ISSUE 005 · 03-07 AUGUST 2026

控制
回到画面里

从 3D 调度到 30 秒长镜头，AI 影像开始处理生成之后的工作。

FLOW STUDIO

SIGGRAPH

SEEDANCE 2.5

SOCIAL FIELD

Original artwork commissioned for Rainwalker Weekly · 2026

EDITOR'S LENS

镜头生成以后，麻烦才真正开始。

这几日，屏幕上出现了许多三十秒的长镜头。人物走过房间，推门，来到泳池；云层卷起，机械巨人从远处逼近。画面长了，旧问题便无处躲藏：手里的杯子忽然换了位置，影子先人一步，奔跑的人在雨里失了骨头。

行业也在转身。Autodesk 把摄像机、角色和场景拉回三维空间；SIGGRAPH 谈的是一帧如何真正交付；Google 让生成后的画面继续接受剪辑指令。本期沿着这条线，看控制如何重新进入制作。

社媒取样

X、小红书与创作者社区用于发现线索；正文只引用可公开回溯的原帖。

报道窗口
03-07 AUG

信息截止
07 AUG · 18:00 CST

阅读时间
28 MIN

页数
40 PAGES

THE WEEK, EDITED

五天里，*控制成为关键词。*

08.04	FLOW STUDIO	Autodesk 展示 3D Editor 与 Canvas，把摄像机、角色和场景调度放回可操作空间。	报道 ↗
08.05	GOOGLE VIDS	Gemini Omni 的视频生成与指令式编辑开始分批进入 Google Vids。	官方 ↗
08.05	DEEPMIND	Demis Hassabis 转任 DeepMind 主席及 Alphabet 首席科学家，管理层开始重新分工。	Axios ↗
08.05-07	SEEDANCE 2.5	即梦、小云雀和第三方入口继续出现新样片，三十秒单镜头进入公开压力测试。	本期审计 ↘
08.07	CREATOR FIELD	长镜头、迷你 DV 日常片与动作比较在 X 和创作者社区集中出现。	现场 ↘
08.10	DEADLINE	CapCut CRE[AI]TE 征集即将截止，创作者只剩一个周末整理作品。	投稿 ↗

时间线按公开发布时间与本期检索时间整理。平台开放范围可能随地区和账号变化。

90-SECOND BRIEF

本周五个判断

01 生成质量仍在涨，产品竞争已经移向控制。

摄像机、角色、场景、节点和剪辑指令，正在成为新的卖点。

05 ↘

02 三十秒长镜头可用了，也更容易穿帮。

Seedance 2.5 的连续性变强，物体位置、影子和快速动作仍会破坏可信度。

13 ↘

03 “最后 20%”成为行业共同难题。

一段好看的样片离可交付画面仍隔着修补、合成、版权和来源记录。

09 ↘

04 平台要的内容更加分化。

Disney 把 TikTok 创作者带进 Verts，真实作者与生成工具可能在同一分发系统里并行。

24 ↘

05 社媒热度适合发现作品，不适合替作品下结论。

本期保留原帖与动态画面，同时把失误、成本和方法写在旁边。

28 ↘

LEAD STORY

生成之后，导演需要一个可以站住脚的空间。

Autodesk 本周展示 Flow Studio 的 3D Editor。角色、环境、动画和摄像机被放进同一空间，创作者可以移动机位、换镜头、调动作，再把结果送入 Canvas 继续处理。

这一步听来寻常。电影制作原本就在三维世界里发生。过去两年的生成工具却常把导演留在文本框外，让空间关系藏在模型的猜测里。

Creative Bloq 报道 ↗

A DIRECTABLE SPACE

机位终于不必藏在提示词里。

在 3D Editor 中，摄像机有位置，人物有骨架，场景有尺寸。镜头为什么看不见人物、动作为什么穿过墙面，都能在空间里找到原因。

Camera	改变机位、焦段与运动路径。
Character	AI 动捕与自动绑定把表演交给可编辑骨架。
Environment	文字或图像可以转为 3D 资产，再进入场景。
Continuity	同一资产可在多个镜头中继续使用。

官方能力列表 ↗

3D EDITOR · CAMERA, CHARACTER, ENVIRONMENT

ONE WORKSPACE, TWO WAYS OF THINKING

先摆场，再接节点。

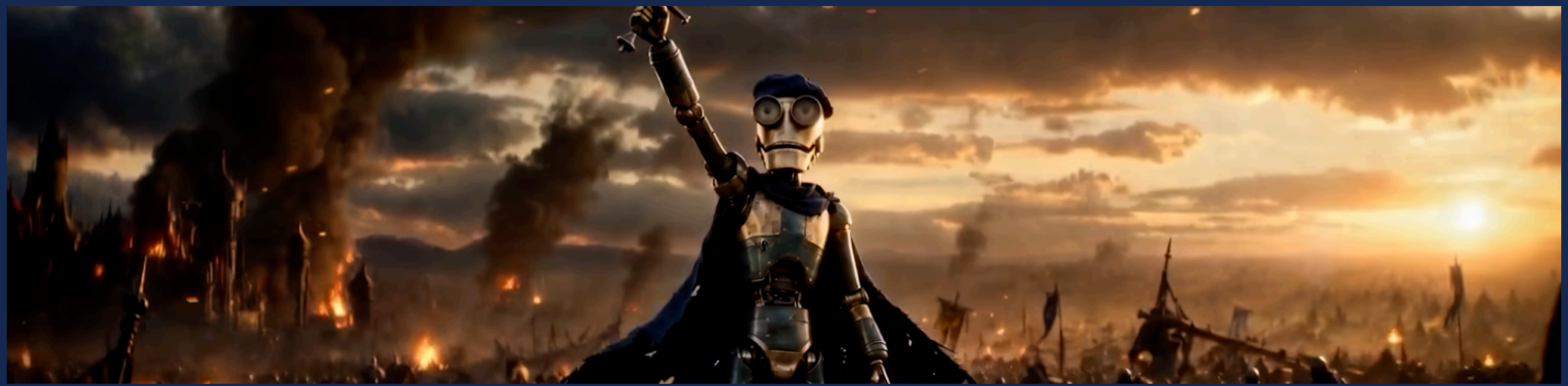
空间里的决定进入 Canvas 后，节点接管版本、生成和后期。导演不必一开始就面对复杂流程，技术人员也不必放弃精细控制。

01
场景
放入角色、环境与道具。

02
调度
确定机位、动作与时间。

03
Canvas
连接参考、模型与版本。

04
交付
进入合成、剪辑和审查。



官方示例显示：同一角色和动作可以在不同镜头中继续调度。

Flow Studio 官方页面 ↗

WHAT CHANGES ON MONDAY

提示词仍要写，制作重心已经往前移。

一条镜头不再只由一句描述决定。角色资产、场景母版、动作数据、机位和版本记录一起进入生产。小团队需要学会管理这些东西，大团队则会重新划分导演、动画、技术美术和合成之间的边界。

FLOW STUDIO · PRODUCT VIEW

独立创作者

可以先用二维工作方式起步，需要时再进入三维调度。

制片团队

资产复用和版本追踪开始直接影响成本。

仍待观察

学习成本、模型开放范围、跨软件交换与最终输出质量。

Autodesk SIGGRAPH 2026 ↗

THE LAST 20%

漂亮的样片之后，还剩最后二成。

SIGGRAPH 今年把许多讨论留给同一件事：生成画面如何进入真正的制作。补帧、材质、合成、来源判断和软件协作，都是那最后二成。

SIGGRAPH 官方专题 ↗

FROM IMPRESSIVE TO USABLE

一帧能交付，要经得起放大。

会议把一场讨论直接命名为“The Last 20%: Getting GenAI to a Deliverable Frame”。标题很坦白：生成的第一眼往往已经够好，真正费工的是边缘、动作、材质和镜头之间的衔接。

> 观众可以原谅风格，却很难忽略突然变形的手、失去重量的脚，或一只凭空换边的杯子。

查看官方议程 ↗

SIGGRAPH 2026 · INDUSTRY STAGE

01

动作是否连续

02

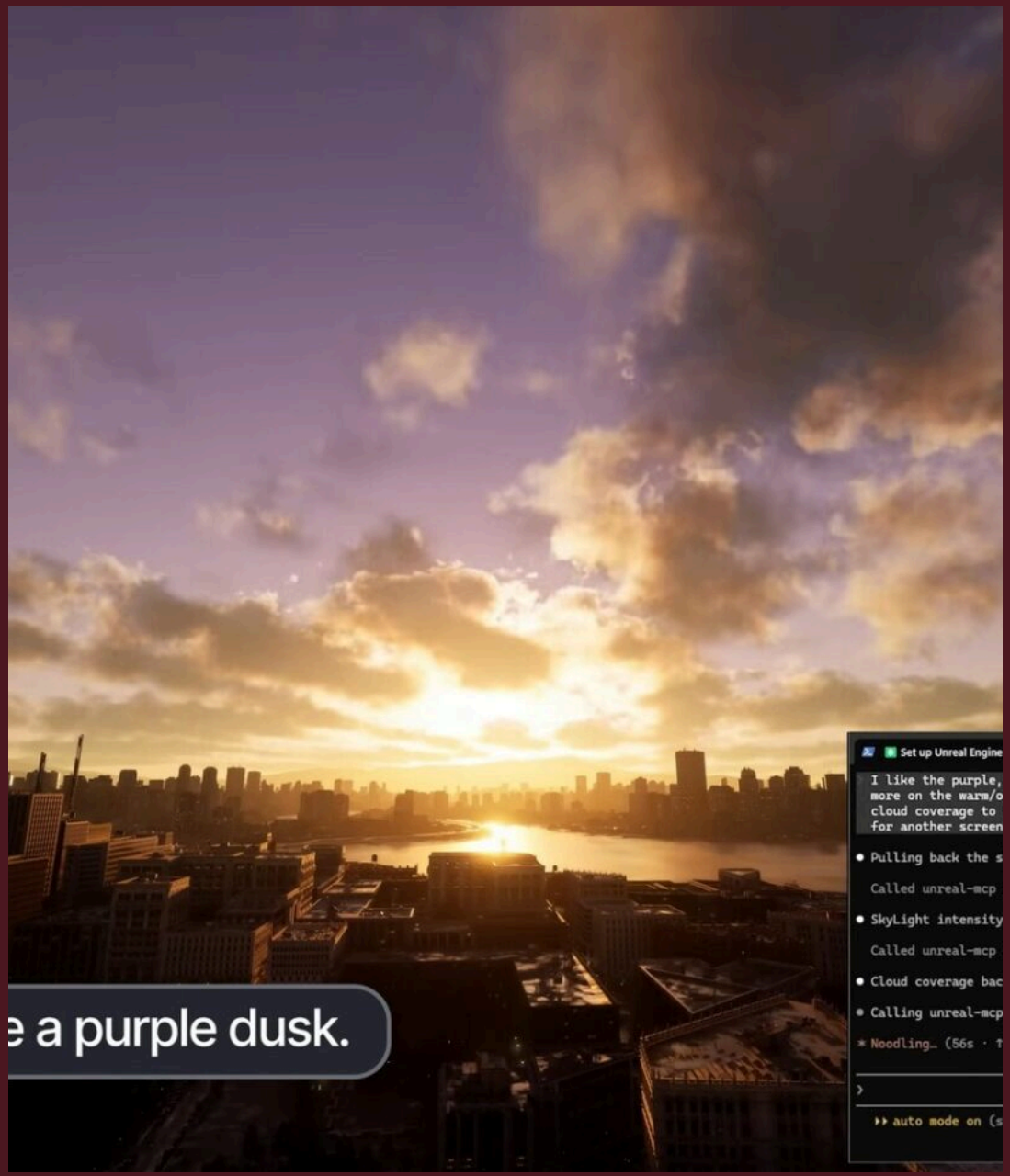
材质能否重打光

03

画面能否进入合成

04

来源与授权能否说明



NVIDIA · MCP-CONNECTED CREATIVE TOOLS

SOFTWARE STARTS TO TALK

Adobe、Blender、Houdini，开始接进同一张桌子。

NVIDIA 在 SIGGRAPH 展示一组与 MCP 相连的创意工具。Adobe、Affinity、Blender、Boris FX Silhouette、Foundry、Houdini 与 Unreal 都在名单里。

MCP 可以把它理解为一套通用接口。代理收到任务后，能够知道不同软件会什么，并把一步步操作交给合适的工具。它能否稳定执行复杂制作，还要看权限、错误处理与版本兼容。

NVIDIA 官方发布 ↗

PROVENANCE

检测合成视频，压缩会让答案变难。

NVIDIA 同时公布 Synthetic Video Detector NIM。厂商测试显示，未经压缩的视频识别率最高；上传平台、反复转码以后，准确度会下降。

这类工具适合风险筛查，不能独自承担最终判断。创作者仍需保存源文件、生成记录、授权与修改历史。

测试条件与完整说明 ↗

VENDOR-PUBLISHED TEST

以上均为 NVIDIA 公布的测试数字，任务设置与独立复核仍需补充。

ONE WEEK OF PUBLIC OUTPUTS

三十秒，足够让优点和破绽一起走到镜头前。

Seedance 2.5 已在中国的即梦与小云雀出现首发，并持续产生样片。第三方平台也开始接入。公开作品显示，人物和场景可以在更长时间里保持方向，物体位置、人体结构和快速运动仍会突然失手。

00 — 10 — 20 — 30 SEC

CREATOR SAMPLE · SHORT EDITORIAL EXCERPT · OPEN ORIGINAL ↗

CASE 01 · PARTY TO POOL

一扇门，把室内和泳池连成一条路。

人物从聚会穿过门口，走到阳光下，最后进入泳池。三十秒里，服装、发色和行动目标大体保持，空间转场也能读懂。

细看仍能发现影子不稳、杯子位置变化、肢体在水中失真。这类错误不再出现在第一秒，它们埋在一段看似顺畅的行动中。

可用处	室内外连续转场
风险	物体与身体连续性

查看原帖与完整视频 ↗

CASE 02 · FAST ACTION

时间变长，动作未必更利落。

一组同题打斗测试里，2.5 的动作连接更长，画面却显得偏软；2.0 的短段落反而更清楚。这个样本不能代表全部模型表现，却提醒创作者：时长、锐度和动作密度要一起测试。

高速场景最怕“每一步都发生了，观众却看不清”。剪辑仍要为重击、反应和方向留出位置。

查看完整比较 ↗

CREATOR COMPARISON · 2.5 / 2.0

THIRTY SECONDS NEED STRUCTURE

长提示词不够，还要给时间安排事情。

一位创作者让镜头连续穿过六个房间。另一组"同提示词比较"则发现，把 2.0 的短提示直接交给 2.5，三十秒会出现松散和重复。时间变多以后，提示需要像场面调度：何时进入、看见什么、转向哪里、怎样结束。

00-05	建立人物、方向与环境
05-12	完成第一次行动或转场
12-20	引入变化，维持空间关系
20-27	完成第二个动作与结果
27-30	收束，给剪辑留下出口

六房间测试 ↗ 同提示词比较 ↗

WHERE IT CAN BE FOUND

入口正在增加，版本说明仍不统一。

中国用户已在即梦与小云雀看到首发和产出；8 月 7 日，Atlas Cloud 与 Venice 也分别宣布接入。不同平台的分辨率、积分、过滤规则和命名方式并不相同。

中国首发

即梦 / 小云雀

当前最早的一批公开样片来自中国应用。

即梦 ↗

第三方

Atlas Cloud

8 月 7 日宣布上线 30 秒单镜头入口。

公告 ↗

第三方

Venice

同日公布可用入口，具体能力随套餐变化。

公告 ↗

谨慎阅读

价格与积分

社区中已有成本记录，但它们会随地区和套餐变化。

用户记录 ↗

本期不把第三方平台公告等同于字节跳动的全球正式发布。

EDITORIAL AUDIT

2.5 的进步，主要写在时间里。

人物能够在更长的行动中保持目标，镜头也更愿意穿过门、楼道和复杂环境。成片感因此上升。与此同时，错误从“立刻看见”变成“看了一会才发现”。审片需要更慢，也更细。

↑ **长段落连贯** 三十秒内的动作链和空间转场更容易成立。	↑ **风格保持** 材质、色调和人物外观可维持更久。	→ **物体逻辑** 杯子、衣服、影子和手仍会改变位置。
→ **快速动作** 高速度下的清晰度与身体结构仍不稳定。	? **成本** 长时长让单次失败更贵，积分体系尚未稳定。	? **正式边界** 地区、平台、API 与版权过滤仍需后续确认。

更长的生成，把导演的时间还回来，也把场记和审片的责任一并还了回来。

ROLLOUT · 05 AUGUST

生成完以后，还可以继续告诉它怎么改。

Google Vids 开始分批上线 Gemini Omni。用户可以生成更高质量的视频片段，也能用一句话要求它改色调、换风格，或移除画面中的警笛声。

这里的变化很实际：AI 视频第一次生成后，不必立刻被当成定稿。它可以进入下一轮编辑。

Google Workspace 官方说明 ↗

AFTER THE FIRST GENERATION

一句话，进入三种编辑。

INPUT

“把这段改成傍晚。”

色调、光线与气氛改变。

STYLE

“转成手绘动画。”

重绘视觉风格，保留主要动作。

AUDIO

“去掉警笛声。”

处理声音，而不必重做全部画面。

对于普通用户，这是少开一个软件。对于专业团队，它更像一条信号：生成、剪辑和声音处理会逐渐出现在同一界面里。

真正的考验在于修改是否局部、结果是否可重复，以及每一次编辑能否保留版本。

功能示例 ↗

WHO GETS IT FIRST

分批开放，也分地区。

Google 采用 Scheduled Release，8 月 5 日起最长约十五天完成推送。商业、企业、教育及部分消费者 AI 套餐在名单中。

GENERATE

面向多个 Workspace 与 AI 套餐逐步开放

EDIT NON-AI VIDEO

欧洲经济区、瑞士、英国、得州与伊利诺伊州在首发阶段受限

产品能力与法律边界如今常在同一张更新表里出现。团队需要把地区可用性写进交付计划。

查看完整适用范围 ↗

《月下之臣》· 预告报道图片

PREVIEW HEAT

《月下之臣》说要在八月来，正式入口仍待确认。

报道把它称为八集、约二百四十分钟的 AI 短剧，并强调全女性制作团队。到本期截稿时，我们没有找到可确认的官方上线地址。

这段空白值得记录。预告、定档报道和平台上线是三件事。长内容项目越来越会制造声量，编辑也要把“说要来”与“已经能看”分开。

查看预告报道 ↗

AVAILABLE TO WATCH

真正能看的项目，才进入下一轮讨论。

《奇谭：纸刃渡荒墟》已于 7 月 23 日上线。公开资料称，制作团队建立了两百余项角色与视觉资产。它给行业留下的问题更加具体：人物在长叙事中能否保持，观众愿意看多久，平台怎样推荐。

本期把它作为背景样本，不拿旧项目冒充本周新发布。

新华社报道 ↗　观看页 ↗

《奇谭：纸刃渡荒墟》· iQIYI

DISNEY × TIKTOK CREATORS

DISTRIBUTION

Disney 把一部分竖屏内容，交给熟悉平台的人来做。

TechRadar 本周报道，入选的 TikTok 创作者可以使用 Disney IP 制作竖屏视频，部分作品还有机会进入 Disney+ Verts。

这是一条面向真人创作者的发行通道。它与 Disney 早前探索生成视频并不冲突，却说明平台仍需要有作者、语气和社群关系的内容。

阅读报道 ↗

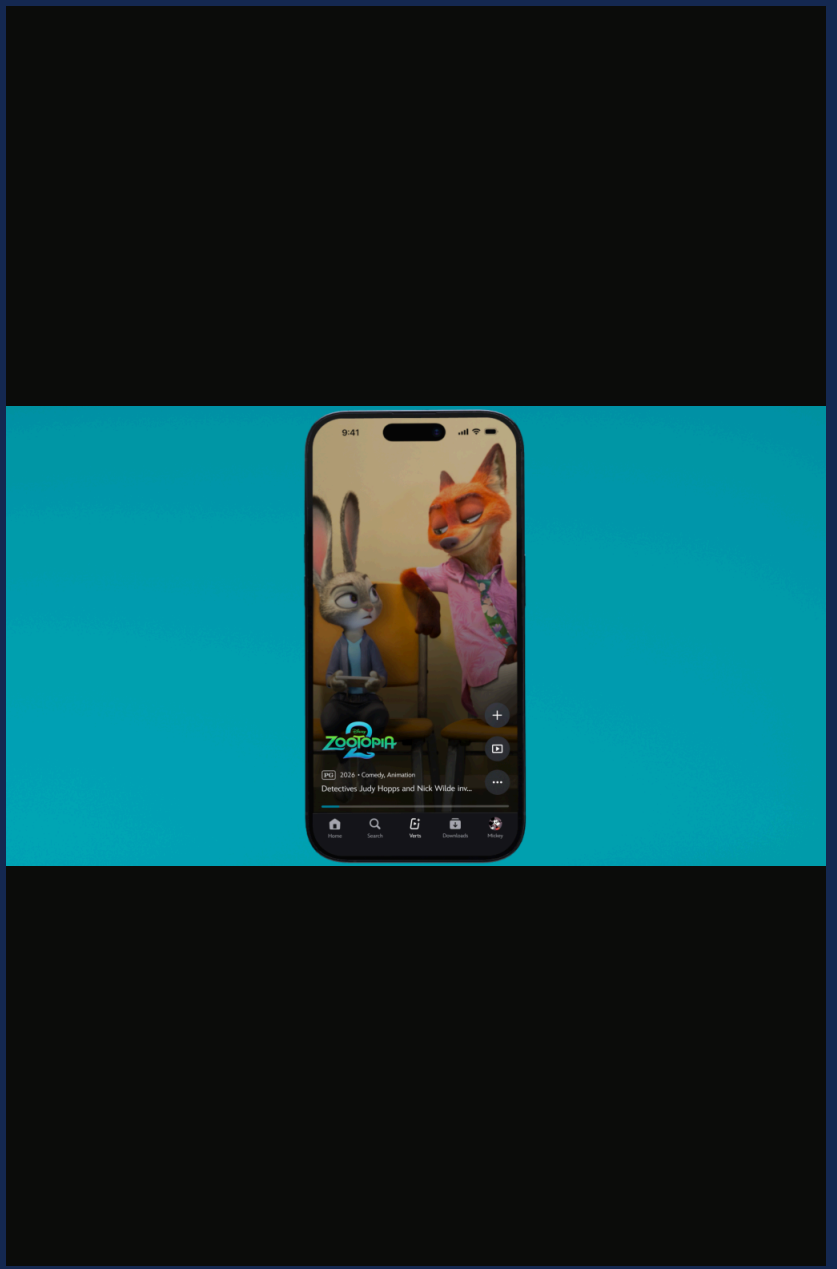
VERTICAL VIDEO INSIDE STREAMING

短视频的门，开进流媒体客厅。

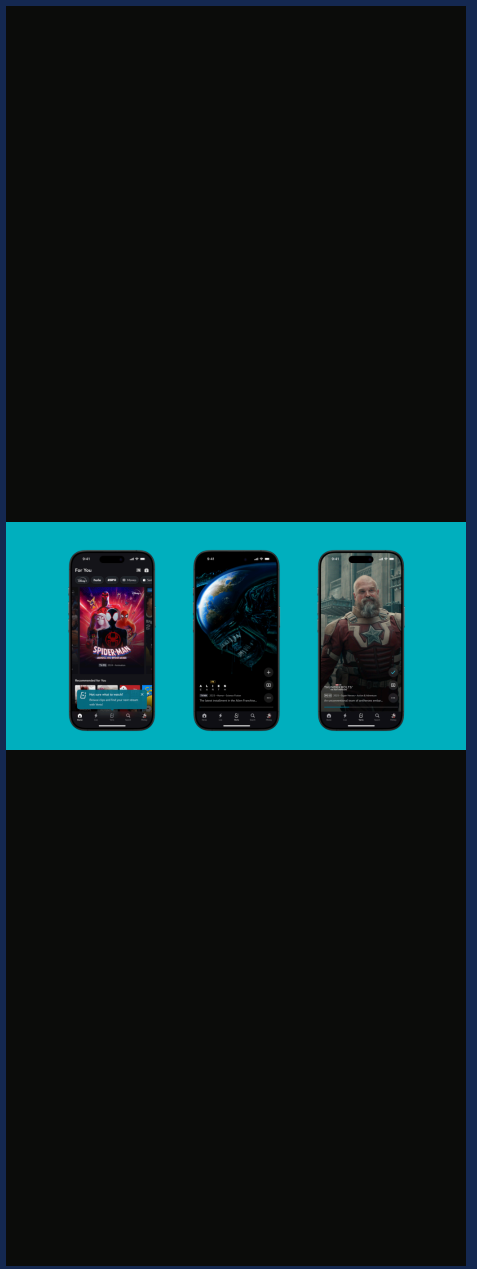
Verts 把竖屏、可连续滑动的视频带进 Disney+。当 TikTok 创作者进入这条内容线，社交平台的叙事节奏也会进入传统 IP 的发行系统。

对 AI 创作者而言，机会不只在"生成一条片"。更重要的是能否围绕角色和世界观持续更新，并让内容适配手机、电视和流媒体界面。

Disney 官方介绍 ↗



VERTS FEED



MOBILE VIEW

CAPCUT AI FESTIVAL · SUBMISSIONS CLOSE 10 AUG

CREATOR DEADLINE

最后一个周末，适合整理作品，不适合仓促重做。

CapCut CRE[AI]TE 将于 8 月 10 日截止。评审关注原创视角、叙事、工艺、类别完成度，以及 AI 是否真正参与了创作。

如果作品已接近完成，此时应把时间留给声音、字幕、片尾信息和来源记录。评审往往能看出一条片是“用了模型”，还是“完成了作品”。

投稿与规则 ↗

AUGUST CALENDAR

两场征集，对应两种作品。

10 AUG

CapCut CRE[AI]TE

适合已经完成的短片、系列、广告与创意作品。重点补齐声音、字幕和制作说明。

官方入口 ↗

23 AUG

AI Film Frontiers

SIGGRAPH Asia 相关征集，关注 AI 影像研究、制作方法与前沿案例。适合有清楚过程记录的项目。

官方入口 ↗

提交前五项

导出文件可播放

字幕与声音复核

片尾署名完整

素材权利可说明

制作方法写得清楚

X / XIAOHONGSHU / CREATOR COMMUNITIES

热度先把作品推到眼前，编辑再问它做成了什么。

本期选取三条可以公开回溯的作品。每一条都保留原帖链接，也只截取几秒作为阅读证据。小红书继续用于发现国内入口和讨论方向；登录限制下无法稳定回溯的帖子，不写进事实栏。

01 看完整作品

02 找制作方法

03 记录失误

04 不拿点赞替代判断

CLOUDSTITCHER · STARJUPI · X

SOCIAL WORK 01

一张角色参考，八格动作分镜。

《Cloudstitcher》让持针守卫在云城上迎战机械巨人。作者说明，作品以一张角色参考和八格动作分镜进入 Seedance 2.5，之后再用 Topaz 放大。

值得看的地方在于信息组织。角色参考负责身份，分镜负责动作与节奏。画面仍有生成式运动的柔软感，但高潮和方向十分清楚。

X 原帖与工作流 ↗

SOCIAL WORK 02

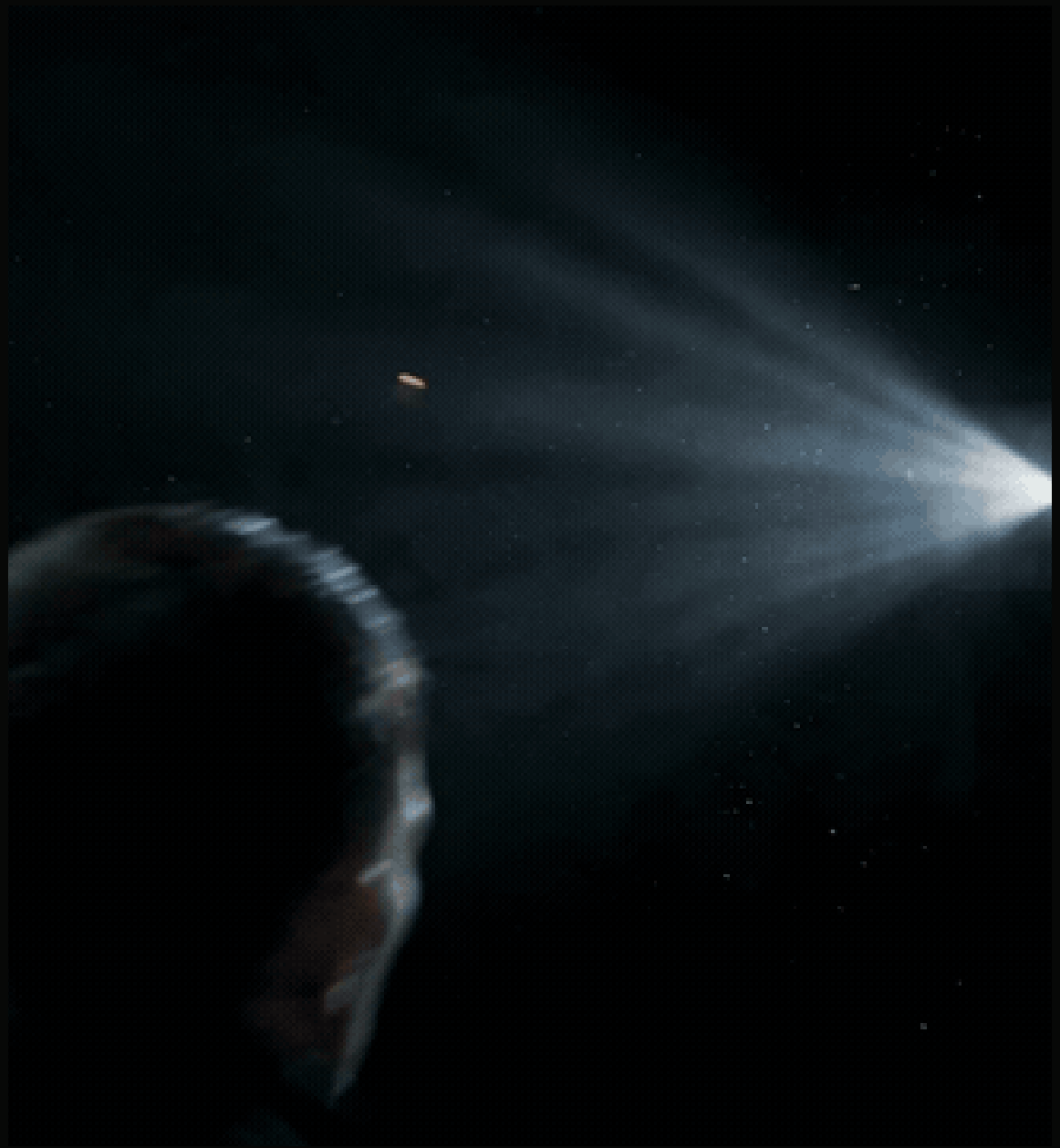
一杯咖啡，比巨兽更考验手和物。

镜头跟着人磨豆、装粉、萃取、打奶。迷你 DV 的噪点与轻微失焦替画面遮住一部分不稳定，也让日常感更自然。

这条样片提醒人们，真实感常来自有意保留的不完美。可它同样要求物体连续：手、杯子、咖啡机和液体都不能随意换位。作者把声音计划写进生成过程，这也是成片成立的重要原因。

查看原帖 ↗

ONE-PASS COFFEE VLOG · CREATOR SAMPLE



58-SECOND EXPERIMENT · STARJUPI · X

SOCIAL WORK 03

五十八秒已经足够，让一个世界露出轮廓。

雨夜、街道、海岸和旧剧场被剪成一条短叙事。单个镜头仍由 Seedance 2.5 生成，作者通过剪辑把多个段落接成长一些的时间。

它的长处在气氛与转场，人物行为相对简单。长内容的下一步仍要回答：角色为什么走进这个世界，又在其中做了什么。

X 原帖 ↗

SCREENSHOT CULTURE · CLAIMS TRAVEL FASTER THAN PROOF

MEDIA LITERACY

“一个周末赚四十万美元”，是社媒最喜欢的速度。

Tom's Guide 追踪了一批在 X 与 TikTok 流传的 AI 创业成功故事。它们常有年轻面孔、惊人收入和模糊截图，却没有足够证据让人核对业务、成本或时间。

AI 影像也会遇到同一种叙事：一条成片被说成“一次生成”，几百次失败、后期修补和付费放大消失在画外。制作报道应把过程写回来。

阅读调查 ↗

NEW STUDY

人们信任真人内容，常因为看见了共同生活。

一项新研究比较 AIGC 与 UGC 的信任来源。用户看真人创作时，会从经验、身份和社群关系里判断可信度；面对 AI 内容时，他们更看重效率、信息整理和工具能力。

这意味着披露并不会自动毁掉内容。读者真正想知道的是：谁在说，依据是什么，这段内容替我完成了什么。

查看研究 ↗

UGC

- 经验
- 身份
- 社群关系

信任来自“与我有关的人”

AIGC

- 效率
- 信息综合
- 工具能力

信任来自“它是否帮到我”

作者感不会因为工具出现而消失，它会转移到选择、说明和责任上。

GOOGLE DEEPMIND

Hassabis 离开日常管理，继续掌管更长的科学方向。

Axios 8 月 5 日报道，Demis Hassabis 将转任 DeepMind 主席及 Alphabet 首席科学家，Koray Kavukcuoglu 负责更多运营工作。Jeff Dean 等人则离开去创办新公司。

对创作者最直接的观察点，是研究成果进入产品的速度是否改变，以及视频、世界模型与代理工具的优先级如何调整。

Axios 首报 ↗ 后续分析 ↗

THREE PAPERS, PLAINLY EXPLAINED

研究正在补上制作流程里的空位。

Video-DeepResearch: Towards the Next-Generation Multimodal Deepresearch Agent

Video-Deepresearch Team
fazii@mail.ustc.edu.cn, wxhuang0616@gmail.com
https://github.com/Osilly/Vision-DeepResearch

Abstract

We introduce Video-DeepResearch (Video-DR), extending multimodal agents from static images to continuous video streams, a setting that demands dense spatiotemporal grounding coupled with open-web exploration. Preliminary evaluations reveal two critical bottlenecks in current models: (1) modality bias, where agents bypass visual tools in favor of textual search, and (2) parametric knowledge leakage, where models rely on internal memory rather than genuine tool-augmented execution. To address these challenges, we propose VIDEO-DEEPRESEARCH, featuring a decoupled perception-exploration pipeline with stage-wise tool unlocking that compels exhaustive cross-frame visual grounding prior to web retrieval. Our framework adopts a two-stage training recipe—supervised fine-tuning followed by Group Relative Policy Optimization (GRPO)—enabling autonomous exploration that breaks the imitation-learning ceiling. Furthermore, we curate VIDEODR-BENCH, a human-AI collaborative benchmark comprising 200 complex, multi-hop VQA instances. Empirical results demonstrate that our Video-DeepResearch-35B-A3B establishes a new state-of-the-art of 64.0% average accuracy, surpassing proprietary Claude-4.5-Sonnet (59.0%) by 5.0 points and significantly outperforming GPT-5 (52.5%) and Gemini 2.5 Pro (57.5%). The 30B-A3B variant achieves 59.3%, competitive with Claude-4.5-Sonnet and demonstrating the effectiveness of our training paradigm even at compact scale.

1 Introduction

Executing long-horizon tasks through active interaction marks a critical milestone in the pursuit of AGI, a capability epitomized by the recent rise of deep research agents (Li et al., 2025b; Wu et al., 2025a; Tao et al., 2025). Leveraging advanced VLMs (Bai et al., 2025; Team, 2026; Comanici et al., 2025), research agents have advanced into unstructured, vision-rich digital wildernesses (Chen et al., 2026; Huang et al., 2026; Ma et al., 2025; Wu et al., 2025b; Feng et al., 2026). This transition shifts the agent's workload from processing clean textual inputs to decoupling critical insights from noisy, heavily redundant multimedia web layouts, redefining the complexity of autonomous exploration.

Previous literature addressing these challenges generally tracks a multi-modal trajectory. Text-based web agents, notably WebGPT (Nakano et al., 2021) and AutoGPT (Yang et al., 2023), established the foundations of iterative knowledge synthesis through programmatic search. To accommodate rich visual layouts, subsequent paradigms shifted toward sensory interfaces; frameworks like WebWatcher (Geng et al., 2025) introduced various vision tools for better search, and recent Vision-DeepResearch (Huang et al., 2026) expanded this frontier to execute multi-step, exhaustive search across highly dense and complex visual contexts.

Despite these advances, current literature bypasses an ecologically valid yet far more formidable setting: ***Video-DeepResearch (Video-DR)***. This paradigm shifts the focus from isolated modalities to a holistic environment where text-based synthesis and dense visual tracking are deeply intertwined, fundamentally redefining the cognitive workload of autonomous agents. Pioneering this Video-DR frontier entails two fundamental bottlenecks that existing methodologies fail to address. First, on the data synthesis front, it remains largely elusive how to construct effective training pipelines that align with the intrinsic properties of video-based deep research. Conventional video datasets focus on localized captions or short-term action labels; conversely, Video-DR requires generative data curation that couples long-horizon decision trajectories with dense, time-varying textual and visual evidence. Second, from an evaluation perspective, establishing a rigorous and high-fidelity benchmark presents a formidable challenge.

Video-DeepResearch

让系统围绕视频继续查找、推理和回答，而非只看一个片段。

ging Creative Intent and Visual Quality: Creator-Driven Recurrent Vid
Generation with Agentic Feedback Loops

Savytski[*1] Aiden Lei[*2] Heding Liu[*3] Warren Yang[*4] Sihan Liang[*4] Alexander Liu[*3] Zhe Zha

Abstract

erative AI has made content creation increas-y accessible, but many AI-generated videos narrative coherence and creative direction. es that become more substantial at longer du-ons. Unlike coding, where AI generation bene-from reliable feedback and techniques such as urrent self-improvement, video generation re-es subjective feedback about plot, scenes, and rative, which naturally motivates approaches incorporate human creative direction. We oduce CHIEF, a human-AI co-creation video eration framework that places the creator at center of human-in-the-loop iterative video nement, and supports them by providing au-atic subjective feedback. The creator incor-ates their creative direction by driving each ation, while their revisions are incorporated specialized refiner agent. The feedback loop enerated by persona-conditioned multimodal Ms that watch generated videos and produce jective critique from the audience perspectives, viding feedback that self-evaluation alone can-capture. To test the effectiveness of our pro-ed framework, we work with high school and lege students with no prior filmmaking experi-e to create videos, from short 1-minute videos complete short 10-minute film with a compli-ed plot.

oduction

Intelligence has become widely accessible to users echnical expertise, enabling the generation of text, and videos from simple natural-language prompts.

contribution [1]Department of Computer Science, y of California, Davis, USA [2]The Harker School dependent Silicon Valley [3]Saratoga High. Correspon-Denis Savytski <dsavytski@ucdavis.edu>, Zhe Zhao edavis.edu>.

Workshop on Human-AI Co-Creativity, Seoul, South opyright 2026 by the author(s).

Increased accessibility made AI a common tool for co creation, with 21% of videos shown to new YouTube being AI-generated (Kapwing, 2025). These videos have high visual fidelity but lack narrative coherence creative direction, issues that become dramatically wor longer durations.

A specific challenge that distinguishes video genera from other fields is the inherent difficulty of objectiv evaluating videos. Generation models often do not duce correct outputs on the first try but can iteratively improve when given a signal that distinguishes good ou from bad (Madaan et al., 2023). In particular, coding automatic feedback signals such as tests and has see markable success through closed-loop self-improver Frameworks that have extended this paradigm to video eration utilize a similar closed-loop structure, substitu the missing automatic signal with human-aligned re models (Liu et al., 2025; Ji et al., 2024) or multiple L judges (Long et al., 2025). However, this feedback si is narrow and captures aggregated preferences or ge critique, rather than real viewer sentiment. This na feedback also reflects a deeper assumption: existing fr works inherit an autonomous closed-loop structure coding, whereas video generation is a creative task should enable creators to express their creative directi

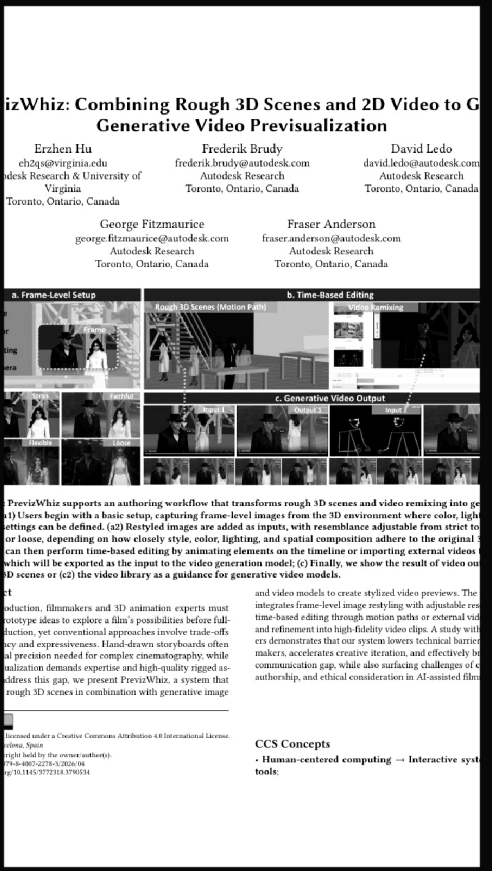
We propose CHIEF (Creator-driven Hybrid Iter Evaluation Framework, Figure 1) — an iterative v generation framework built around human-AI co-crea CHIEF is based on two main insights. First, we replac tonomous self-improvement with human-in-the-loop re ment, enabling the creator to express their creative direc Second, we use recent advances in LLM-based human ulation (Argyle et al., 2022) to support creators with div automatic feedback that captures real viewer sentimen

At a high level, CHIEF operates as an iterative loop system generates a video, agents simulate user feedb and the creator provides revisions. The **Video Gen tor** converts written text into model-specific prompts uses prompts to generate video. The **Feedback Agents** multimodal LLM agents to watch videos and simulate viewers with diverse backgrounds. The **Feedback Tr lator** provides the creator with feedback and translate

1

CHIEF

把创作者反馈放进循环，让工具从修改中学习制作偏好。



izWhiz: Combining Rough 3D Scenes and 2D Video to G
Generative Video Previsualization

Erzhen Hu
eh2qs@virginia.edu
desk Research & University of Virginia
Toronto, Ontario, Canada

Frederik Brudy
frederik.brudy@autodesk.com
Autodesk Research
Toronto, Ontario, Canada

David Ledo
david.ledo@autodesk.com
Autodesk Research
Toronto, Ontario, Canada

George Fitzmaurice
george.fitzmaurice@autodesk.com
Autodesk Research
Toronto, Ontario, Canada

Fraser Anderson
fraser.anderson@autodesk.com
Autodesk Research
Toronto, Ontario, Canada

a. Frame-Level Setup
b. Time-Based Editing
Rough 3D Scenes (Motion Path)
Video Remixing
c. Generative Video Output

PrevizWhiz supports an authoring workflow that transforms rough 3D scenes and video remixing into ge a1) Users begin with a basic setup, capturing frame-level images from the 3D environment where color, light settings can be defined. (a2) Restyled images are added as inputs, with resemblance adjustable from strict to or loose, depending on how closely style, color, lighting, and spatial composition adhere to the original can then perform time-based editing by animating elements on the timeline or importing external videos which will be exported as the input to the video generation model; (c) Finally, we show the result of video ou 3D scenes or (c2) the video library as a guidance for generative video models.

et

oduction, filmmakers and 3D animation experts must ototype ideas to explore a film's possibilities before full-duction, yet conventional approaches involve trade-offs ncy and expressiveness. Hand-drawn storyboards often al precision needed for complex cinematography, while ualization demands expertise and high-quality rigged as-ddress this gap, we present PrevizWhiz, a system that rough 3D scenes in combination with generative image and video models to create stylized video previews. The integrates frame-level image restyling with adjustable res time-based editing through motion paths or external vid and refinement into high-fidelity video clips. A study with ers demonstrates that our system lowers technical barrier makers, accelerates creative iteration, and effectively b communication gap, while also surfacing challenges of authorship, and ethical consideration in AI-assisted film

licensed under a Creative Commons Attribution 4.0 International License.
ezinna, Spain
right held by the owner/author(s).
79-8-4007-2278-3/2026/04
g/10.1145/3772318.3790834

CCS Concepts
• **Human-centered computing** → **Interactive syst tools**;

PrevizWhiz

面向预演，把文字、分镜与镜头规划连成可讨论的前期流程。

WHAT THE PAPERS ARE REALLY ASKING

机器要看懂制作，先得看懂人的修改。

看视频

理解镜头里发生了什么。

→

问问题

沿着任务继续搜索与比较。

→

接反馈

记住创作者为什么要求修改。

→

做预演

在昂贵生成前先把镜头讲清。

三篇论文看似分散，实际都在处理同一件事：让系统理解过程。未来的工具若只会交出一张结果图，很快就会显得笨重；能读懂反馈、版本和镜头意图的工具，才可能进入团队。

论文仍处在研究阶段。创作者现在可以先做好一件朴素的事：把为什么改、改了什么、哪一版有效写下来。

Video-DeepResearch ↗ CHIEF ↗ PrevizWhiz ↗

NEXT SEVEN DAYS

本周可以做的五件事

01 做一次 30 秒压力测试

用一个明确动作链，记录物体、人物和影子的连续性。

02 把提示拆成时间段

为开始、变化、结果和出口分别安排秒数。

03 保存失败版本

失败画面会告诉你模型在哪些地方失去空间关系。

04 补齐作品说明

准备投稿时，写清模型、参考、后期和声音来源。

05 把原始链接留在旁边

收藏 X 或小红书案例时，同步记录作者与检索日期。

这份清单对应本期报道中的实际缺口，七天后可复查。

WATCH NEXT

下周，看这些事情有没有落地。

01 Seedance 2.5

是否出现更清楚的官方版本说明、API、价格与地区范围。

MODEL / ACCESS

02 Flow Studio

3D Editor 与 Canvas 的真实上手成本、交换格式和公开测试。

WORKFLOW

03 Google Vids

指令式编辑是否能稳定修改局部，而不破坏其余画面。

EDITING

04 《月下之臣》

是否出现官方观看入口，以及八集长内容的实际交付形态。

CHINA AIGC

05 CRE[AI]TE

入选作品会怎样描述 AI 的参与，评审是否重视制作过程。

CREATOR ECONOMY

06 社媒样片

寻找能公开复现的方法，也继续记录高热度作品的失败部分。

X / XIAOHONGSHU

TRACEABLE READING

来源在每条资讯旁边，这里再给一张总表。

官方材料负责确认功能与日期；媒体报道负责补充语境；X、小红书与创作者社区负责发现现场。社区样片从不被写成统一评测。

01 Creative Bloq · Flow Studio

02 SIGGRAPH 2026 · AI program

03 NVIDIA · SIGGRAPH releases

04 Google Workspace · Vids

05 TechRadar · Disney × TikTok

06 CapCut CRE[AI]TE

07 Tom's Guide · fake AI success

08 X · Cloudstitcher

09 X · 58-second experiment

10 Creator sample · party to pool

11 Creator sample · coffee vlog

12 Video-DeepResearch

13 CHIEF

14 PrevizWhiz

素材说明

官方图像用于新闻报道与评论；社媒动图为短时编辑截取，原作者与原帖均在页面中标明。完整视频请前往原始链接观看。

检索截止：2026-08-07 18:00 CST。动态平台上的内容与可用性可能变化。

END OF ISSUE 005

画面可以跑得更远。
下一步，要让它知道自己从哪里来，又该往哪里去。

03-07 AUGUST 2026