arXiv:2603.07155v1 [cs.HC] 7 Mar 2026

NarrativeLoom：通过多角色协作即兴创作增强创意叙事

Yuxi Ma*[123]
人工智能研究院，
北京大学
北京，中国
yxma@stu.pku.edu.cn

Yongqian Peng*[124]
人工智能研究院，
北京大学。
中国北京
yqpeng@stu.pku.edu.cn

Fengyuan Yang[124]
人工智能研究院，
北京大学
中国北京
fyyang@stu.pku.edu.cn

Siyu Zha
未来实验室，清华大学
大学
中国北京
zhasiyu22@mails.tsinghua.edu.cn

Chi Zhang[2]
智能科学与技术学院
北京大学
中国北京
chizhang.cz@pku.edu.cn

Zixia Jia[2]
北京通用人工智能研究院
人工智能（BIGAI）
中国北京
jiazixia@bigai.ai

Zilong Zheng✉[2]
北京通用人工智能研究院
（BIGAI）
中国北京
zlzheng@bigai.ai

Yixin Zhu✉[235]
心理与认知科学学院，
认知科学，北京大学
中国北京
yixin.zhu@pku.edu.cn

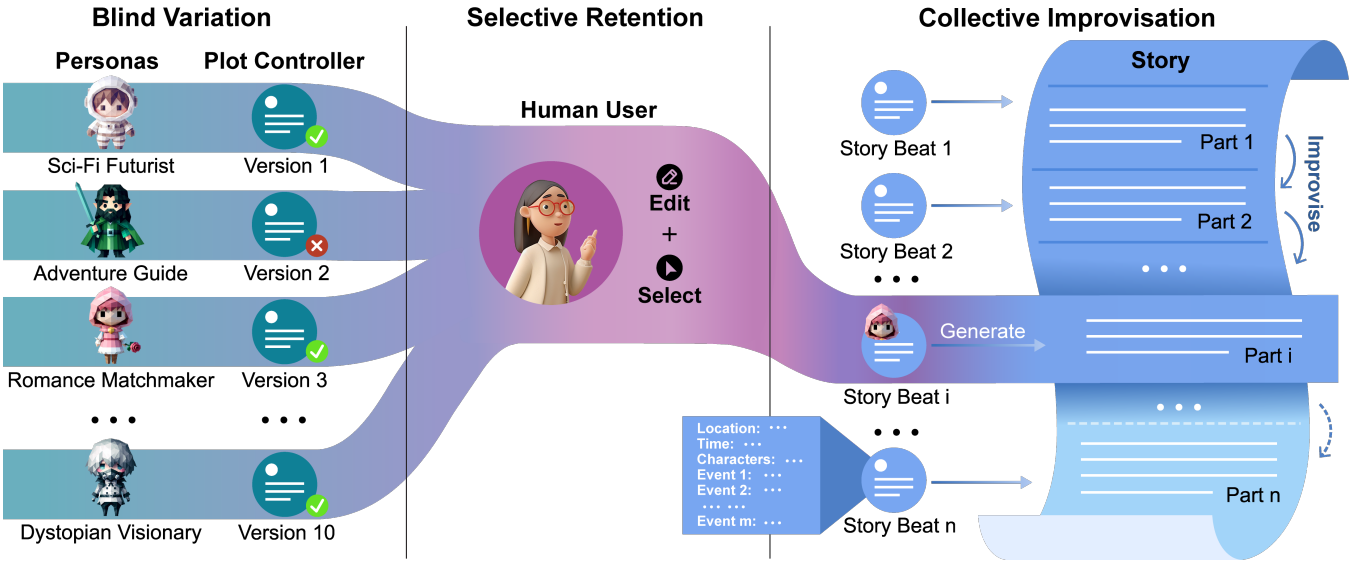


图 1：NarrativeLoom 的架构。NarrativeLoom 是一个受盲目变异与选择性保留（BVSR）理论启发的多角色共同创作系统。我们对 BVSR 的计算实现包含三个相互关联的阶段：(i)盲目变异：十个专门的讲故事者角色设定分别为每个故事节拍生成多样的叙事可能性，情节控制器则通过一致性验证确保基本连贯性；(ii) 选择性保留：人类用户评估、选择生成的备选方案中最有前景的故事节拍，并可选择进行编辑修改，从而发挥创作主导作用；(iii) 集体即兴创作：选定的故事节拍——其中包含结构化的叙事元素（地点、时间、角色、事件）——依次转化为连贯的叙事片段，并通过人类与人工智能的协作迭代，逐步构建完整故事。

*两位作者对本研究的贡献相同。[1]亦任职于北京大学心理与认知科学学院。[2]亦任职于通用人工智能国家重点实验室。[3]亦任职于北京大学北京市行为与心理健康重点实验室。[4]亦任职于北京大学元培学院。✉通讯作者[5]亦任职于北京大学人工智能研究院

允许免费制作本作品全部或部分的数字或纸质副本，供个人或课堂使用，前提是不得制作或分发副本

用于营利或商业利益，且副本须在首页载明本声明及完整引文。必须尊重本作品中第三方组件的版权。对于所有其他用途，请联系所有者/作者。*CHI '26, 2026年4月13日至17日，西班牙巴塞罗那* © 2026 版权由所有者/作者持有。
ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3790416

摘要
大语言模型在人工智能辅助讲故事方面展现出巨大潜力，但现有工具生成的叙事作品往往可预测且缺乏原创性。为解决这一局限，我们提出了 NarrativeLoom，这是一种基于坎贝尔盲目变异与选择性保留理论的多角色协同创作系统。NarrativeLoom 部署专门的人工智能（AI）角色来生成多样化的叙事选项（盲目变异），用户则作为创意总监对其进行选择和润色（选择性保留）。我们对 50 名参与者开展了一项对照研究，结果发现，与 NarrativeLoom 合著的故事不仅被用户认为更新颖、更多样，也经专家客观评定，在托兰斯创造力测验的所有创造力维度上均显著更佳：流畅性、灵活性、原创性和阐述。故事篇幅显著更长，场景设定更丰富，对话也更多。写作专长表现为一个调节变量：初学者从结构化支架中获益更多。这表明理论指导的协同创作系统具有重要价值，也凸显了根据不同用户专业水平进行调整的重要性。项目页面：https://ppyyqq.github.io/narrativeloom。

CCS概念
• **以人为中心的计算** → **交互设计系统与工具**;**协作交互**; 协作与社会计算系统及工具; **人机交互（人机交互）**。

关键词
人机交互、大语言模型、叙事、创意支持工具

ACM参考文献格式：
Yuxi Ma、Yongqian Peng、Fengyuan Yang、Siyu Zha、Chi Zhang、Zixia Jia、Zilong Zheng和Yixin Zhu。2026。NarrativeLoom：通过多角色协作即兴创作增强创意叙事。载于*2026*年*CHI*计算机系统中的人为因素会议论文集（*CHI '26*），*2026*年*4*月*13*日至*17*日，西班牙巴塞罗那。ACM，美国纽约州纽约市，19页。https://doi.org/10.1145/3772318.3790416

1 引言
叙事是人类文化的基石，也是分享知识、价值观与经验的主要媒介 [39, 58]。从亚里士多德的《诗学》 [4] 到现代剧本，有效的叙事都依赖于连贯性与惊喜之间的平衡。它们需要“突转”——出乎意料的情节转折，这些转折吸引观众的方式并非依靠随机性，而是揭示更深层、潜藏的逻辑 [26]。新颖性与恰当性的结合体现了人类创造力，也是计算创造力研究的核心关注点 [28, 50]。

大语言模型（LLM），如 GPT-4 [1] 和 Gemini [60]，已经改变了计算叙事生成的格局。与早期的基于规则的系统相比，这些模型展现出更高的流畅性和语境理解能力 [2, 63]。然而，人机交互（HCI）领域的研究者指出，当前的人工智能叙事系统往往仍然较为保守 [10, 16, 43, 68]。由于其架构经过优化，旨在进行下一个词元预测，它们擅长生成具有统计概率的续写，但往往会抑制对创意深度至关重要的出人意料的偏离 [10, 11, 26, 28]。行业专业人士观察到，这一局限 [3, 15, 16, 19, 43] 往往会导致可预测的、陈词滥调式的叙事 [11, 35, 50, 67],，进而造成想法同质化。

我们开展了一项形成性研究，探究写作者如何感受这些局限。研究揭示了对创作多样性的渴望与对叙事控制的需求之间存在张力。当前工具往往无法化解这一问题；它们要么提供可预测的发展路径，要么生成与写作者构想脱节的内容。这凸显了人机交互领域的一项关键挑战：如何让基于大语言模型的系统在保持用户能动性的同时，生成有意义的惊喜。

为满足这些需求，我们采用了坎贝尔的盲目变异与选择性保留（BVSR）理论 [8]。BVSR描述了一个两阶段过程：先生成不受约束的想法（“盲目变异”），再有意筛选出有潜力的想法（“选择性保留”）。这一框架为系统提供了理论蓝图，使其能够提供更广泛的可能性，同时保留用户的创作主导权。基于这一基础，我们开发了NarrativeLoom，通过多角色协作即兴创作将BVSR付诸实践 [52, 54]。为实现“盲目变异”，系统采用由专门化AI角色组成的协作团队，每个角色都运用独特的体裁感知叙事视角来生成多样的故事节拍。为促进“选择性保留”，用户则作为创意总监，负责挑选、编辑并整合这些故事节拍。我们的方法与传统工具的不同之处在于：（i）将生成过程与筛选过程分开；（ii）通过专业化角色而非参数化采样实现变异；（iii）在支撑探索性的同时保留用户的主导权。

我们的研究结果表明，与单声部聊天机器人相比，NarrativeLoom 基于 BVSR 的多角色设定方法显著提升了创作成果，生成的故事更长、内容更丰富，场景设定更多，对话比例也更高，从而突出了“展示胜于讲述 [33]”。用户认为 NarrativeLoom 能提供更多样的叙事可能性，同时保持较高的可用性；对角色设定的策略性运用还揭示了不对称的转变模式，其中某些角色设定充当“发起者”，另一些则充当“发展者”。专业写作专家对 NarrativeLoom 生成的故事在所有创造力维度——流畅性、灵活性、原创性和阐述——上的评分都显著更高，并称赞它能够创造出意想不到的叙事转折和心理层次复杂的角色。写作经验是一个重要的调节因素：新手写作者从 NarrativeLoom 的结构化支架中获益更多，这表明它尤其有助于培养写作者的叙事直觉。

本研究为人机交互与计算创造力领域提供三项主要贡献：（i）一个**理论框架**，将坎贝尔的盲目变异与选择性保留（BVSR）理论落实于人机协同创作之中，提供一种原则性机制，以平衡人工智能驱动的多样性与人类的执行控制；（ii）一种**系统设计**，采用多角色集体即兴创作架构，通过专业化角色的功能异质性而非随机参数采样来构建创造性变异；以及（iii）**实证洞见**表明，写作专长与创作阶段

均会显著调节系统的有效性，因而需要自适应界面，以适配不同的技能水平与不断变化的创作情境。

2 相关工作

2.1 叙事中的人机共创

叙事中的人机共创已从基础的创造力支持，转向在写作者与计算系统之间重新分配能动性的协作伙伴关系。近期的人机交互研究考察了这些互动的社会与认知动态。例如，Gero 等人 [27] 展示了写作者如何在将人工智能视为工具还是协作者之间把握边界，而 Dhillon 等人 [19] 发现，不同的脚手架层级会同时改变创作过程与叙事质量。尽管大语言模型具有很高的流畅性，专业作家仍对创意同质化，以及在 Wordcraft [35, 68]等系统中维持真实声音的困难表示担忧。

为应对这些局限，人机交互社区发展出了超越文本生成、提供更多控制的交互范式。Chung 等人 [15] 提出了 TaleBrush，该系统通过线条级草图来塑造故事弧线，表明视觉模态能够有效引导创作结果。针对结构控制的需求，Lu 等人 [41] 提出了 WhatELSE，这一系统允许作者通过规划来管理叙事抽象，从而确保开放式情节的因果合理性。同样，Mirowski 等人 [43] 主张在剧本创作中，工具应作为创意伙伴而非生成器；而 Chakrabarty 等人 [11] 则研究了迭代式编辑界面，以使人类意图与模型输出保持一致。

尽管取得了这些创新，在长篇叙事中维持结构连贯性仍然很困难。层级式方法通过自上而下规划来应对这一问题， [66], Mirowski 等人[43] 认为，僵化的结构与写作的涌现性相冲突，往往会产生可预测的模式。Chakrabarty 等人[10] 将此现象描述为“创造力的虚假承诺”：为下一词预测优化的模型会陷入“概率陷阱”[26]。这会导致内容在技术上虽娴熟，却趋于保守，缺乏引人入胜的叙事所必需的偏离。连贯性与多样性之间、能动性与辅助之间的这些张力表明，我们需要既能支持叙事多样性、又能保留作者控制权的方法。

2.2 即兴叙事

即兴叙事是指在没有预设要素的情况下构建即兴故事 [52, 54]。这一过程需要持续适应，并会带来不确定性，因此有别于层级式方法 [29, 59]。这种框架与分布式创造力理论 [53, 54], 相契合，该理论认为，创造力源于交互动态，而非孤立的行为。人机交互领域已将这些原则应用于协作系统；例如，Kim 等人 [38] 开发了 Ensemble 系统，在领导者与群体之间非对称地分配创意责任。

在 Ensemble [38], 中，主导作者通过“场景提示”维持宏观构想，而贡献者则在指定边界内生成内容。这表明，战略性约束可以将注意力集中在特定叙事元素上，从而支持创作自由。我们的系统通过将故事节拍用作导航约束，并以专门化AI角色取代人类群体，扩展了这一原则。这种配置将领导者与贡献者之间的互动转变为一种由特定创作声音调节的人机伙伴关系，使即兴涌现得以在结构化框架内发生。

近期研究考察了人工智能在分布式创造性流程中的能力。Wang等人 [64] 通过独奏表演提示（SPP）研究了“认知协同”，其中单个大语言模型采用多个协作角色。他们的研究结果表明，使用专业化角色能够减少幻觉，并在复杂任务中保持推理能力。虽然 SPP 评估的是任务解决能力，而非叙事生成，但它为将分布式创造力应用于协作即兴叙事提供了计算基础。

3 形成性研究

为了解人机协作叙事中的挑战，我们开展了一项形成性研究，对不同经验水平的写作者进行了深入访谈。我们的目标是找出写作者需要支持的具体方面，从而为设计用于协作叙事创作的基于AI的系统提供依据。

3.1 方法

3.1.1 参与者。写作者是不同媒介叙事创作的主要实践者。我们通过专业网络招募了五位参与者，以确保经验范围多样，涵盖新晋写作者（2 年）至资深专业人士（15+ 年；见表 1）。所有参与者此前都使用过 AI写作工具，主要用于头脑风暴、研究和故事开发。

3.1.2 实验流程。我们通过视频会议进行了半结构化远程访谈，每场约一小时。实验协议考察了参与者在故事发展各阶段的经验：构思、规划、起草和修订。我们探讨了他们当前对人工智能工具的使用情况、具体策略、挑战，以及他们对人工智能如何支持创意叙事的看法。参与者讨论了协作写作经验，以及他们对在人机伙伴关系中保持自主性的期望。

3.2 主要发现

分析发现，写作者在协作叙事中寻求支持的领域有四个：

通过分段单元管理叙事结构。参与者在管理长篇叙事的连贯性和创作势头时始终面临困难，尤其是在平衡即兴创造力与结构组织方面。W1 将创作过程描述为“螺旋式推进”，写作者“反复打破简单的自上而下的线性结构”，。W4 强调，引人入胜的写作包含“随机性”，而这种随机性“无法用高概率经验来解释”，这表明有效的结构支持必须能够容纳意料之外的创作发展。写作者自然而然地采用分段式方法来应对这种复杂性。W2 使用“故事节拍”

表 1：参与者人口统计信息和专业背景。此表详细列出了五位研究参与者的性别、年龄、职业、教育程度和写作经验年限。

ID	性别	Age	职业	教育程度	写作经验（年）
W1	Male	34	编剧	电影制作艺术硕士	8
W2	Male	34	作家	创意写作文学硕士	15+
W3	女性	38	电影制片人／编剧	文学学士	10+
W4	女性	27	自由撰稿人	人类学文学硕士	2
W5	Male	32	内容创作者/教育者	医学学士	7

作为离散的结构单元，在不限制探索性的前提下维持叙事推进。W1 指出，虽然“短篇故事不需要如此严格的结构要求”，但较长的作品可以从“将其拆解为熟悉的故事结构，例如三幕结构或八段式结构”中获益，这些结构提供了导航路标，却不会规定创作内容。

寻求超越单声部叙事的多元视角。所有参与者都表示，单声部人工智能系统存在创作局限，认为这类系统生成的内容重复，缺乏真正新颖的叙事元素。W2解释说，现有人工智能系统“只是继续你正在做的事情”，而不是为故事带来“新的节拍或新元素”。写作者重视在创作过程中接触多重视角。W5主张来自不同学科背景的“多个观点”，而W3强调，个人的创作能力天生“有限”，并且能够从多元输入中受益。W1描述了人工智能在“寻找一个伙伴”方面的潜力，这个伙伴可以提供“不同的可能性”并进行创造性的“组合”，这表明有效的AI叙事系统应提供多样化的创意备选方案。

在人工智能辅助过程中保留创作所有权。参与者虽然认可人工智能辅助，但强烈强调要维护创作所有权并保留即兴创作能力。W3清楚地表达了这一需求：“我是做出判断的人，我是做出最终决定的人……所以这是我的故事。”然而，参与者批评了当前人工智能工具低效的交互模式。W2描述了“不断提示它大约半个小时”却没有产生有益结果的挫败感，凸显出需要更有效的协作工作流，在提供必要支持的同时保留创作主体性。

协调多声部贡献者之间的一致性。在管理多个创意贡献者之间的连贯性时，挑战随之而来。W2 强调了协作中的困难，并指出不同的写作风格可能会“彼此冲突”。当多个声音共同参与同一叙事时，维持一致的角色发展、情节逻辑和主题连贯会变得愈发复杂，需要超越个人写作过程的谨慎协调。W3 指出，人工智能在长篇协作作品中维持一致性方面存在局限，并称当前系统生成连贯长篇内容的能力有限。这一挑战不仅涉及基本的事实核查，还包括在漫长的协作创作过程中协调更深层的叙事要素，例如人物发展弧、主题一致性和语气维持。

3.3 设计目标

基于这些发现，我们确定了人类与人工智能协作式故事创作的四个设计目标：

DG1：通过叙事单元构建创意发展过程。 系统应将故事创作分解为离散且定义明确的叙事单元（例如故事节拍）。这些单元可作为创意路标，让写作者在长篇叙事中兼顾即兴发挥与结构组织。

DG2：通过多元叙事声音拓展创意探索。 为克服单一模型生成内容的重复性，系统应采用多个各具专长的创意声音。这些声音应提供出人意料的方向和多元视角，以增强探索性。

设计目标3：通过选择性控制赋予用户能动性。系统应将用户定位为创意总监，由他们评估、选择并润色人工智能生成的元素。相关机制应增强人类决策，而非完全自动化创作过程。

DG4：支持协作输入中的叙事连贯性。 系统应监测并支持多元贡献之间的叙事连贯性。它必须帮助写作者识别并解决角色弧光和主题元素中的不一致之处，以确保整个协作过程中的质量。

4 系统设计

4.1 理论基础

基于形成性研究，我们发现需要一个能够平衡创意多样性、人类能动性与叙事连贯性的框架。我们采用了坎贝尔的盲目变异与选择性保留（BVSR）模型 [8], ，该模型将进化原理应用于创造力。BVSR认为，创作过程需要两个截然不同的阶段：生成多种备选方案（盲目变异），以及系统地评估并保留有前景的选项（选择性保留）。

盲目变异：这一阶段涉及独立于既有模式或统计可能性来生成备选方案。这可以防止系统收敛到可预测的输出，而这正是大语言模型下一个词元预测已知的局限。通过借助专业化角色实现变异，系统能够在结构和因果层面产生多样性，而不只是改变表层语言特征。这一方法直接支持设计目标*2*。

选择性保留：在这一阶段，我们会评估并保留有潜力的变体。由于这需要结合上下文进行理解以及领域专业知识，系统将这一职责交给用户。这样一来，用户便成为

主要决策者，从而支持设计目标 *3*。在此框架中，人工智能提供生成支持，而人类用户则引导叙事走向。

迭代循环：盲变异与选择性保留理论认为，创造力源于反复的变异—选择循环。每次选择都会为后续变异提供叙事语境。这一观点启发了我们的基于故事节拍的架构（设计目标 *1*）：系统针对每个故事节拍生成多个变体，由人类选择来引导叙事推进。这种结构通过算法变异与人类策展的相互作用来实现连贯性，而非仅依赖预设的叙事约束，从而支持设计目标 *4* 。

4.2 设计原则

本节详细介绍如何通过设计决策，在协同创作叙事系统中实现盲变异与选择性保留理论。

4.2.1 用户工作流设计。 NarrativeLoom 采用三阶段工作流程（见图 2），以**故事节拍**为基础——故事节拍是叙事推进的离散单元，包含场景设定、角色和事件。在剧本创作中，故事节拍是叙事弧的基本组成部分 [44], ，因此是变异—选择动态过程的有效单元。

探索与构思： 用户提供初始叙事输入（“火花”）并定义故事参数。这将建立初始状态，同时为后续变化保留广阔的探索空间。

协作式故事创作： 此阶段通过多角色人格生成实现变化。十个专业化角色同时生成不同的节拍备选方案。界面会显示每项提案的结构细节和设计理念。随后，用户根据叙事目标评估这些备选方案，并选择性地保留其中内容。用户可以修改选定的节拍，然后将其扩展为 800–1000 字的散文。

迭代： 用户重复生成与选择流程，以构建叙事。系统整合先前的上下文，同时保持生成的多样性。基于检索增强生成的一致性系统会将故事历史编码为语义嵌入。在生成新的节拍时，该机制会识别逻辑不一致，并调整排序，优先呈现连贯的选项，同时保留其他备选方案。

4.2.2 故事节拍架构 *[DG1]*。

设计理念。 为便于管理，叙事被组织为离散的故事节拍。我们的设计优先考虑语义透明度，将每个叙事片段拆分为场景设定、角色和关键事件。这种模块化设计支持精确修改——例如更改某个角色的特定动作——而无需重新生成整个场景。这种精细度为模型提供了清晰的因果锚点，并确保上下文连贯性。

4.2.3 多角色设定生成，促进创意多样性 *[DG2]*。

设计理念。 为提高输出多样性，系统采用基于体裁而非基于风格的角色设定。以风格为导向的方法主要影响词汇选择，而基于体裁的角色则影响叙事逻辑与因果结构。体裁决定事件的类型以及叙事的底层逻辑。例如，悬疑 角色设定整合了信息不对称和线索策略性布置等结构要素。

角色设定设计。 我们依据三项标准选取了十种角色设定：

- **类型覆盖**：以成熟的文学框架为基础，涵盖主要叙事类别 [21]。
- **叙事差异化**：这些角色设定涵盖了不同的创作方式，包括情节驱动型（例如，冒险向导）、角色中心型（浪漫撮合师）以及世界构建角色（奇幻世界构建者）。
- **互补功能**：各角色设定侧重于不同的叙事元素，例如氛围（恐怖氛围营造师）或社会评论（反乌托邦构想者）。

这种多样性使 NarrativeLoom 能够同时探索叙事空间中的不同区域，而不是围绕单一叙事轨迹聚集，仅作细微变化。完整的角色设定规格详见表 2。

4.2.4 面向创作主体性的选择性保留界面 *[DG3]*。

设计理念。 该界面通过多层设计，将用户定位为最终评估者。这种设计能够根据用户的具体创作目标，支持不同程度的参与。

交互设计。 NarrativeLoom 通过三种模式保留创作主体性：

- **直接选择：** 这种设计允许写作者一键采用偏好的故事节拍，从而支持流畅的起草流程。该模式适用于创作势头强劲的时刻，此时首要目标是快速推进，而非进行细致斟酌。
- **比较评估：** 这种设计通过并排比较，鼓励用户进行反思性决策。为尽量降低认知负荷，界面会同时呈现两个备选方案，并通过渐进式披露提供查看完整候选内容的途径。
- **多阶段编辑：** 这种设计将人工智能生成的建议视为启发性的起点，而非定稿文本。它允许写作者润色单个故事节拍和扩展叙事作品，确保最终输出根植于作者独特的表达风格与创作意图。

4.2.5 通过软约束进行一致性管理 *[DG4]*。

设计理念。 管理连贯性需要在叙事稳定性与创造性变化之间取得平衡。我们没有采用会自动过滤掉不一致选项的硬约束，而是实现了一个排序系统。该系统优先考虑逻辑上连贯的路径，同时允许用户在选项有特定创作目的时选择偏离叙事的路径。

4.3 技术实现

4.3.1 节拍生成管线。

技术架构。 节拍生成采用三层提示架构（见图 3）。**元提示层**

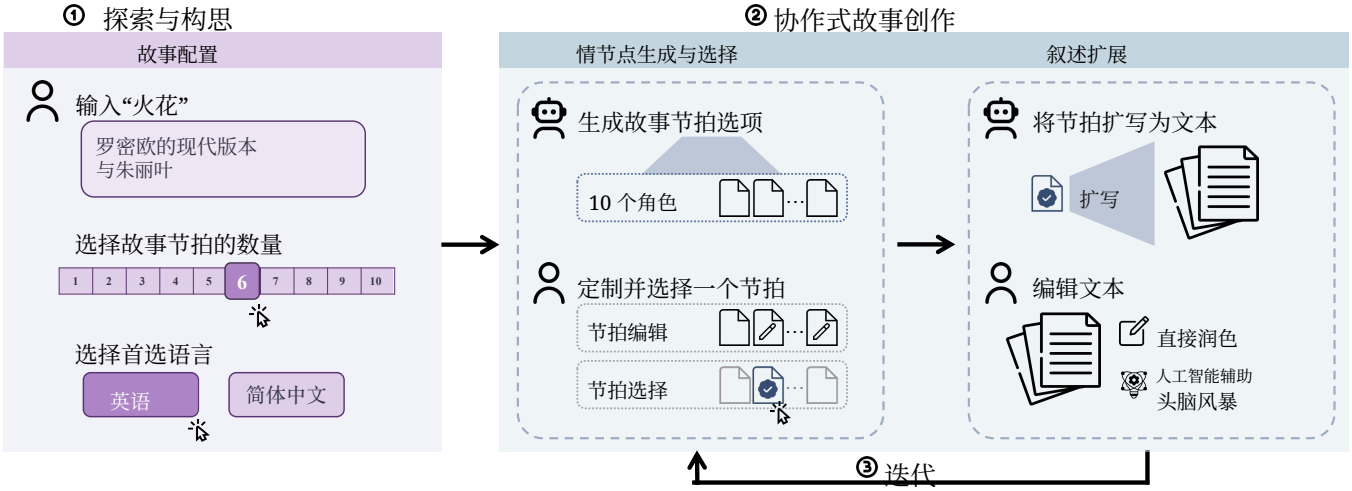


图 2：NarrativeLoom 的用户工作流程。该流程由三个相互衔接的阶段组成：（i）探索与构思，用户通过输入“火花”并选择语言和故事长度等参数来初始化叙事；（ii）协作式故事创作，系统利用 10 个各具特色的角色生成多样的情节点选项，使用户能够定制、选择并将情节点扩展为完整叙事，还可在迭代至下一个故事节拍之前编辑文本；（iii）迭代，用户通过重复节拍选择和叙事扩展流程，逐步构建故事。

表 2：NarrativeLoom 中的讲故事者角色设定。十个专业化的讲故事者角色设定构成 NarrativeLoom 的生成式 Ensemble，每个角色都具备独特的类型专长和叙事能力。

角色设定	叙事专长
奇幻世界构建者	专注于构建丰富而富有想象力的奇幻世界，其中包含错综复杂的魔法体系和神话生物，以及多元的文化。
科幻未来学家	专注于打造可信且富有创新性的科幻设定，融入先进技术、太空旅行和未来社会。
悬疑解谜师	协助构思复杂而引人入胜的悬疑故事，巧妙埋下线索、误导线索和情节转折让读者一直沉浸其中，直至故事迎来结局。
浪漫撮合师	擅长创作引人入胜的浪漫情节线，确保角色化学反应真实自然，并且让关系随着叙事推进自然发展。
历史研究员	擅长将准确的历史细节和背景融入叙事作品，让历史小说焕发生机并使读者沉浸于特定的时代背景之中。
恐怖氛围营造师	通过运用描写性语言和节奏，帮助恐怖叙事作品营造紧张感和悬念，创造令人不安的氛围元素，从而提升读者的参与度。
冒险向导	专注于创作扣人心弦的冒险故事，设计精彩的动作场面、危险的障碍，以及推动角色发展的高风险挑战。
喜剧幽默家	专注于在叙事作品中融入幽默与机智，运用文字游戏、情景喜剧和角色互动来提升叙事体验。
反乌托邦构想者	擅长构建反乌托邦背景，并探索其社会和政治影响，助力创作发人深省且具有警示意义的叙事框架。
魔幻现实主义 conjuror	协助将奇幻元素融入日常现实，创作出既扎根现实又充满奇趣的叙事作品，在普通情境中呈现非凡事件。

为多角色设定系统建立协作式故事讲述框架。**上下文整合层**将压缩后的故事历史与当前节拍状态相结合，以维持叙事连贯性。**生成约束层**规定输出的结构要求和连贯性标准。

数据模式与结构化输出。 每个故事节拍都会生成为一个结构化 JSON 对象，包含三个关键字段：用于表示时空背景的故事背景、活跃参与者角色，以及包含 3–5 个关键行动的关键_事件。这种结构化形式既便于一致地处理，也能在用户选择时保持可理解性。系统默认采用 Hauge [30]的框架，设置为六拍结构，不过用户可以配置为 1–10 拍。节拍的复杂程度会根据叙事位置调整：开篇节拍包含 3–4 个事件以进行世界构建，而高潮节拍则扩展至 4–5 个事件，以增强戏剧张力。用户还可以通过交互式优化进一步调整复杂程度。

特定阶段的生成逻辑。 Sparkle 到节拍生成会将用户最初的叙事种子转化为第一个故事节拍，后续的节拍生成则在已确立的故事元素基础上继续构建。

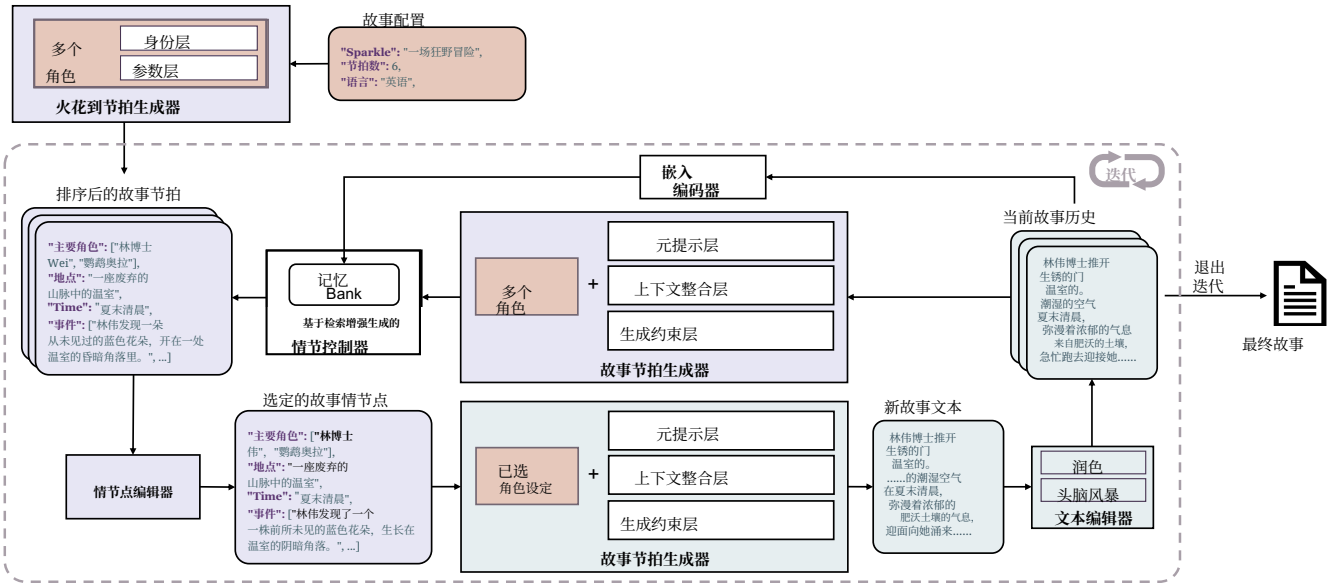


图3：NarrativeLoom的技术流程。该系统通过多角色人格生成将用户火花转化为故事节拍，并在节拍生成与文本生成两个阶段采用三层提示架构（元提示、上下文整合、生成约束）。基于检索增强生成的情节控制器确保叙事一致性，同时用户通过节拍选择和文本润色进行迭代。紫色表示节拍级操作，绿色表示文本级操作。矩形表示功能模块，圆角表示数据。

文本扩展流程。 用户选择一个节拍后，系统会将其扩展为 800–1000 字的叙事文本。扩展过程采用相同的三层提示架构，并将选定的角色设定作为参数传入。这种参数化方法无需使用多个特定角色人格生成器，从而降低成本和延迟。在整个生成过程中，系统会保留最多 8000 个 token 的叙事历史，以保持上下文连贯性。

4.3.2 角色实例化与并行生成。

基于提示词的角色实例化。 每个角色设定都通过多层提示词模板来定义。**身份层** 确立了类型特定的写作理念与创作重点，而**参数层** 则定义了叙事要素的定量约束（例如，词汇多样性、对话与叙事比例）。这种结构并非提供静态的内容指令，而是在更深层次上塑造角色设定的生成行为。每个角色设定都作为对GPT-4o的一次独立调用来实例化，这些层在各次生成中驱动一致的、特定于角色设定的变化。

并行 *API* 协调。 不同于依次生成，全部十个角色通过并行 API 调用同时生成故事节拍。这种并行架构确保每个角色都能独立生成备选方案，不受其他输出的影响。每次 API 调用都包含完整的故事上下文（最多 8000 个令牌的叙事历史）以及角色专属指令（约 500 个令牌的角色定义提示词）。

4.3.3 界面架构与交互机制。

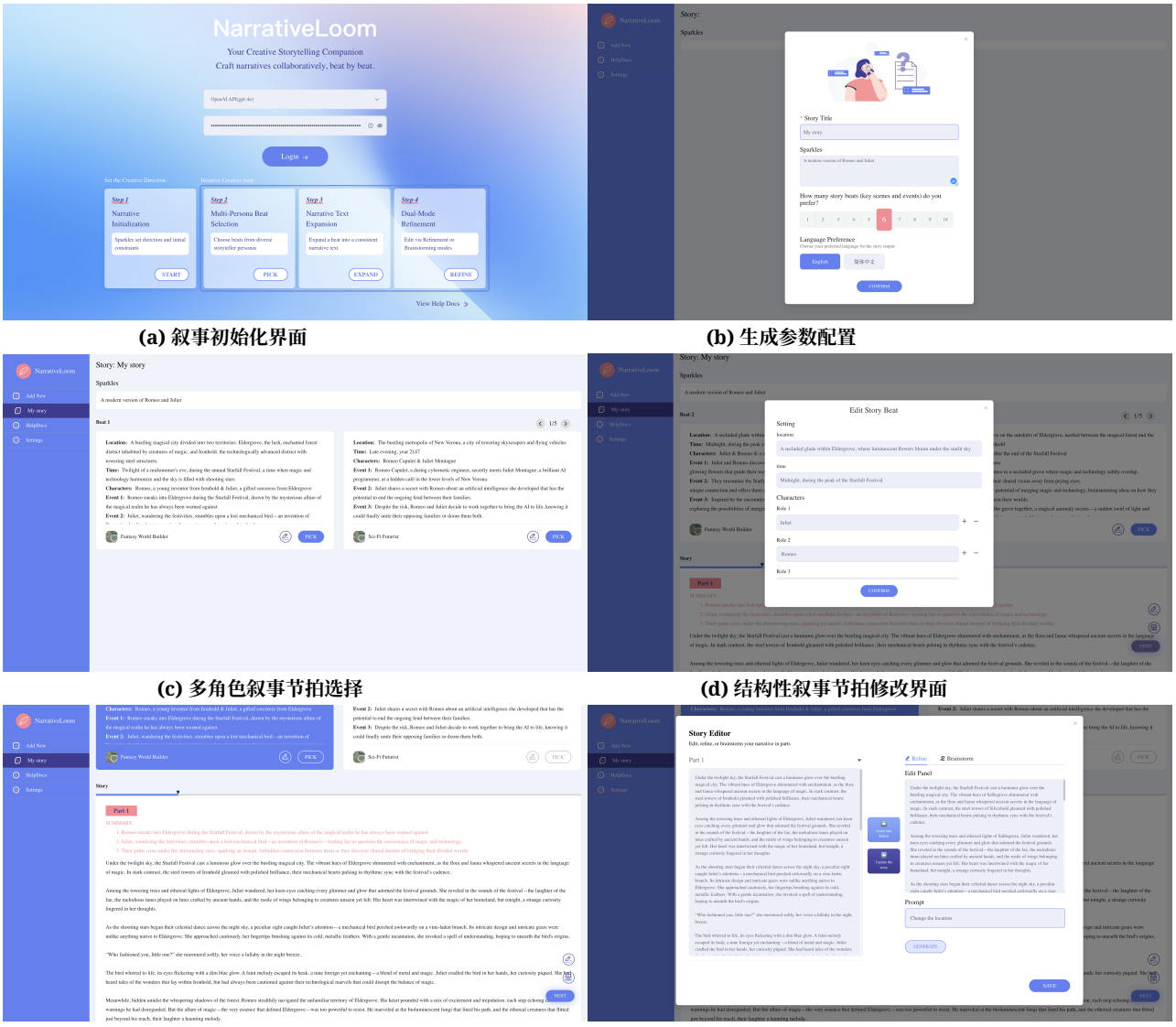
• **头脑风暴模块：** 该模块通过允许用户与人工智能互动以生成叙事灵感，促进开放式探索。它作为对话伙伴，帮助用户克服创作瓶颈，探索其他情节发展，或进一步阐述角色动机，而无需直接修改当前草稿。

• **润色模块：** 该模块通过允许用户提交具体的修订请求，提供直接的编辑支持。用户可以指示人工智能改写现有散文，以调整节奏、丰富描写细节，或修改对话以增强角色一致性。通过迭代修订，确保最终文本体现用户具体的风格与叙事意图。

界面概览。 该界面使用 Streamlit 构建，并采用三面板布局，以支持多角色叙事探索（见图 4）。左侧边栏用于管理故事作品集，中央工作区展示由角色生成的内容以供比较，下方面板则通过双模式编辑支持叙事完善。在中央工作区中，每个角色的节拍提案都显示在专属面板内，其中包括角色特定理由和模板化内容。系统会将当前会话数据保存在内存中，同时将用户的故事作品集持久化存储。

节拍选择显示。 中央工作区使用基于检索增强生成的情节控制器（详见第 4.3.4 节），按一致性对节拍提案进行排序。界面优先显示一致的选项，同时保留全部十个角色设定的输出，供用户探索。

编辑与润色。 系统支持在两个层级进行编辑，以确保整个创作过程中实现人机对齐。**节拍级编辑** 允许用户在文本扩展前修改场景设定、角色和事件，这些结构性修改会贯穿整个生成管线。**文本级编辑** 可通过人工干预或两个专用的人工智能模块来润色散文：



(a) 叙事初始化界面 (b) 生成参数配置

(c) 多角色叙事节拍选择 (d) 结构性叙事节拍修改界面

(e) 叙事文本扩展 (f) 双模式优化机制

图 4：实现 BVSR框架的 NarrativeLoom 界面。六个面板展示了系统的功能组件：(a) 叙事初始化界面；(b) 生成参数配置；(c) 多角色叙事节拍选择；(d) 结构性叙事节拍修改界面；(e) 叙事文本扩展；(f) 双模式优化机制。该工作流支持迭代变异与选择，同时保留人类创作主体性。

4.3.4 基于*RAG*的情节控制器实现。

技术实现。 情节控制器通过一致性检查系统维持叙事连贯性，该系统借助 LlamaIndex [40]实现检索增强生成（RAG）。系统结合语义检索与语言模型推理，检测新生成的叙事节拍与既有故事元素之间的逻辑不一致。

故事历史索引。 故事历史通过基于向量的记忆机制进行索引。每个已完成的故事片段都被视为独立文档，并使用 OpenAI t ext-embedding-ada-002 模型进行编码。此过程会生成 1536 维嵌入向量，用于表示此前叙事发展的语义内容。这些嵌入向量存储在 LlamaIndex向量存储中，使系统能够在评估后续故事节拍时检索相关的历史上下文。

一致性检查流程。对于角色设定生成的每个新故事节拍，系统都会执行上下文检索和逻辑验证。故事节拍内容会转换为查询并进行嵌入，然后使用余弦相似度与向量存储库进行比较：

$$\mathrm{sim}(\mathbf{v}_{\mathrm{beat}}, \mathbf{v}_i) = \frac{\mathbf{v}_{\mathrm{beat}} \cdot \mathbf{v}_i}{\|\mathbf{v}_{\mathrm{beat}}\| \|\mathbf{v}_i\|},$$

其中，$\mathbf{v}_{\text{beat}}$ 是验证查询的嵌入，$\mathbf{v}_i$ 代表第 i个故事历史嵌入。系统会将检索到的上下文连同以下提示词提供给 GPT-3.5-turbo：“新故事节拍为：*[节拍内容]*。新故事节拍中的事件是否存在逻辑错误？请简要回答 *[是]* 或 *[否]*。如果回答 *[是]*，请简要描述错误。”大语言模型会检查该节拍是否与先前事件、角色状态（例如，未作解释的复活）、时间顺序或既定的世界规则相矛盾。如果发现错误，系统不会丢弃该节拍，而是将其标记为不一致，供用户查看。

5 用户研究

我们开展了一项被试内用户研究，以评估 NarrativeLoom 对协作式故事创作的支持效果。我们将所提出的系统与传统的单角色聊天机器人界面进行比较，以考察其对创作过程、叙事质量和用户参与度的影响。我们假设，与基线聊天机器人相比，NarrativeLoom 提供的结构化指导和多样化叙事选项能够提升叙事能力。

5.1 参与者

参与者通过 Prolific 招募，资格要求为年满 18 岁、英语为母语且至少拥有本科学位。设定这一标准是为了确保参与者具备完成复杂叙事任务所需的熟练程度。在最初招募的 109 人中，有 94 人通过了强制熟悉度测试；该测试作为流程性核查，用于确认参与者理解界面操作方式和核心叙事规则。

为确保数据质量，若参与者与系统互动时间少于 20 分钟，或在 NarrativeLoom 条件下完成的故事节拍少于两个，则将其排除。经过这一筛选流程，最终有 **50 名参与者** 纳入分析。最终队列中有 24 名女性、26 名男性，平均年龄为 34.82 岁（SD = 10.03，范围：22–71 岁）。报告的族裔包括白人（n=35）、亚裔（n=7）、黑人（n=6）、混血（n=1）和未披露（n=1）。

参与者的教育背景多元，涵盖艺术与人文学科（n=11）、商科（n=11）、自然科学（n=10）、社会科学（n=10）及其他领域（n=8）。在学位方面，33 名参与者持有本科学位，15 名持有研究生学位，2 名持有博士学位。报酬为每小时 9 英镑。本研究的实验协议经所在大学的机构审查委员会（IRB）批准。

5.2 研究流程与实验协议

我们采用被试内设计，让参与者分别与实验系统（NarrativeLoom）和基线系统进行交互。为减轻顺序效应，我们在参与者之间对系统呈现顺序进行了平衡 [49]。研究分为以下五个连续阶段：

1. **人口统计数据与背景（10 分钟）：** 参与者报告了人口统计数据、写作经验和体裁偏好。

2. **创意热身（5 分钟）：** 参与者拟定了初步的故事构思（“火花”），并用两分钟写了一篇短篇故事，以建立无辅助创作的基线。.

3 **熟悉与训练（10 分钟）：** 参与者查看了实验协议的信息图，并回答了两道理解性问题，以确认他们理解任务要求。

4. **系统交互（40 分钟）：** 参与者在每个系统上使用 20 分钟，每个系统使用前均先完成 3 分钟的教程。
   - **NarrativeLoom 条件（20 分钟）：** 参与者使用我们的多角色系统（GPT-4o API）。在每个叙事轮次中，系统通过匿名化AI人格生成多个续写选项。参与者可以选择、编辑或在这些选项的基础上继续创作，但不知道具体的人格来源。参与者必须使用自己的火花至少完成两个故事节拍。

   - **聊天机器人条件（20 分钟）：** 作为对照，参与者使用自定义聊天机器人界面与聊天机器人交互，该界面由相同的模型驱动，并使用相同的火花。故事通过与单个AI智能体进行直接对话式提示来创作。

5. **任务后评估（10分钟）：** 每种条件结束后，参与者都完成了一份关于其系统使用经验的问卷。

5.3 数据收集与分析

我们收集了定量和定性数据，以评估创意产品、创作过程和叙事质量。

5.3.1 创意产品与创作过程。我们采用经调整的指标来衡量创意产品和创作过程的质量，这些指标源自创意产品语义量表（CPSS） [46] 和创造力支持指数（CSI） [9, 14]。由于原始量表针对的是不同的创意领域——CPSS 面向一般创意产品，CSI 则面向更广泛的创造力支持工具——因此必须进行领域专门化调整，以用于叙事评估。

我们的调整包括三项关键修改：
（1）**领域专门化：** 我们将抽象的创造力维度转化为叙事专属构念。例如，CPSS 中通用的“解决能力”维度被具体化为叙事“连贯性”（即不存在情节不一致），而 CSI 中的“探索性”则被重新界定为叙事“多样性”（即故事可能性的范围）。

(2) **量表简化：** 我们将原有的71项CPSS和12项CSI精简为六个聚焦维度，并采用5点李克特量表，在保持比较评估构念效度的同时，减轻参与者疲劳。

(3) **情境相关性：** 我们重新设计了问题，使其符合AI辅助叙事情境。例如，CSI中的“协作”因子被调整为“定制化”，用于衡量系统能否使人工智能生成内容契合用户的创作愿景，而非衡量人际协作。

参与者对六个维度的经验进行了评分，这些维度分为两类：

- **产品指标：**
  – 新颖性 （改编自 CPSS 的“原创性”）：生成的叙事作品所具有的惊喜感和原创性程度

表 3：专家评估者人口统计信息和专业经验。四位专家评审员的资料，包括其职业、教育程度以及专业写作年限。

ID	性别	Age	职业	教育程度	写作经验（年）
E1	女性	32	编剧	电影制作艺术硕士	12
E2	Male	34	编剧兼导演	电影制作艺术硕士	8
E3	女性	38	创意制片人	中国文学文学学士	18
E4	Male	31	记者和作家	比较文学文学学士	10

表 4：托兰斯创造性写作测验（TTCW）评估框架。该系统评估包含 14 项二元标准，涵盖四个创意维度（流畅性、灵活性、原创性和阐述），用于评估叙事质量。

维度	测试问题
流畅性	时间处理是否恰当且平衡? 故事是否体现出对场景与概述之间平衡的把握? 故事是否巧妙运用了隐喻或文学手法? 结尾是否自然且令人信服，而非显得突兀? 故事元素是否相互配合，构成一个统一的整体?
灵活性	故事是否呈现了多样且令人信服的视角? 故事是否有效地平衡了内在性与外在性? 故事是否包含既出人意料又恰如其分的情节转折?
原创性	读者能从这个故事中获得独特的创意吗? 这个故事是否原创，且没有陈词滥调? 这个故事在结构或形式上是否展现了创新?
阐述	作家是否在感官层面让虚构世界显得可信? 角色的塑造是否达到了恰当的复杂程度? 故事是否具有多个意义层次?

– 多样性 （改编自 CSI 的“探索性”）：所提供的叙事可能性的范围和多样程度– 连贯性 （改编自 CPSS 的“解决能力”）：逻辑一致性，且不存在情节矛盾

- **过程指标：**
– 定制化 （改编自 CSI 中的“协作”）：系统对用户偏好和创作方向的响应程度– 参与度 （改编自 CSI 中的“享受”）：故事创作过程中的持续兴趣与投入– 可用性 （改编自 CSI 中的“成果值得付出”）：相对于创作成果质量而言，交互的易用程度

5.3.2 叙事质量分析。为补充主观评分，我们使用 spaCy [32] 和 NLTK [6]对生成故事进行了计算文本分析。我们分析了四项属性：

- **词数：** 衡量叙事长度和发展的基本指标。
- **Gunning Fog 指数：** 用于衡量文本可读性和复杂程度的指标。
- **对话比例：** 以角色对话与叙事阐述形式呈现的文本比例。
- **地点数量：** 不同空间指称的数量，反映故事背景的丰富程度。

5.3.3 角色设定选择追踪。对于NarrativeLoom条件，我们追踪了用户在故事创作过程中选择的AI人格。这使我们能够分析人格偏好、选择频率和选择序列的模式。

5.4 专家评审与反馈

为验证我们的研究发现，我们邀请了四位具有丰富创意写作经验的专家，对生成故事的叙事质量进行评估（人口统计信息见表3）。专家分为两组（E1-E2、E3-E4），每组独立评估同一组10对故事（共20个故事），这些故事对从用户研究数据集中随机抽取。因此，两组专家总共评估了20对、40个故事。为尽量减少偏差，所有故事均经过匿名处理，且展示顺序经过随机排列。

在评估中，我们使用了TTCW协议 [10], ，该协议通过14项二元测试（见表4）对四个创造力维度进行系统评估。每位专家独立对每项测试给出通过或未通过的二元评定，并附上简要理由。故事的最终创造力得分为通过测试的总数（0–14）。完成TTCW评估后，专家通过强制选择比较，选出他们认为每对故事中更优秀的一篇。

最后，我们对所有专家进行了半结构化访谈，以收集他们对各系统间感知差异的定性见解。访谈探讨了五个关键方面：(i) 各系统在创造性质量方面的显著差异，(ii) 质量差距最大的 TTCW 维度，(iii) 区分高质量与低质量故事的模式，(iv) 反复出现的创造性失误或局限，以及 (v) 改进 AI叙事系统的优先事项。我们对访谈进行了录音和主题分析，以识别专家视角中的共同模式。

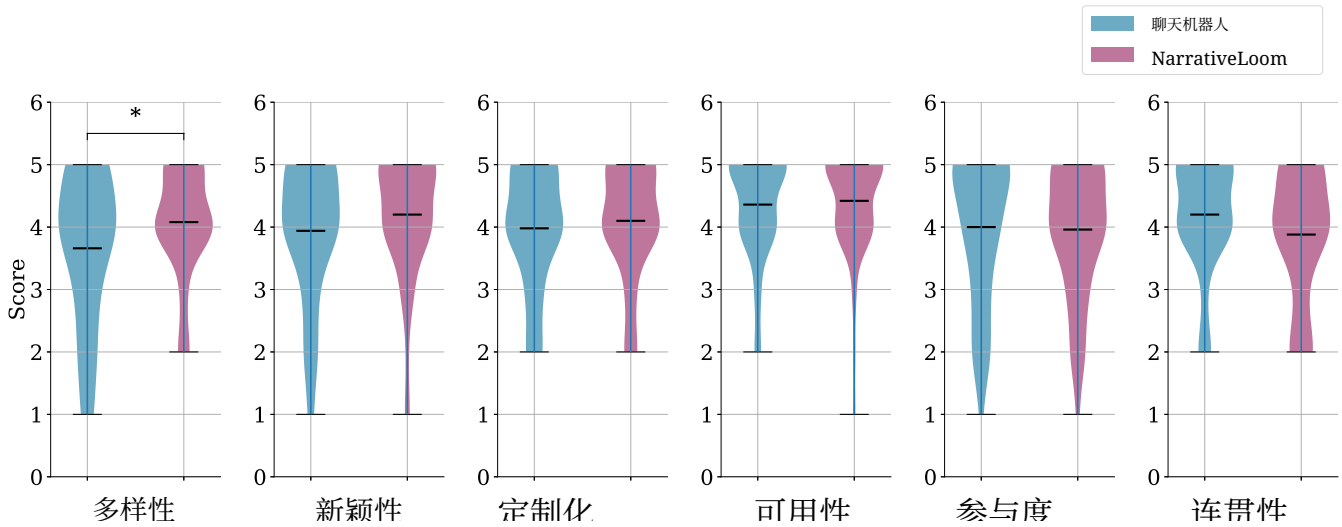


图 5：NarrativeLoom 与聊天机器人的用户评估得分对比。小提琴图展示了用户在关键维度上的评分分布，评分采用五点评分量表。NarrativeLoom 在多样性和新颖性等以创造力为重点的指标上，始终取得更高的中位数得分；在可用性、参与度和连贯性方面，表现则不相上下。星号表示统计显著性：* 表示 $p < 0.05$。

6 研究发现

6.1 多样化叙事图景的策略性探索

为了评估 NarrativeLoom 是否通过多样的叙事可能性成功增强了创造性探索性，我们分析了用户对创造力的感知，以及用户策略性地使用多角色系统的潜在模式。

用户感知的创意增强。 在用户感知的创意增强方面，NarrativeLoom 相较于基线展现出具有实际意义的优势（见图 5）。使用配对样本 t 检验进行分析，证实**多样性** 的提升（见第 5.3.1 节）具有统计显著性（$M = 4.08$， $SD = 0.89$ 对比。$M = 3.66$， $SD = 1.23$; $t(49) = 2.14$， $p = 0.037$，Cohen 的 $d = 0.39$）。就**新颖性** （见第 5.3.1 节）而言，尽管评分更高（$M = 4.20$, $SD = 0.98$ $M = 3.94$ $SD = 1.10$ $d = 0.25$ 对比。，；Cohen 的 ），但未达到统计显著性，效应量仍显示出有利于本系统的趋势。

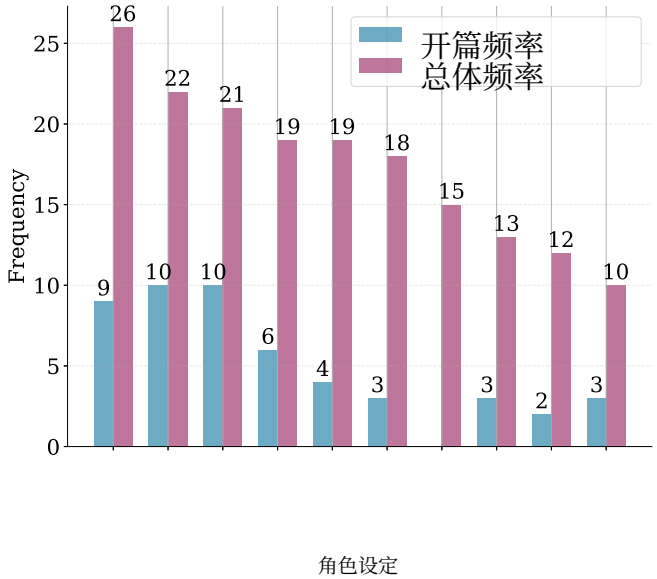
用户一致称赞 NarrativeLoom 的创意多样性，P23、P24 和 P34 指出，它提供了 “ 更多的情节走向选择 ”，生成的想法 “ 似乎更加多样、新颖。”用户尤其看重这种结构化的创意探索方式。P9 解释说，NarrativeLoom “能让不同的节拍层层加入故事，而不是像聊天机器人那样只创作故事的一个部分。”。

使用频率与叙事角色。 对角色设定使用模式的分析显示，用户会根据叙事阶段，有策略地选择特定角色设定来承担专门的角色（见图 6a）。历史类和反乌托邦类角色设定最常用于故事开篇，而悬疑类角色设定的总体使用频率最高。这种反差表明，用户会直觉地为角色分配不同功能：具有鲜明世界构建原型的角色设定（历史类和反乌托邦类）充当“发起者”，负责建立故事背景和初始冲突；而悬疑类角色设定则成为主要的“推进者”，在故事展开后用于处理情节复杂性并推动叙事发展。

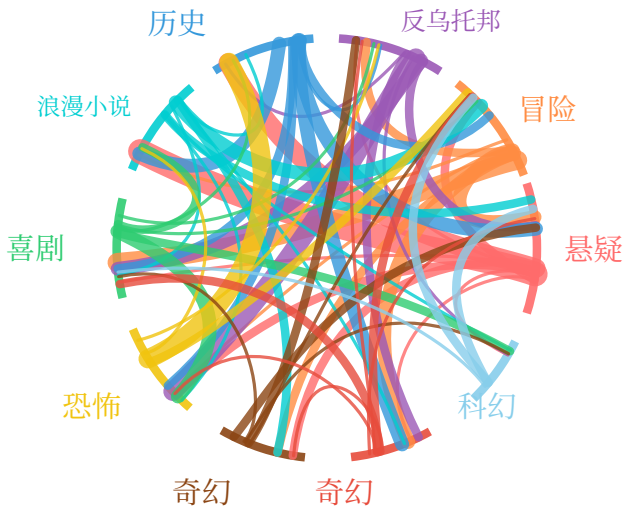
转场与叙事流。为了解故事如何发展，我们分析了角色之间的转移网络（见图 6b）。这些转移并非随机，而是形成了一个结构化网络，其特点是存在高频路径和明显的方向偏好。从悬疑到浪漫小说的转移最为常见，其他突出的路径还包括冒险到喜剧，以及反乌托邦到恐怖。这些路径往往体现了互补的类型组合，表明用户积极融合不同类型，以创作更丰富的叙事作品。此外，这种转移具有显著的不对称性，正向与反向转移之间的平均差异在统计上具有显著性（$M = 1.24 \pm 0.92$， $t(45) = 8.95$， $p < 0.001$）。例如，从悬疑到浪漫小说的转移（5 次）比反向转移（2 次）多一倍以上，而从历史到奇幻的路径（3 次）则从未出现反向转移。这种方向性表明，用户会引导故事沿着特定轨迹发展，并将多角色系统用作有目的、结构化地推动叙事升级的工具，而非仅仅追求简单的多样性。

6.2 平衡协作创作能动性与脚手架支持

从定量结果来看，NarrativeLoom 在三个维度上的表现与基线相当（见第 5.3.1 节）：**定制化** (NarrativeLoom: $M = 4.10, SD = 0.92$ 对比 聊天机器人: $M = 3.98, SD$ =0.97)、**可用性** (NarrativeLoom: $M = 4.42$, $SD = 0.83$ 对比 聊天机器人: $M = 4.36, SD = 0.84$)，以及**参与度** (NarrativeLoom: $M = 4.00, SD = 1.15$ 对比 聊天机器人: $M = 3.96$, $SD = 1.04$)，其中



(a) 角色使用分工。分组柱状图揭示了不同角色的专门分工。历史和反乌托邦等角色是主要的“发起者”，开篇使用率较高；而悬疑是关键的“推进者”，总体使用率占主导。



(b) 叙事流网络。转移图显示，角色之间的切换并非随机。边的粗细表示出现频率，揭示了常见且往往不对称的路径（例如，从悬疑到浪漫小说）。

图 6：用户有策略地选择角色设定，并以结构化模式在它们之间切换。数据揭示了两种关键的用户行为：(a) 为角色设定分配专门的“发起者”或“开发者”角色；(b) 沿着合乎逻辑且相邻的类型路径在角色设定之间切换。

差异无统计学意义。这表明，尽管复杂性有所增加，管理多个角色设定并未损害可用性。除了这些定量上的相似之处，我们的定性分析还揭示了 NarrativeLoom 所赋能的独特创作控制形式。我们确定了共同能动性增强的三个关键维度。

分布式创作权威（共享能动性）。NarrativeLoom 的多角色架构将创作主导权分配给系统和用户。它并非只提供单一的线性建议，而是由不同角色生成多样的叙事方向，拓展创作空间。用户仍可自行选择、组合或舍弃这些选项，从而保持明确的编辑控制权。参与者看重这种共享能动性的平衡；例如，P23 表示：“我喜欢智能体提供的不同选项”，而 P24 则强调，与通常会引导故事沿单一路径发展的传统聊天机器人相比，这种方式“在情节走向上提供了更多选择”。

响应式协作完善（协商式能动性）。用户将能动性体验为一种协商过程，他们的反馈会积极引导系统输出的方向。多位参与者强调，系统能够将他们的反馈融入后续生成内容。例如，P19指出，系统“在构思引人入胜的情境和角色的同时，也考虑了编辑修改和评论”。这种迭代式响应体现了用户输入与系统创造力如何共同塑造故事不断演进的走向。

赋能式创作伙伴关系（增强能动性）。通过增强用户的表达能力，NarrativeLoom让用户获得了更强的能动性感。参与者描述，系统“根据我的提示词生成了内容详尽且富有创造力的故事（*P30*）”，并创作出“与其他故事相比，读起来流畅自然得多，也更加充实（*P24*）”。P45则通过强烈的参与度，表达了这种增强的能动性：“我很喜欢系统为我生成的不同故事。它们引人入胜，真的很有意思。”

6.3 借助分段式结构应对创造力与连贯性权衡

NarrativeLoom 将叙事拆分为易于处理的创作单元，这种结构化方法在故事发展和丰富性方面带来了显著优势（见图 7）。与聊天机器人基线相比，NarrativeLoom 让用户创作的故事显著**更长**（M = 3803.36, SD = 1109.50 字；聊天机器人基线为，M = 1907.88, SD = 1304.44 字；t 49 = 9.160 p < .001 ( ) , ）。这种结构也有助于构建更丰富的叙事环境：NarrativeLoom 创作的故事中包含显著更多的**地点**（M = 3.86, SD = 2.67 对比 M = 2.44, SD = 2.54; t 49 p = ( ) = 3.276, ）, .002 且对话占比更高（, 对比 M = 0.16, M = 0.30 SD = 0.13; SD = 0.12 t(49) = 5.675, p < .001）。此外，最终生成的叙事作品更易于阅读，可读性 得分更低（更佳）（M = 7.45, SD = 0.54 对比 M = 8.17, SD = 1.16; t(49) = −4.212, p < .001）。

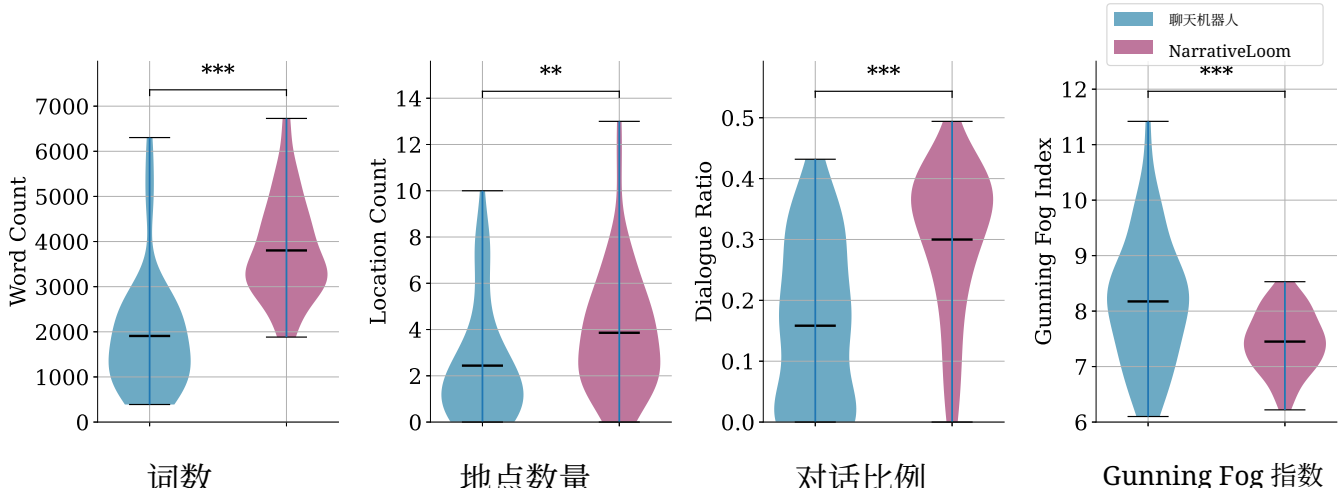


图 7：NarrativeLoom 与聊天机器人的叙事特征比较。小提琴图对比了聊天机器人基线（蓝色）与 NarrativeLoom（粉色）。NarrativeLoom 生成的故事显著更长、包含更多地点、对话占比更高，且可读性更强（Gunning Fog 指数更低）。星号表示统计显著性： 表示 $p < 0.01$，*** 表示 $p < 0.001$。**

定性反馈证实了这一分段式流程的价值。正如 P10 所说："我喜欢 NarrativeLoom 分段进行叙事创作过程的方式。聊天机器人给人的感觉更像是非此即彼，而这个系统则一次专注于一个时刻。"然而，这种促进创造力的多角色设定、分段式架构，也带来了维持叙事连贯性的挑战。

事实上，与 NarrativeLoom（$M = 3.88, SD = 1.05$）相比，聊天机器人基线的连贯性 评分更高（见图 5；$M = 4.20, SD = 0.92$）。尽管这一差异在统计上并不显著，但其效应量处于小到中等水平（Cohen's $d = 0.32$）。这是一种预期中的权衡：聊天机器人生成的是连贯的单一声音叙事，而我们系统的多角色设计以及显著更长的故事增加了潜在的片段断裂点，使连贯性更容易受到挑战。

尽管有这些结果，用户仍表示 NarrativeLoom 保持了逻辑一致性。P28 指出："总体而言，建议和生成的文本都不错，故事生成过程也很流畅。"P38 则观察到，系统"生成了引人入胜的片段，并将其串联起来，形成了流畅的故事"。这些定性反馈表明，系统能够妥善处理连贯性与创造性的权衡，在提升叙事丰富度的同时，维持了可接受的故事连贯性。

6.4 写作专长塑造对 AI写作辅助的偏好

用户特征，尤其是写作经验，影响了他们与叙事系统的互动，这表明有必要提供个性化的 AI写作辅助。我们观察到，初学者和资深写作者对这两个系统的评价模式有所不同（见图 8），不过交互效应在统计上并不显著。

在新颖性方面，初学者对 NarrativeLoom 的评分高于聊天机器人基线（NarrativeLoom：$M = 4.26$，$SD = 0.90$；聊天机器人：$M = 3.78$，$SD = 1.18$；Cohen's $d = 0.457$）。相反，资深者没有明显偏好，对两个系统的评分相近（聊天机器人：$M = 4.00$，$SD = 1.00$；NarrativeLoom：$M = 3.94$ $SD = 1.18$，）。

可用性评分呈现出不同的模式。初学者对 NarrativeLoom 略有偏好（$M =$ NarrativeLoom：4.39 $SD = 0.97$ $M = 4.26$ $SD = 0.90$ $d = 0.140$，；聊天机器人：，；Cohen's），而资深写作者认为聊天机器人的可用性更高（聊天机器人：$M = 4.67$，$SD = 0.47$；NarrativeLoom：$M = 4.39$，$SD = 0.76$；Cohen's $d = 0.441$）。

用户偏好也受到其具体写作阶段的影响。P8总结了这一差异："聊天机器人适合已经大致知道自己想写什么故事、只是需要帮助将其实现的人。但对于那些想写作却不确定该讲述什么故事的人来说，*NarrativeLoom*系统更为合适。"这表明，人工智能辅助应与创作过程的不同阶段相匹配。NarrativeLoom中的结构化BVSR框架通过帮助写作者识别叙事可能性，支持构思和探索阶段。相比之下，传统聊天机器人可能在发展阶段更为有效，此时写作者专注于执行和完善已经确立的概念。

这些结果表明，采用普适性的AI写作辅助方法可能并非最佳选择。观察到的偏好模式——初学者优先看重NarrativeLoom的新颖性，而资深作家优先看重聊天机器人的可用性——表明，支持系统应根据用户专业水平及其当前所处的写作过程阶段进行定制。

6.5 专家评估结果

为补充大规模用户研究，四位创意写作专家评估了叙事质量。我们从数据集中随机抽取了 20 对故事，每对包含一个由 NarrativeLoom 生成的故事和一个由聊天机器人基线生成的故事（$N = 40$）。专家们几乎一致地偏好 NarrativeLoom。专家指出，两个系统都无法始终生成完美的作品，但他们认为 NarrativeLoom 生成的叙事作品质量更高，也更引人入胜。

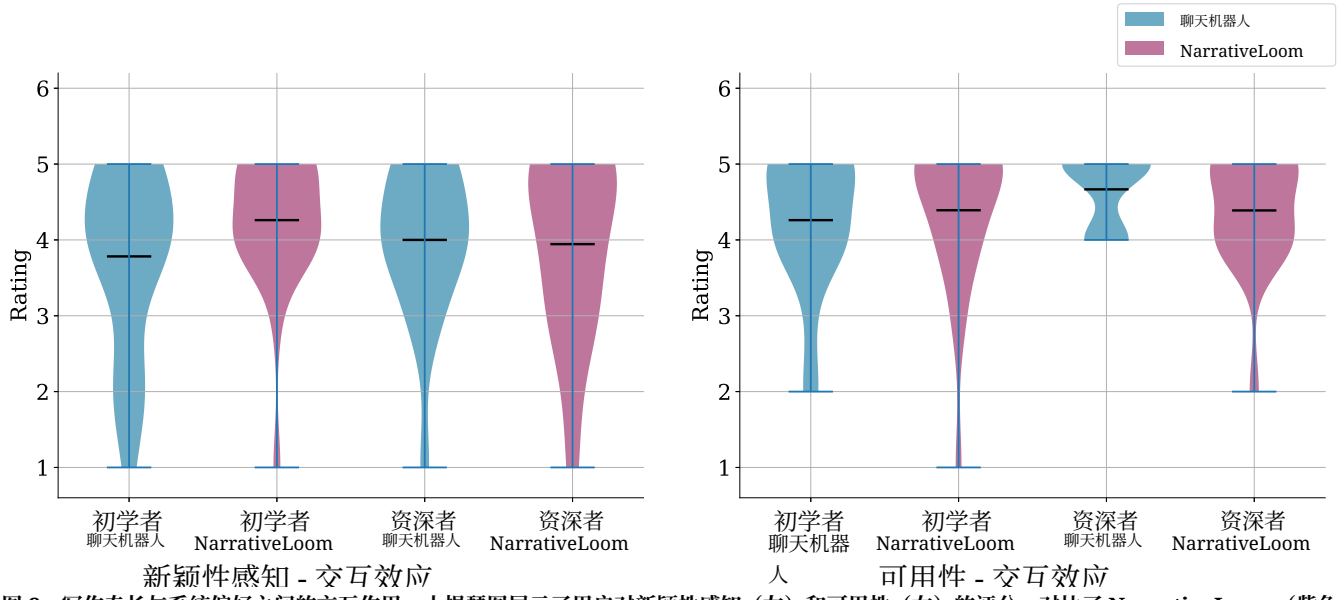


图 8：写作专长与系统偏好之间的交互作用。小提琴图展示了用户对新颖性感知（左）和可用性（右）的评分，对比了 NarrativeLoom（紫色）与聊天机器人基线（蓝色）。新手写作者认为 NarrativeLoom 更新颖，而资深写作者则认为聊天机器人更易用。

定量偏好与创造力得分。在直接比较中，专家更偏好 NarrativeLoom，在 40 个强制选择情境中，有 38 个选择了 NarrativeLoom 生成的故事。这一偏好也反映在 TTCW 得分中（图 9）。配对 t 检验显示，NarrativeLoom 的总体创造力得分更高（$M = 9.72$ 满分 14 分，$SD = 2.1$），而聊天机器人基线为（$M = 5.00$ 满分 14 分，$SD = 1.8$；$t(79) = 13.68$，$p < 0.001$）。对各项维度的分析显示，NarrativeLoom 在流畅性（$M =$ 4.38 5 2.40 t 79 $= 12.93$ $p < 0.001$ 分对比；( )，）、灵活性（$M =$ 2.33 3 1.15 t 79 $= 12.61$ $p < 0.001$ 分对比；( )，）、阐述（$M =$ 1.95 3 1.07 t 79 $= 8.31$ $p < 0.001$ 分对比；( )，）和原创性（$M = 1.07$ 分，满分 3 分对比0。38；$t(79) = 6.16$，$p = 0.001$）方面均占优势。此外，NarrativeLoom 在所有用户专业水平层级中都优于聊天机器人基线；这一优势的幅度在初学者与资深写作者之间没有显著差异（图 10）。尽管专家样本量较小，限制了广泛的统计推断，但这些结果与我们的大规模用户研究一致。

创意独创性与不可预测性。专家的定性反馈指出，NarrativeLoom 在三个主要方面创作出了更出色的叙事作品，首要之处在于创意创新。基线版本常因墨守成规而受到批评；E1 指出，它“经常依赖传统故事结构和陈词滥调元素。”相比之下，NarrativeLoom 因其叙事惊喜而备受赞誉。正如 E4 所说，“这些故事会带你去一个意想不到的地方……聊天机器人创作的故事多少有些可预测。”E3 也表达了类似看法，他指出，“*NarrativeLoom* 生成的这个故事有点出乎意料。我读到 *[Sparkle]* 时并没有料到这一点。”专家还提到了具体的新颖想法，E1 特别指出，“*NarrativeLoom* 创作的故事常常有一些巧妙的想法或亮点，是我以前没见过的，而且不那么陈腐。例如以“友谊”为主题的故事为例，朋友们相遇后，一位时间旅行者出现了。”

叙事技艺更胜一筹：“展示”胜于“讲述”。第二个主题涉及 NarrativeLoom 对沉浸式场景叙事的运用。E1 对比两个系统时这样描述聊天机器人：“它有点像一本说明手册，直接从‘上帝视角’把事情一一摆出来，只是简单地罗列事件。”相比之下，E1 指出：“*NarrativeLoom* 不会一次性呈现所有信息，而是从一个场景开始，让人感觉故事正以一种缓慢而引人入胜的方式娓娓道来。”这与 E2 的评价一致：NarrativeLoom 的写作“深入人物内心，细节丰富，场景铺陈充分”，而聊天机器人则“更偏第三人称、客观，像是在做总结”。E4 最直接地阐明了这一区别：“*NarrativeLoom* 的写作水平高出许多。它更擅长展示，而不是讲述。”

角色深度和心理现实主义。NarrativeLoom 能够塑造出具有心理深度的角色，而他们认为基准系统缺乏这一特质。E1 观察到“*NarrativeLoom* 故事中的大多数角色都设定得很好，”E4 则总结了这种偏好，直接指出 NarrativeLoom“角色和对话都更出色。”E2 进一步阐述了这一区别，指出聊天机器人角色“缺乏内在深度和心理复杂性，没有内心的犹豫或心理活动。”相较之下，NarrativeLoom 塑造的人物更加细腻；E2 特别称赞 NarrativeLoom 的“角色发展很出色，”尤其是反派，他们给人的感觉是“真实的人，而不是人为拼凑出来的角色。”这种营造可信内心世界的能力，是专家评审中反复提及的一项显著差异。

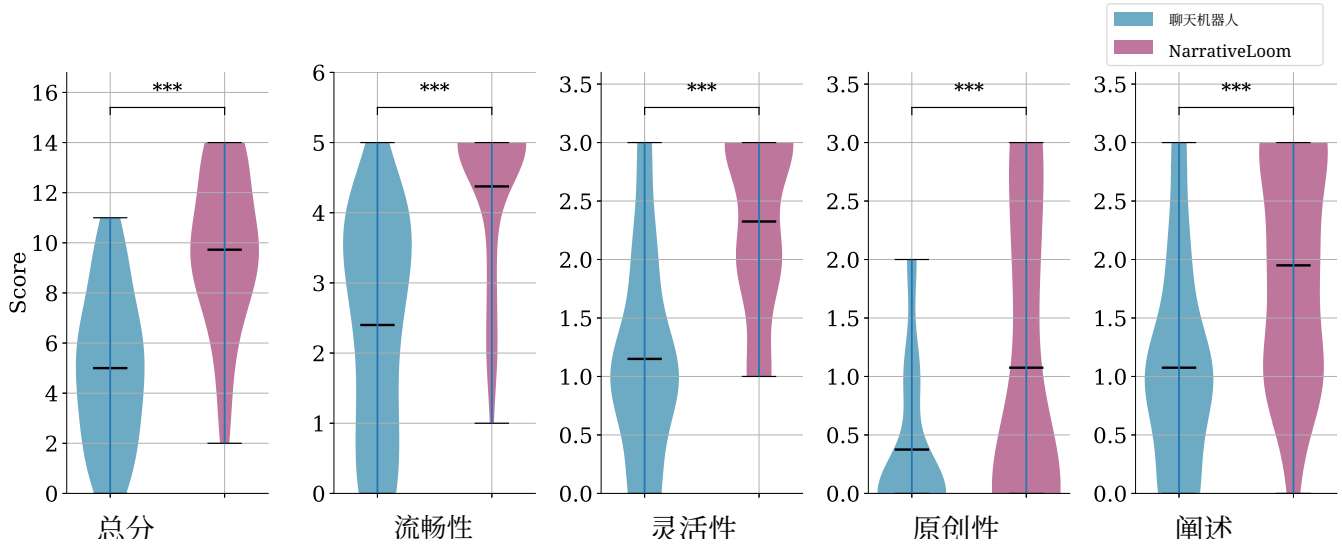


图 9：使用TTCW框架的专家评估结果。小提琴图展示了四个维度及总体表现的创造力得分分布，对比了NarrativeLoom（紫色）与聊天机器人基线（蓝色）。NarrativeLoom在所有创造力维度上均取得了统计显著的提升。星号表示统计显著性：*表示 $p < 0.001$。**

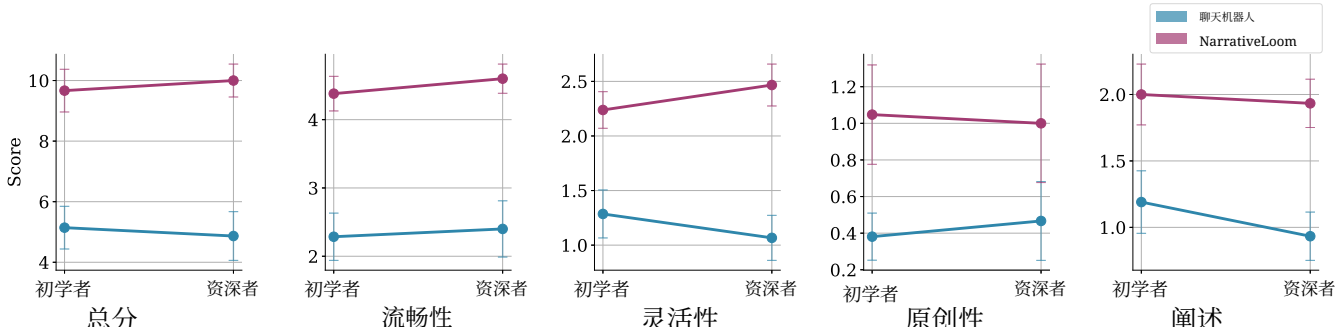


图 10：基于写作专长对参与者叙事进行的专家TTCW评估。图中显示了初学者和资深写作者使用NarrativeLoom（紫色）和聊天机器人（蓝色）时的平均得分（$\pm SE$）。评估标准包括总分、流畅性、灵活性、原创性和阐述。对于这两组用户，NarrativeLoom在所有指标上的表现均优于聊天机器人。

7 讨论

我们的研究表明，与聊天机器人系统相比，NarrativeLoom 显著提升了创意叙事表现（见第 6.5 节）；用户会策略性地利用不同角色来承担特定的叙事任务（见第 6.1 节），创作出篇幅更长、细节更丰富的故事（见第 6.3 节），并在系统提供结构化引导的同时保持创作主体性（见第 6.2 节）。这些发现揭示了坎贝尔的盲变与选择保留理论 [8] 如何为人工智能系统的设计提供一套有原则的框架，既拓展创作可能性，又保留用户能动性。

7.1 平衡创意多样性与认知可管理性

我们的研究发现，人工智能辅助创意写作存在一个根本性的矛盾：如何在拓展叙事可能性的同时，保持认知可管理性。当前系统通过介入不同的叙事层次来应对这一问题。通过参数化采样在**话语层** 运作的系统（Wordcraft [68]）能够实现风格变化，但无法改变底层故事逻辑。**情感轨迹层**（TaleBrush [15]）提供可视化的故事弧线控制，但会限制叙事具体性。**结构规划层** 提供两种方式：一种是僵化的自上而下的级联（Dramatron [43]），会限制自发性；另一种是可配置抽象（WhatElse [41]），则需要复杂的预先规划。

我们基于类型的角色人设方法介入的是**故事内层**——即故事事件与因果逻辑这一层面。这一区分很重要，因为类型会塑造叙事可能空间 [13]。通过实例化多个类型角色，我们解决了统计质心问题 [5],：单一模型会从训练分布的中心进行采样，因而收敛到可预测的输出。每个角色都从不同的叙事区域进行采样：当参与者运用喜剧来缓解紧张，或运用悬疑来提升情节复杂度时，他们接触到的是截然不同的叙事逻辑。这也解释了为何专家认为，与基线故事相比，NarrativeLoom 的故事更有可能带领读者前往意想不到的地方，而基线故事仍然

可预测——类型干预改变的是可能发生什么，而不只是如何讲述。

基于节拍的设计为创作脚手架提供了特定的粒度。Dramatron 的 [43] 严格规划确保了连贯性，但需要大量认知上的预先规划。Wordcraft 的 [68] 开放式生成具有灵活性，却可能让用户应接不暇。我们的方法采取由中而外的立场——通过语义锚点实现结构化即兴创作，而不预设结果。这一方法处于最近发展区 [62]之内：它足够复杂，能够支持精妙的叙事作品，同时又足够模块化，便于管理。NarrativeLoom 中显著更长的故事、更丰富的对话和更多样的场景设定验证了这种平衡。连贯性方面的轻微取舍，是提升创作探索性的可接受代价。

7.2 时间综合与协作涌现

BVSR 中创造力的源头仍是一个持续争论的话题。经典理论将变异视为创造力的主要驱动力——坎贝尔 [8] 和 Simonton [55, 56] 认为，创造力源于大量盲目变异——而近期计算研究则表明，创作主体性存在于人类选择之中。在这一观点下，内部预测模型引导策略性选择，而非随机过滤 [20, 72]。

我们的结果表明，NarrativeLoom 中的创造力源于算法变异、人类选择与情境综合之间的三元互动。角色转换模式（见图 6b）显示，用户并非仅仅选择选项；他们通过连续选择构建叙事轨迹，其中每一次选择都会约束后续的可能性。非对称转换表明，用户利用叙事可供性 [61] ，这些可供性产生于多个角色的交汇。这一观察与 Zhou 和 Lee [70], 的看法一致，他们在文生图情境中指出，人类创造力通过策展得以显现。这进一步支持 Epstein 等人 [22]的论点，即生成式人工智能将创造力从直接操控转变为迭代式规范，从而使用户能够保持对叙事轨迹的控制。

然而，我们的研究发现了一种与人机交互设计相关的区别：与图像生成 [70], （即从并行备选方案中进行选择）不同，叙事创作需要时间综合。用户必须将所选节拍整合为连贯的序列结构。这一时间维度促进了协作涌现 [52], ，即创意产品源于人类与人工智能之间的结构化即兴。NarrativeLoom 故事中更高的对话比例和提升的地点多样性表明，这种综合将创造性阐述放大到超越个体能力的水平，从而支持分布式认知 [34]理论。这些结果表明，共创系统应作为进行时间综合的平台，而非简单的生成工具，并同时利用人类策展与人工智能变异。

7.3 写作者对创造性支持的需求程度各不相同

我们的研究结果表明，写作者会根据其**专业水平**和**创作阶段**（见第6.4节）需要不同类型的创造性支持，这对当前人工智能写作工具所采用的“一刀切”方法提出了挑战。这与先前关于写作者希望AI协作者扮演多种角色的研究一致 [11, 27], ，同时也揭示了创造力支持研究中经常被忽视的一个关键设计考量 [14]。

与创造性领域中的专家与新手差异相一致 [23, 24], ，我们观察到了不同的偏好模式。新手写作者偏好 NarrativeLoom 的结构化探索，受益于能够激活不同叙事空间的多角色脚手架 [57] ，并在自身资源有限时帮助他们克服创作瓶颈 [65, 69]。相反，专家写作者偏好精简的类聊天机器人界面，能够无缝融入他们既有的工作流程。这与专长逆转效应相符 [36]——对新手有益的脚手架，对拥有高度成熟图式且力求保持创作心流的专家而言，却会成为多余的认知负荷 [17]。

关键的是，我们的TTCW评估表明，这些偏好差异并未转化为质量差异——两组使用 NarrativeLoom 都取得了相当的创作提升。这一悖论揭示了一个关键区别：创造性支持作用于两个层面——**工作流程整合** （专家倾向于尽量减少干扰）和**认知刺激** （结构化多样性对所有人都有益）。多角色系统的价值在于通过系统性变化来对抗认知固着，这一益处不受专业水平影响。

除了静态的专业水平之外，不同创作阶段的支持需求也会动态变化。写作者在构思阶段可能会受益于 NarrativeLoom 的探索性脚手架，但在修订阶段则更偏好精简的辅助 [12, 25, 35]。这表明，未来的创意人工智能应采用**自适应双通道设计**：在保留结构化变异带来的认知益处的同时，允许用户自定义界面，从提供充分脚手架（初学者／构思阶段）到尽量减少干扰（专家／完善阶段）。关键原则是将创作脚手架与界面复杂度解耦——保留前者的价值，同时尽量减少后者造成的干扰。

7.4 创意人工智能系统的设计启示

7.4.1 理论指导设计：对模型规模的另一种思路。我们的结果表明，创意人工智能系统若要提升性能，受益于审慎设计的理论指导框架，而非仅仅依赖扩大参数规模 [7, 48]。我们构建的结构化多角色模型以坎贝尔的变异与选择性保留（BVSR）框架 [8],为基础，其成功表明，有原则的设计可能比单纯提升计算能力更有效。这与一种观点一致：如果缺乏根本性的结构创新，当前的人工智能方法可能会面临收益递减 [42]。我们的方案并未追求更大的模型，而是表明，通过融入多样化的创作视角，系统能够探索训练语料中往往较少涉及的叙事空间，例如需要新颖类型组合或非传统情节发展的空间。未来的创意人工智能可采用模块化系统，设置针对特定创作功能进行优化的专用组件，从而提供更有针对性的辅助。

7.4.2 促进创造性涌现的结构化即兴架构。基于节拍的分段方法应对了长篇生成中的一项核心挑战：在不抑制

涌现的同时保持连贯性。与可能限制探索的分层规划系统 [43] 或有损连贯性的无约束生成 [31], 不同，我们的模型实现了渐进式连贯性——在可管理单元内保持局部一致性，同时兼顾涌现式全局发展。这些发现表明，创意人工智能应将复杂任务分解为离散的上下文感知单元，在保留自发性的同时，为扩展创作提供导航结构 [53]。未来的系统可利用自适应机制，根据项目类型、用户偏好和上下文自动确定最优边界，从短篇作品扩展到长篇作品，同时兼顾新颖性与逻辑性。

7.4.3 创意主体性的明确角色分离。 NarrativeLoom 中人工智能辅助与人类主导权之间的平衡，凸显了明确划分角色的必要性。通过让人工智能负责生成变体，并将选择权保留给人类，我们的方法在利用计算速度的同时，维护了用户能动性。因此，创意人工智能系统应在机器贡献与人类贡献之间设置透明的界限，避免采用会削弱创作者信心的黑箱模式 [18, 37, 71]。未来的设计应加入明确的角色指示器，为生成和选择设置独立界面，并允许用户协商这些界限，自行决定将多少创作空间交由系统处理。

7.5 局限性与未来工作

尽管我们的结果展示了基于盲变异与选择理论的计算创造力的有效性，但仍有若干理论和方法上的局限值得考虑。依赖基于体裁的角色设定专门化虽然有助于叙事多样性，但可能倾向于沿用既有文学惯例，而未能充分探索非传统的创作可能性 [13]。我们的评估框架虽然全面，却不可避免地反映了西方叙事传统和评估标准。鉴于有证据表明，不同文化语境中的创造力表现形式和评估方式存在显著差异，跨文化验证对于确立其更广泛的适用性至关重要 [45]。此外，观察到的专业知识效应也引发了关于人工智能辅助创造力长期发展影响的问题。初学者虽受益于系统支持，但其长期发展影响仍不明确。纵向研究有助于确定此类脚手架究竟能促进技能习得，还是会无意中造成依赖，从而可能阻碍独立创造能力的发展。这一担忧与教育情境中关于技术支架的既有讨论相呼应 [47],，即过度依赖外部支持可能削弱掌握所需的内在认知努力。研究这一发展轨迹，对于设计能够发挥发展工具作用、而非沦为认知拐杖的系统至关重要。

未来的工作将优先推进四项系统改进。首先，我们将**突破基于体裁的专业化方式，拓展面向功能的模块**，开发专注于世界构建、角色心理和叙事节奏的角色设定。这种混合系统旨在支持非常规的创意组合，同时保留本研究中观察到的体裁多样性优势。其次，我们计划**实现自适应加权机制**，从用户选择模式中学习并调整角色设定的突出程度。当系统察觉创造性发散减弱时，该方法会主动推荐互补的替代选项，以帮助缓解创造性固着。第三，我们将**纳入具有文化意识的变体**，使其以西方叙事之外的文学传统为基础进行训练。这需要与文化专家合作，确保在全球不同文化背景下真实呈现故事结构和惯例 [45]。最后，我们将**开发专业水平自适应界面**，根据用户熟练度动态调整支架支持。系统将追踪选择速度、编辑频率和连贯性评分，为初学者提供探索支持，并为专家提供精简工具 [51],，以适应不同的用户能力和任务需求。

8 结论

在 NarrativeLoom 中成功应用 BVSR 理论，表明心理创造力框架能够为计算创造力系统提供有效蓝图。通过明确区分变体生成与选择性保留，并借助专业化角色实现结构化多样性，我们构建了一种创造性伙伴关系模型，在增强人类创造力的同时保留创作主体性。研究中观察到的、因专业知识而异的偏好凸显了创意人工智能中自适应设计的重要性，而策略性的角色使用模式则表明，只要提供适当的交互框架，人类便能与人工智能集群有效协作。

这些结果表明，创意人工智能的下一个前沿不仅在于更智能的模型，更在于尊重个体创造性认知细微差异的、有原则的设计。随着人工智能能力不断发展，以理论为指导的方法对于确保人工智能继续成为创作过程中的伙伴至关重要——这将有助于构建真正拓展人类想象力、同时维护人类表达与自主权内在驱动力的系统。

致谢

我们谨向 Zhen Chen 致谢，感谢他创作了精美的插图，为本文增色不少。我们也感谢 Yujia Peng、Guangyuan Jiang 和 Yizhou Wang 在本研究开展过程中提供的宝贵反馈与真知灼见。本研究部分得到了国家科技创新2030重大项目（2025ZD0219400）、国家自然科学基金委员会（62376009、62376031）、北京大学通用人工智能国家重点实验室、北大-冰极人工智能联合实验室、武汉市重大科技专项（2025060902020304）、湖北省具身智能基础模型研发计划，以及武汉东湖高新技术开发区国家智能社会治理综合实验基地的资助。

参考文献

[1] Josh Achiam, Steven Adler,Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Floren- cia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat 等。2023。GPT-4 技术报告。*arXiv* 预印本 *arXiv:2303.08774* (2023)。[2] Arwa I Alhussain 和 Aqil M Azmi。2021。自动故事生成：方法综述。*Comput. Surveys* 54, 5 (2021), 1–38。

[3] Barrett R Anderson、Jash Hemant Shah 和 Max Kreminski。2024。大语言模型对人类创造性构思的同质化效应。载于创造力与认知会议。[4] 亚里士多德。1942。亚里士多德的《诗学》。北卡罗来纳大学出版社。[5] Emily M Bender、Timnit Gebru、Angelina McMillan-Major 和 Shmargaret Shmitchell。2021。论随机鹦鹉的危险：语言模型会不会过于庞大？载于ACM公平、问责与透明度会议。[6] Steven Bird、Ewan Klein 和 Edward Loper。2009。使用 Python 进行自然语言处理：借助自然语言工具包分析文本。" O'Reilly Media, Inc."。[7] Tom Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared D Kaplan、Prafulla Dhariwal、Arvind Neelakantan、Pranav Shyam、Girish Sastry、Amanda Askell 等。2020。语言模型是少样本学习者。载于神经信息处理系统进展（NeurIPS）论文集。[8] Donald T Campbell。1960。创造性思维中的盲目变异与选择性保留，及其他知识过程中的盲目变异与选择性保留。《心理学评论》67, 6 (1960), 380。[9] Erin A Carroll 和 Celine Latulipe。2009。创造力支持指数。载于ACM计算机系统中的人为因素会议（CHI）。[10] Tuhin Chakrabarty、Philippe Laban、Divyansh Agarwal、Smaranda Muresan 和 Chien-Sheng Wu。2024。艺术还是人工伪造？大语言模型与创造力的虚假承诺。载于ACM计算机系统中的人为因素会议（CHI）。[11] Tuhin Chakrabarty、Philippe Laban 和 Chien-Sheng Wu。2025。人工智能写作能否挽救？通过编辑修改缓解特异性问题，并改善写作过程中的人机对齐。载于ACM计算机系统中的人为因素会议（CHI）。1–33。[12] Tuhin Chakrabarty、Vishakh Padmakumar、Faeze Brahman 和 Smaranda Mure- san。2024。大语言模型时代的创造力支持：一项涉及专业作家的实证研究。载于创造力与认知会议。[13] Daniel Chandler。1997。体裁理论导论。[14] Erin Cherry 和 Celine Latulipe。2014。通过创造力支持指数量化数字工具的创造力支持。ACM人机交互汇刊21, 4 (2014), 1–25。[15] John Joon Young Chung、Wooseok Kim、Kang Min Yoo、Hwaran Lee、Eytan Adar 和 Minsuk Chang。2022。TaleBrush：使用生成式预训练语言模型勾勒故事。载于ACM计算机系统中的人为因素会议（CHI）。[16] Elizabeth Clark、Anne Spencer Ross、Chenhao Tan、Yangfeng Ji 和 Noah A Smith。2018。机器参与其中的创意写作：关于口号和故事的案例研究。载于智能用户界面国际会议。[17] Mihaly Csikszentmihalyi 等。1997。心流与发现和发明心理学。HarperPerennial，纽约39 (1997), 1–16。[18] Nicholas Davis、Chih-Pin Hsiao、Yanna Popova 和 Brian Magerko。2015。面向计算协作与共同创作的生成式创造力模型。《数字时代的创造力》(2015), 109–133。[19] Paramveer S Dhillon、Somayeh Molaei、Jiaqi Li、Maximilian Golub、Shaochun Zheng 和 Lionel Peter Robert。2024。塑造人机协作：语言模型协同写作中的不同支架层级。载于ACM计算机系统中的人为因素会议（CHI）。[20] Arne Dietrich 和 Hilde Haider。2015。人类创造力、进化算法与预测性表征：思维试验机制。《心理学通报与评论》22, 4 (2015), 897–915。[21] Steven Earnshaw。2014。《创意写作手册》。爱丁堡大学出版社。[22] Ziv Epstein、Aaron Hertzmann、人类创造力研究者、Memo Akten、Hany Farid、Jessica Fjeld、Morgan R Frank、Matthew Groh、Laura Herman、Neil Leach 等。2023。生成式人工智能的艺术与科学。《科学》380, 6650 (2023), 1110–1111。[23] K Anders Ericsson。2018。经验、练习与刻意练习对专家卓越个人表现发展的不同影响。载于《剑桥专长与专家表现手册》，K Anders Ericsson、Robert R Hoffman、Aaron Kozbelt 和 A Mark Williams（编）。剑桥大学出版社，745–769。[24] K Anders Ericsson 和 Jacqui Smith。1991。迈向一般专长理论：前景与局限。剑桥大学出版社。[25] Ronald A Finke、Thomas B Ward 和 Steven M Smith。1996。创造性认知：理论、研究与应用。麻省理工学院出版社。[26] Giorgio Franceschelli 和 Mirco Musolesi。2024。论大语言模型的创造力。《人工智能与社会》(2024), 1–11。[27] Katy Ilonka Gero、Tao Long 和 Lydia B Chilton。2023。创意写作中人工智能支持的社会动态。载于ACM计算机系统中的人为因素会议（CHI）。[28] Pablo Gervás。2009。叙事与创造力的计算方法。《人工智能杂志》30, 3 (2009), 49–49。[29] Hans Hansen、Daved Barry、David M Boje 和 Mary Jo Hatch。2007。真相还是后果：即兴集体故事创作。《管理学探究杂志》16, 2 (2007), 112–126。[30] Michael Hauge。2011。畅销电影剧本写作。布鲁姆斯伯里出版社。[31] Ari Holtzman、Jan Buys、Li Du、Maxwell Forbes 和 Yejin Choi。2020。《神经文本退化的奇特案例》。载于《国际学习表征会议》（ICLR）论文集。[32] Matthew Honnibal、Ines Montani、Sofie Van Landeghem 和 Adriane Boyd。2020。《spaCy：面向工业级应用的 Python 自然语言处理》。https: //github.com/explosion/spaCy。[33] David Howard 和 Edward Mabley。1993。《剧本创作工具：写作者的剧本创作技艺与要素指南》。Macmillan。[34] Edwin Hutchins。2000。分布式认知。《国际社会与行为科学百科全书》138, 1 (2000), 1–10。[35] Daphne Ippolito、Ann Yuan、Andy Coenen 和 Sehmon Burnam。2022。借助人工智能写作助手进行创意写作：专业作家的视角。arXiv 预印本 arXiv:2211.05030 (2022)。[36] Slava Kalyuga。2009。专长逆转效应。载于《自适应多媒体学习中的认知负荷管理》。IGI Global Scientific Publishing, 58–80。[37] Anna Kantosalo 和 Hannu Toivonen。2016。创造性人机协作的模式：交替式与任务分工式协同创造力。载于《国际计算创造力会议》。[38] Joy Kim、Justin Cheng 和 Michael S Bernstein。2014。Ensemble：探索领导者与群体在创造性协作中的互补优势。载于《计算机支持的协同工作与社交计算会议》。[39] Randee Lipson Lawrence 和 Dennis Swiftdeer Paige。2016。祖先的智慧：通过叙事进行教学与学习。《成人与继续教育新方向》149, Spring (2016), 63–72。[40] 2022 10 5281 1234 Jerry Liu。[41] 2025 LlamaIndex。https://doi.org/ 。/zenodo。Zhuoran Lu、Qian Zhou 和 Yi Wang。WhatELSE：以可配置的抽象层级塑造叙事空间，实现人工智能驱动的交互式叙事。载于《ACM计算机系统中的人为因素会议》（CHI）。[42] Gary Marcus。2020。人工智能的下一个十年：迈向稳健人工智能的四个步骤。arXiv 预印本 arXiv:2002.06177 (2020)。[43] Piotr Mirowski、Kory W Mathewson、Jaylen Pittman 和 Richard Evans。2023。与语言模型协同创作电影剧本和戏剧剧本：行业专业人士的评估。载于《ACM计算机系统中的人为因素会议》（CHI）。[44] Michael Z Newman。2006。从节拍到弧线：迈向电视叙事诗学。《The Velvet Light Trap》58, 1 (2006), 16–28。[45] Weihua Niu 和 Robert J Sternberg。2001。文化对艺术创造力及其评价的影响。《国际心理学杂志》36, 4 (2001), 225–241。[46] Karen O'Quin 和 Susan P Besemer。1989。修订版创意产品语义量表的编制、信度与效度。《创造力研究杂志》2, 4 (1989), 267–278。[47] Roy D. Pea。2004。脚手架及相关理论概念的社会与技术维度：学习、教育与人类活动。《学习科学杂志》13, 3 (2004), 423–451。[48] Yongqian Peng、Yuxi Ma、Mengmeng Wang、Yuxuan Wang、Yizhou Wang、Chi Zhang、Yixin Zhu 和 Zilong Zheng。2025。探究并诱发视觉-语言模型中的组合创造力。载于认知科学学会年会（CogSci）。[49] William D Perreault。1975。控制顺序效应偏差。《公共舆论季刊》39, 4 (1975), 544–551。[50] Mark O Riedl 和 R Michael Young。2006。作为探索性创造力的故事规划：拓展叙事搜索空间的技术。《新一代计算》24 (2006), 303–323。[51] Gavriel Salomon、David N Perkins 和 Tamar Globerson。1991。认知伙伴：借助智能技术拓展人类智能。《教育研究者》20, 3 (1991), 2–9。[52] R Keith Sawyer。2000。即兴与创作过程：杜威、科林伍德与自发性的美学。《美学与艺术批评杂志》58, 2 (2000), 149–161。[53] R Keith Sawyer。2014。群体创造力：音乐、戏剧与协作。Psychology Press。[54] R Keith Sawyer 和 Stacy DeZutter。2009。分布式创造力：集体创作如何在协作中涌现。《美学、创造力与艺术心理学》3, 2 (2009), 81。[55] Dean Keith Simonton。1999。创造力即盲目变异与选择性保留：创造过程是达尔文式的吗？《心理学探究》(1999), 309–328。[56] Dean Keith Simonton。2023。创造力的盲目变异与选择性保留理论：BVSR 的最新进展与现状。《创造力研究杂志》35, 3 (2023), 304–323。[57] Ut Na Sio、Kenneth Kotovsky 和 Jonathan Cagan。2015。固着还是灵感？关于范例在设计过程中的作用的元分析综述。《设计研究》39 (2015), 70–99。[58] Andrea J Stone。1995。《来自冥界的图像：Naj Tunich 与玛雅洞穴绘画传统》。德克萨斯大学出版社。[59] Theresa Jean Tanenbaum 和 Karen Tanenbaum。2008。即兴与表演：与故事互动的模型。载于《国际交互式数字叙事会议》。

[60] Gemini 团队、Rohan Anil、Sebastian Borgeaud、Yonghui Wu、Jean-Baptiste Alayrac、Jiahui Yu、Radu Soricut、Johan Schalkwyk、Andrew M Dai、Anja Hauth 等。2023。Gemini：一系列能力强大的多模态模型。*arXiv* 预印本 *arXiv:2312.11805* (2023)。[61] Brad R Turner。2025。叙事可供性：故事能做什么，不能做什么。《管理学会年鉴》ja (2025), annals–2023。[62] 列夫·谢苗诺维奇·维果茨基和Michael Cole。1978。《社会中的心智：高级心理过程的发展》。哈佛大学出版社。[63] Yuxin Wang、Jieru Lin、Zhiwei Yu、Wei Hu 和 Börje F Karlsson。2023。借助结构化知识增强进行开放世界故事生成：综合综述。《神经计算》(2023), 126792。[64] Zhenhailong Wang、Shaoguang Mao、Wenshan Wu、Tao Ge、Furu Wei 和 Heng Ji。2024。释放大语言模型中涌现的认知协同：通过多角色自我协作实现的任务求解智能体。载于北美计算语言学协会分会：人类语言技术（*NAACL-HLT*）。[65] Robert W Weisberg。2006。创造性思维中的专长模式：来自案例研究的证据。载于《剑桥专长与专家表现手册》，K Anders Ericsson、Neil Charness、Paul J Feltovich 和 Robert R Hoffman（编）。剑桥大学出版社，761–787。[66] Kevin Yang、Dan Klein、Nanyun Peng 和 Yuandong Tian。2023。DOC：通过详细大纲控制提升长篇故事连贯性。载于年会计算语言学协会（*ACL*）。[67] Lili Yao、Nanyun Peng、Ralph Weischedel、Kevin Knight、Dongyan Zhao 和 Rui Yan。2019。规划并写作：迈向更优质的自动叙事。载于*AAAI*人工智能大会论文集。[68] Ann Yuan、Andy Coenen、Emily Reif 和 Daphne Ippolito。2022。Wordcraft：使用大语言模型创作故事。载于智能用户界面国际会议。[69] Huan Yuan、Kelong Lu、Mengsi Jing、Cuirong Yang 和 Ning Hao。2022。创造力枯竭中的示例：示例特征与个体差异在创造力中的作用。《人格与个体差异》189（2022），111473。[70] Eric Zhou 和 Dokyun Lee。2024。生成式人工智能、人类创造力与艺术。《*PNAS Nexus*》3，3（2024），pgae052。https://doi.org/10.1093/ pnasnexus/pgae052[71] Jianlong Zhu、Antonios Liapis、Sebastian Risi、Rafael Bidarra 和 G Michael Youngblood。2018。面向设计师的可解释人工智能：以人为中心的混合主动式共创视角。载于*IEEE*计算智能与游戏会议。[72] Yuxi Zhu、Simone M Ritter、Barbara CN Müller 和 Ap Dijksterhuis。2017。创造力：在创意筛选中，直觉加工优于审慎加工。《实验社会心理学杂志》73（2017），180–188。