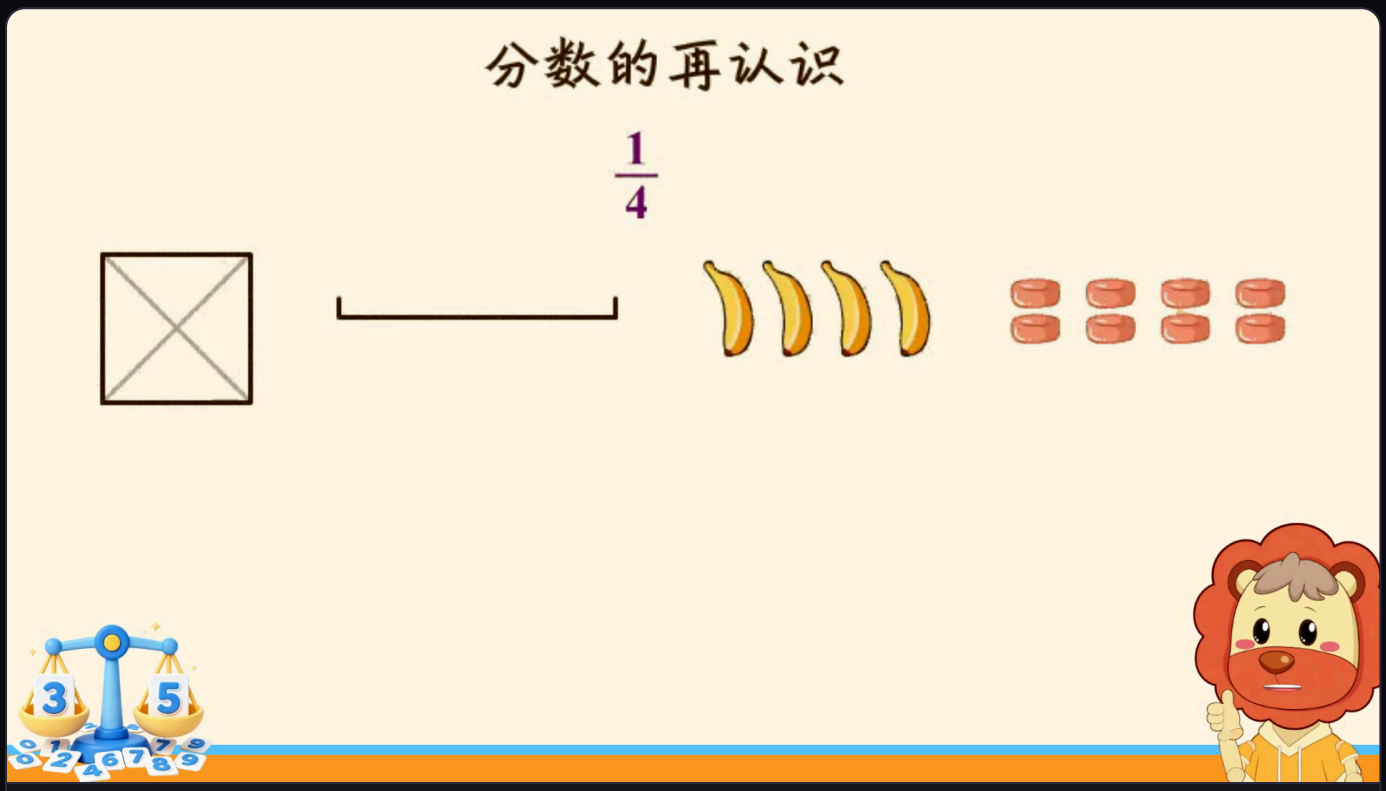
可作为情境探究和交互原型的观察方向；教学使用还需要验证规则、因果与知识准确性。

研究预览 [A8]

由此引出本月项目：面对教学视频资源的实际制作需求，如何组合AI能力和制作工具，形成可交付、可复用的流程？

从技术迭代，落到教学视频生产

模型与工具的持续进步，为教学视频提供了新的制作方式。面向悦学资源处的课程资源需求，本月以AI教学视频项目探索内容处理、角色表达与流程复用。



教学视频项目 · 动作速览（约20秒）

网页版可切换完整成片；下列规格来自已交付样片的核验记录。

1080p	139.4s	8类	19次
1920×1080 · 30帧/秒	完整教学样片	角色动作	旁白触发动作

项目要回答的问题

如何让已有课程内容、统一的角色表达和具体的画面要求协同工作，并让这套制作方法可以继续用于同类教学资源？这使技术选择从“能生成什么”，进一步走向“怎样稳定地完成指定制作任务”。

视频生成模型与程序化制作

视频生成模型负责依据提示与参考素材生成或编辑画面；HyperFrames、Remotion负责用代码组织素材、布局和时间关系，再渲染为视频。两类能力可以在同一流程中协作。 [B4] [H1] [H2]

比较维度	视频生成模型	HyperFrames / Remotion
工作方式	通过模型推理生成或修改视听内容。	把已定义的素材、图层和时间规则转成视频帧。
主要控制入口	提示词、参考素材及模型提供的编辑参数。	代码、配置、组件和帧或时间函数。
文字与图形	生成后的文字、细节与关系需要逐项复核。	可将文字、图形和现有内容作为明确元素编排。
动作与时序	可按模型能力约束动作或时间，仍需检查输出。	可规定出现时点、位置、动作范围及进退场条件。
修改与复用	通过再次生成或编辑迭代，结果随模型与输入变化。	修改指定规则或素材；固定环境后可按帧复现和检查。
适合的生产任务	探索新镜头、场景、视觉素材与表演。	制作有明确内容、布局、时序和模板要求的视频。

HYPERFRAMES BY HEYGEN

HTML描述画面与时间关系

以HTML、CSS和JavaScript组织视频、音频与图层，通过时间轴编排动画并逐帧渲染。 [H1]

REMOTION

React组件随帧编号变化

以React组件组合文字、图片、视频和声音，由帧编号驱动动画与画面状态。 [H2]

本项目采用程序化制作路线，实际实现基于HyperFrames。

教学资源改造具有明确的内容、角色和时序要求。项目用HyperFrames编排分层动画，结合FFmpeg与语音转换完成处理，并将操作步骤封装为Skill。Remotion提供同类路线的另一种实现方式。

04 / FROM BUSINESS NEED TO SKILL

面向悦学资源处：从需求到Skill

将教学资源制作需求拆成内容、表达、时序与复用要求，再转成可执行的处理规则。程序化制作承担明确的控制逻辑，AI辅助完成需求拆解、代码编写与迭代。

业务要求	对应的实现方式	沉淀的能力
统一资源呈现	将画面处理、角色声音和装饰布局分别配置。	形成可复用的视觉与声音设定。
保障教学内容清晰	围绕文字、公式、图形和关键读音安排复核。	保留画面与声音的检查记录。
让角色配合讲解	依据旁白设置动作时点，检测与课程内容的相交。	角色按规则表达，并在需要时避让。
支持同类视频继续使用	把输入、参数、素材与处理步骤组织为同一入口。	替换原始视频后复用制作方法。
便于交付与后续维护	将默认素材、脚本和说明一并打包，保留中间项目。	交付可安装调用、可继续修正的Skill。

01 / INPUT

原始教学视频

读取课程素材；需要时提供新的样音或配置。

02 / PROCESS

画面与声音处理

执行素材处理、配音转换及内容比对。

03 / COMPOSE

HyperFrames合成

编排角色、动作和避让规则，完成渲染。

04 / OUTPUT

成片与核验记录

输出MP4，并保留检查结果与可编辑项目。

REUSABLE SKILL / TEACHING-VIDEO-POLISH

把一组制作操作，组织为一个可调用的入口。

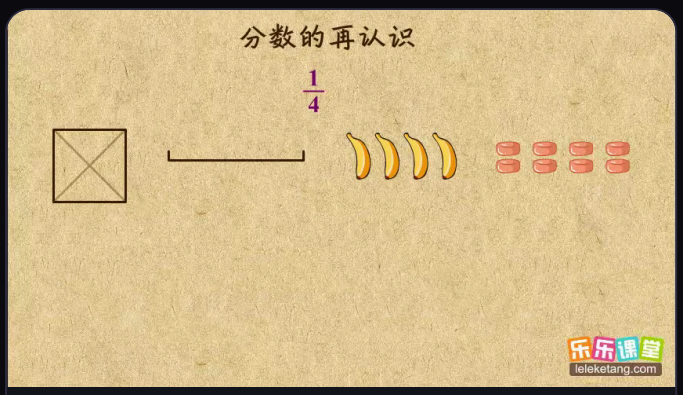
输入课程视频 → 执行既定规则 → 检查结果 → 输出完整成片

这次交付既包含完成的教学样片，也包含执行同类任务的Skill：将处理脚本、默认素材、配置与验证步骤共同保存，便于继续使用和调整。

05 / SKILL BUILD & DELIVERY

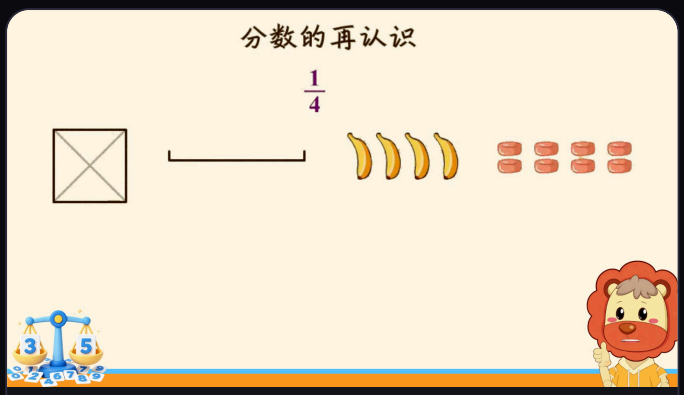
从样片制作，到Skill搭建与交付

围绕同一套需求逐步完成画面处理、声音转换、角色动画与验收，再把可复用的部分收拢为teaching-video-polish Skill。以下为项目样片的原始素材与交付画面对照。



输入：原始教学素材

作为课程内容与时序的核对依据。



输出：项目交付样片

完成统一表达与角色动作编排。

01 / FINISHED VIDEO

完整教学样片

约2分19秒、1080p；用于查看整体制作效果。

02 / MOTION PREVIEW

20秒动作预览

集中展示角色表情、手势与讲解配合方式。

03 / REUSABLE SKILL

Skill便携包

约6.1MB、102个条目，包含脚本、说明与默认素材。

动作逻辑随Skill保存 / 8类动作

- 01 / 招呼
- 02 / 讲解
- 03 / 指向
- 04 / 思考
- 05 / 肯定
- 06 / 提醒
- 07 / 比较
- 08 / 告别

交付内容同时覆盖“结果”和“方法”。

成片与预览用于查看效果；Skill保留素材、配置和处理步骤，便于继续处理同类视频。便携包已包含样音、装饰和动作素材，首次使用仍需按说明准备运行环境和模型。

配套网页版提供完整成片、动作预览与Skill便携包的下载入口。

06 / OUTCOMES OF AI EXPLORATION

探索结果：把AI能力沉淀为工具

本月探索从技术观察出发，经过业务需求拆解、制作路线选择和样片验证，最终形成了可观看的项目产出与可重复调用的Skill。

01 / CHOOSING THE TOOL

明确制作路线

把画面生成与程序化制作分开评估。对有明确内容、布局和时序要求的教学资源，采用程序化合成承担控制逻辑。

02 / BUILDING THE WORKFLOW

把需求转成规则

AI辅助拆解需求、编写处理与动画逻辑，并在关键画面检查中迭代；让动作、口型和避让进入可检查的流程。

03 / DELIVERING THE SKILL

沉淀可复用交付

将一次样片制作扩展为带默认素材、参数、运行说明和验证步骤的Skill，支持同类任务继续使用。

已形成的结果	对应的项目证据	体现的能力
可观看的教学样片	1080p、139.4秒，8类动作、19次旁白触发。	把需求落实为具体的视听产出。
可检查的制作过程	全片解码错误0，既定避让规则下时序违规0。	为结果保留核验依据和修正入口。
可继续使用的Skill	102个条目，默认素材与脚本随包交付。	将制作经验组织为可调用的工具。

本次AI探索的落点：从“完成一条视频”走向“交付一项制作能力”。

围绕悦学资源处的需求，项目已完成样片、预览和Skill包的搭建与交付。后续可在更多类型的教学资源上记录人工干预、返工原因和总制作成本，继续验证复用的稳定性与实际收益。

本期证据覆盖样片与工程交付。规模化人效、制作成本降幅及学习效果尚未形成测量结论。

资料来源与核验说明

技术事实以官方发布与文档为主。报告期为2026年10月1日至31日，资料核验于2026年11月3日；路线选择与业务映射为本报告结合项目情况的分析。

A · 国际技术迭代

[A1] Google　10-01 / 10-13

Gemini 4 Argon：由分阶段发布走向开放使用

区分首次宣布与面向开发者、企业及符合条件的订阅用户开放。

10月1日发布说明 ↗
10月13日开放公告 ↗

[A2] Anthropic　10-07

Claude Haiku 5.5

成本降幅为厂商报告，具体收益取决于任务与配置。

Haiku 5.5 官方发布 ↗
模型规格 ↗

[A3] OpenAI　10-08 / 10-26

GPT-6.1 Sol 速度档位与 GPT-6.1 Terra

Sol 为既有模型增加档位；Terra 为本月新模型，支持图像输入与文本输出。

API 更新日志 ↗
GPT-6.1 Terra 模型文档 ↗

[A4] Meta　10-12

Muse Wonder / Muse Wonder Full

分别按 API、Muse Code 与订阅访问范围描述；本报告不将其记为开放权重。

Meta 官方发布 ↗

[A5] OpenAI　10-16

GPT-Live：用文本描述创建声音

文本生成声音与基于样音的声音克隆属于不同输入方式。

API 更新日志 ↗
Custom voices 官方文档 ↗

[A6] Google　10-19

Gemini Notebook：幻灯片与视频共创

公告称在随后一周全球逐步推出；Review 为可选择的生成前审阅。

共创功能官方说明 ↗

[A7] Google　10-22

Veo 3.1 Preview 的 Gemini API 入口下线

仅指列出的三个 Preview 模型 ID；不等于全部 Veo 3.1 服务停止。

Veo API 官方文档 ↗

[A8] Google DeepMind　10-27

Genie 4：交互式世界模拟

研究预览，面向可信测试者；面向更多用户和开发者的开放属于后续安排。

Genie 4 官方研究发布 ↗

同一系列合并，采用可核验的官方事件或文章日期。制作框架文档用于说明技术路线，不计入本月13项技术更新。

国内来源、制作框架与项目证据

框架能力来自官方技术文档，项目实现与交付范围来自随附的技能包和核验记录。

[B1] 阿里 / QwenCloud — 10-12 / 10-21 / 10-22

Qwen Audio 3.1 TTS 系列

Flash、Plus、Next 为不同定位的条目，合并为一项系列更新。

官方模型发布日志 ↗

[B2] DeepSeek — 10-17

DeepSeek-V4.1-Flash-1017 与搜索 API

API 模型别名与价格保持不变；性能和耗时变化为厂商报告。

DeepSeek 官方更新日志 ↗

[B3] MiniMax — 10-19

MiniMax M3.1 Flash

API、M Plan、MiniMax Code 可用；权重发布为后续计划。

MiniMax 官方发布 ↗

[B4] 字节 / BytePlus — 10-27

Seedance 2.5 4K

日期按可核验官方文章记录；基础版本为7月发布，4K文章描述精选影视客户试用。

BytePlus 官方说明 ↗
7月31日基础版本发布 ↗

[B5] 阿里 / QwenCloud — 10-27

Qwen Audio 3.1 Realtime Flash

实时双工语音模型；对本项目的意义为应用推演，并非项目实际选型。

官方模型发布日志 ↗

[H1] HyperFrames by HeyGen — 技术说明

HTML合成与逐帧渲染

使用HTML描述素材、图层和时间关系，按指定帧渲染；为本项目实际采用的制作框架。

HyperFrames 官方合成文档 ↗

[H2] Remotion — 技术说明

以React和帧编号编排视频

以React组件、帧编号和媒体组件组织视频；作为程序化制作的同类方案参照。

Remotion 官方核心概念 ↗

[P1] 项目交付物

完整成片 · 动作预览
teaching-video-polish Skill便携包

[P2] 工程核验记录

规格、动作、解码与便携包核验
来源与中间文件复用记录

需求背景按本次提供的悦学资源处业务诉求整理。项目实际采用HyperFrames；Remotion用于同类技术路线的比较。交付范围以样片、Skill包与核验记录为准。