arXiv:2601.05401v1 [cs.HC] 8 Jan 2026

原型采样：通过画布驱动的视觉人工智能生成实现采样与原型制作的自由形式融合

Alicia Guo
Autodesk Research
多伦多，ON，加拿大
alicia.guo@autodesk.com

David Ledo
Autodesk Research
多伦多，ON，加拿大
david.ledo@autodesk.com

George Fitzmaurice
Autodesk Research
多伦多，ON，加拿大
george.fitzmaurice@autodesk.com

Fraser Anderson
Autodesk Research
多伦多，ON，加拿大
fraser.anderson@autodesk.com

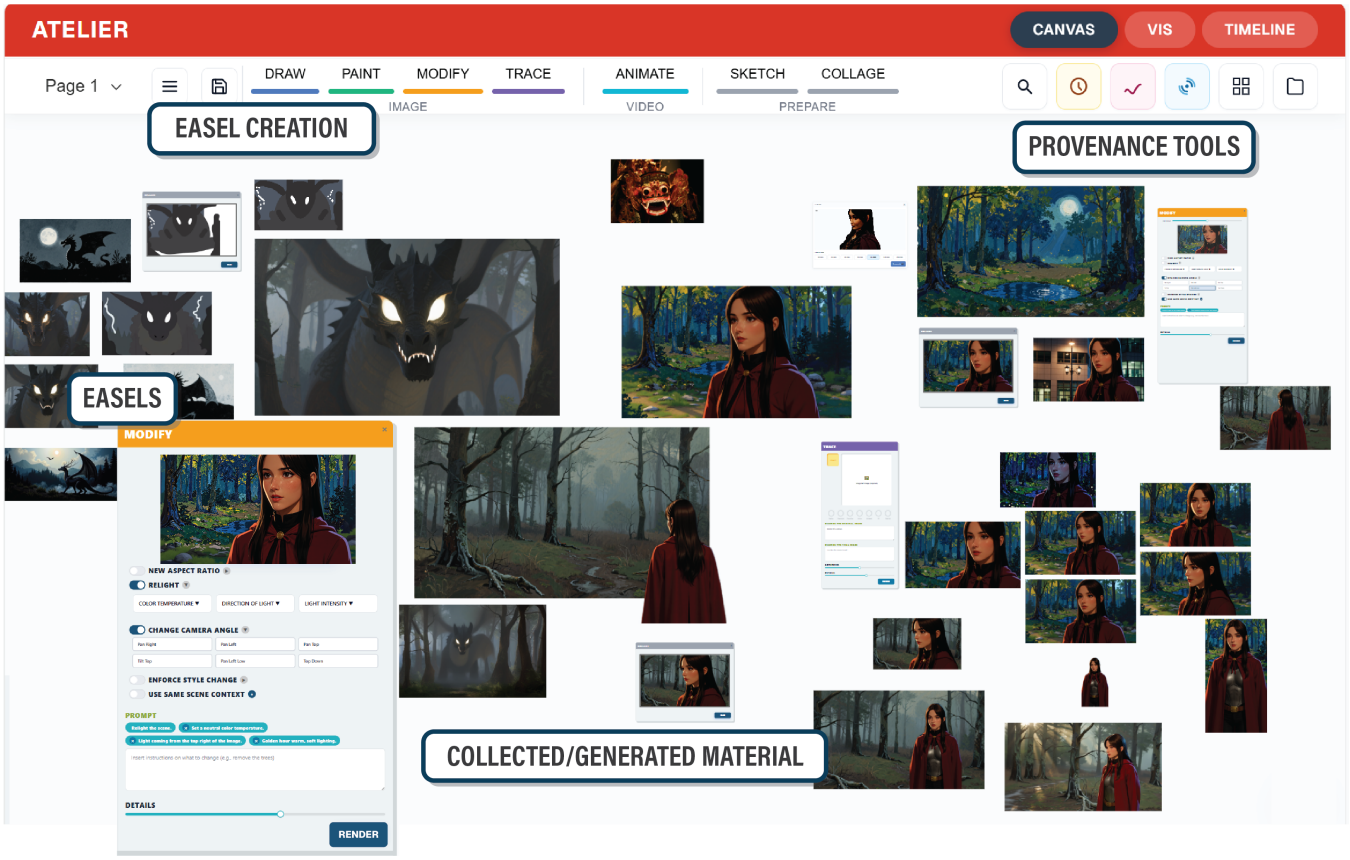

图 1：Atelier 是一个将原型采样付诸实践的系统——通过紧密结合采样与原型制作来创作人工智能视觉内容。它让收集的素材与生成的素材共存，提供一组封装复杂工作流的「画架」，并通过溯源促进收藏与反思。

摘要

创造力是一种不断演进的处理过程，依赖通过采样和原型制作进行探索，以构建问题。这些活动汇集知识，为解决方案营造背景，并回答问题。借助生成式人工智能，从业者可以超越对现有媒体的采样，转而即时生成和混合新媒体。我们将这种融合称为「原型采样」。基于现有文献，我们对原型采样作出定义，并通过 Atelier 将其付诸实践。Atelier 是一个类画布系统，能够利用用于视觉创作的多种生成式图像与视频模型。Atelier：（1）将思考与创作的空间融为一体，让参考资料与生成的素材共存于同一空间；（2）提供多种封装的技术工作流，专注于当前活动；（3）通过交互式可视化、智能搜索和集合，实现对涌现过程的探索。原型采样作为一种视角，重新诠释了创意工作，强调处理过程本身，以及看似彼此无关的想法如何紧密交织成最终解决方案。

CCS 概念
• **以人为中心的计算 → 人机交互（HCI）**。

关键词
创造力支持工具、生成式人工智能与创造力、人工智能图像与视频、采样、原型制作、基于画布的交互

1 引言
生成式人工智能的快速发展，催生了利用自然语言作为主要方式生成丰富媒体内容（例如图像、视频、三维模型、音频）的新能力。随着时间推移，更多机制进一步提升了表现力和控制力：（1）提供多模态输入，例如使用图像来辅助构建新生成图像的结构 [105], ；（2）显示会影响最终生成结果的相关参数。生成式人工智能界面涵盖广泛，服务于不同受众。一端是简化的类聊天系统，借助对话语言快速提供解决方案。其次，有些工具提供一定程度的输入和控制选项。例如，在 Midjourney [63] 中，用户可以输入提示词、选择宽高比、提供风格参考等。最后，处于这一光谱最末端的是程序化工具，它们提供种类齐全的模型、输入和参数设置，但需要用户具备专业知识。ComfyUI [20], 是其中一种流行系统，采用基于节点的编程方法，在特定时刻组合并控制多个模型。这些系统将关键的原子操作封装为节点，使多个模型、算法乃至在线服务能够相互连接。尽管这种范式内容丰富且表现力强，但它主要侧重于通过程序化思维解决问题，并要求用户搜寻大量模型，从中选出最适合当前任务的模型。

这一光谱忽略了一个相当大的缺口：创意从业者需要在保持一定控制权的同时，通过能够最好地表达其意图的媒介开展创作。这意味着要让他们的媒体和处理过程成为核心，并在需要的时间和地点提供易于使用的工具。聚焦于创作过程，便有可能发挥生成式工具的一种潜在特质——能够结合多种多样的素材与思维，利用它们生成新素材，并继续参与广泛且不断涌现的重组过程[21], ，与眼前的创意问题展开真正的对话 [83]。这提出了两种关键创作活动的交织：采样与原型制作。并行考察这两种活动，有助于重新思考创作媒介：从业者以眼前的创意问题为视角观察世界、筛选信息，并积极将其转化为局部解决方案，随着时间推移，这些方案将共同塑造最终成果。

为探索如何支持采样与原型制作之间的相互作用，我们设计了 Atelier（图 1），这是一个画布式界面，其中包含源材料和生成内容，以及一组以活动为中心的小组件，提供一系列精心挑选的输入和参数，用于执行生成式人工智能操作。Atelier 提供多种机制，让用户通过集合、搜索、历史记录和交互式可视化与内容互动，从而反思并重新诠释作品集，以探寻处理过程中可能涌现的特征。一项面向 5 位创意专业人士、持续 4 小时的扩展首次使用研究，进一步展示了该系统的表现力，以及它在现实场景中的潜在用途。具体而言，本文提出以下贡献：

(1) 对“原型采样”的界定——将采样与原型制作视为创作过程中的一项联合活动，而非彼此分离的阶段。(2) Atelier，一个将原型采样付诸实践的类画布系统，旨在支持媒体生成，尤其侧重视觉内容（图像与视频）。(3) 一套用于生成人工智能图像与视频的抽象概念和新颖工作流，聚焦于创作可控性，并通过一系列使用场景加以展示。

2 原型采样
创作过程通常被描述为从业者在解决创意问题时经历的一系列离散阶段 [42, 82]。关于创造力与设计处理过程的表述不胜枚举 [1, 22, 42, 46, 82] ，它们都将处理过程划分为从问题到解决方案的若干独立步数或阶段。然而，Sawyer模型 [82]是少数承认创造力并不一定按既定顺序发生的模型之一；相反，从业者会开展不同的“实践领域”。在这里，这些阶段是经过简化的，旨在强调处理过程中最常见的主要活动（图2）。这些活动包括：询问（发现问题）、学习（获取知识）、观察（收集相关信息）、玩耍（酝酿）、思考（生成创意）、融合（组合创意）、选择（挑选创意）和制作（外化创意） [82]。Sawyer模型强调，创意工作具有高度可变性，并取决于所开展的活动及执行活动的人。创作过程中进行的许多活动涵盖多个阶段或实践领域。两种常见的活动是采样 和原型制作。

采样指的是收集、整理并转化素材的行为 [88] ，而不仅仅是收集素材。创意从业者会持续对现实世界进行采样 [35, 56], ，不断收集素材和信息，从灵感到与当前问题相关的片段，无所不包。采样是一种“机会性吸收”：人们日常生活时，会遇到与尚未解决的问题相关的物体或情境 [53]。基于这一前提，创意问题成为观察世界的视角；创作者会留意周围环境中的信息，将新信息与现有问题和任务联系起来，并将其融入其中。例如，音乐人可能会录下某个环境中的声景，剪辑音频、去除噪声，并将其用作一首乐曲中的节拍；也可能从自己喜欢的歌曲中找到一段和弦进行，从而获得乐曲创作的灵感。平面设计师可能会因为喜欢明信片的颜色而将其保留下来。最终，素材会经过收集、筛选，有时还会被整理成特定形式（例如情绪板），然后加以利用。作为创意问题解决的一部分，Stemasov等人 [88] 将既有文献中的采样描述为一种开放式、有意识的行为，有助于构建思路。采样为创意工作提供背景，将现有素材集合综合为新的设计，提供潜在的原始素材，收集设计决策，并为重新诠释问题提供触发点。

原型制作则着眼于实际创建，以便更好地理解问题。原型被视为有明确目的的

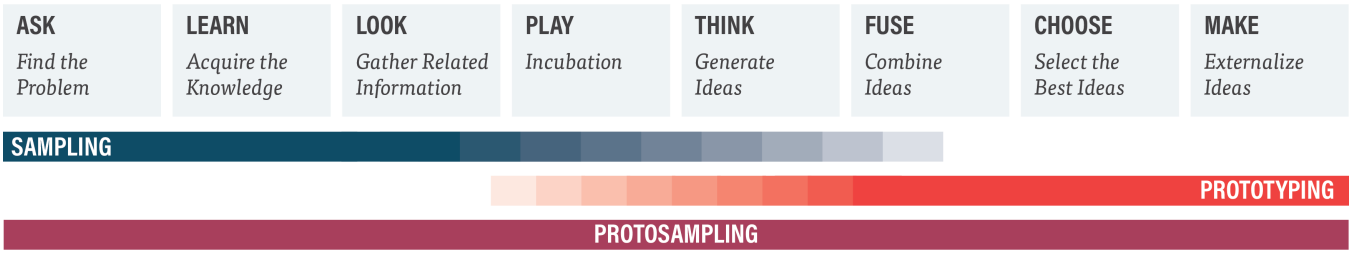

图 2：根据 Sawyer 的创造性处理过程，[82]，将采样、原型制作与构思联系起来，并提出原型采样作为这两种活动之间的桥梁。

设计构想的体现，它们穿越设计空间，从而获得关于最终设计某些方面的有意义知识 [51]。原型制作可以发挥不同作用，例如探索、实验或演化 [29]。特别值得关注的是探索设计空间的能力——探索与实验之间存在一道“模糊”的边界 [29]。虽然 Buxton 将草图与原型 [14], 区分开来，但许多设计文献 [22, 34, 51, 54]倡导以 Buxton 描述素描的方式进行原型制作：原型能够帮助探索设计空间、测试各种方案，并为设计决策提供依据 [51]。事实上，Stolterman 认为，素描是一种严谨的方法，可用于探索和迭代构想。原型的优势在于它们是不完整解决方案，充当“过滤器”，用于检验构想的质量并回答特定问题，而无需创建最终设计 [51]。现有研究通常将原型制作描述为探索设计的特定方面，例如功能、结构、行为、外观与感受、角色、实现等 [21, 33, 51]。Lim等人阐述了原型如何检验设计空间的一个或多个属性 [51], , 例如通过处理颜色或形状等属性来探索外观。虽然“原型制作”一词常用于设计领域，但它也适用于更广泛的创作实践。动画中的原型制作可能包括服务于各种目的的不同活动：分镜脚本用于检验摄像机角度以及故事如何转化为视觉内容，动态分镜则用于探索节奏。为了制作分镜脚本，艺术家可能会绘制一些小型缩略草图来观察乐曲创作。这就是 Logan 和 Smithers [54] 认为原型彼此相互关联的原因。

从广义上说，采样和原型制作都是探索性活动，使从业者能够参与“问题构建” [42], 这一处理过程：问题中的开放性要素逐渐收紧并重新界定，从而推动解决方案的形成。Cross [22] 将问题构建描述为一个始于宏大、高层次目标的处理过程，随后经历一段段密集活动，并在其后进行反思性沉思；直到某个顿悟突然汇聚新的理解，重新定义问题。问题与解决方案共同演进。

生成式人工智能的影响。 尽管目前人们对生成式人工智能对创作过程的影响仍缺乏充分理解，但事实是，它已经从根本上改变了这一过程。文本、图像、视频、三维模型、音乐等媒体都可以快速生成。素材可以在新的生成结果中重新混合和重复使用。此外，这一处理过程意味着，更多部分解决方案能够在更短的时间内以更密集的节奏被创建出来。例如，Ledo 的作品 [48] 展示了生成一段数秒钟的动画需要超过 600 次生成。这种人工智能驱动的创作必然会包含对处理过程的记录，其中每次生成都体现了在当时节点上对问题的部分理解。采样变得更像原型制作，因为素材不仅被收集，也被创建；而原型制作也越来越像采样，因为收集到的和生成的素材都可以组合并重复使用。鉴于这一循环可能正在变得更加紧密，我们认为有必要将这两个概念连接起来：原型采样。

2.1 程序化行动

要实施原型采样，就必须将收集信息所进行的意义建构与外化所采取的行动联系起来，并使这种联系能够实际运作。幸运的是，已有理论和研究对其中许多原则进行了抽象。心理表征被称为概念，并具有带有相应值的属性 [9, 86] （例如，球有大小、颜色等）。我们概述并归类了呈现创意的不同方式，以此作为理解原型采样的视角（图 3）。

2.1.1 衍生。Gero [33] 和 Cross [21] 概述了创意可能经历的渐进变化。我们也考虑创意如何融入创意问题 [58, 89]。值得注意的是，在原型采样中，这些程序性操作类别并非互斥。

变异。对概念进行转变的渐进变化。这可能涉及改变一个或多个属性。例如，引入一个物体并改变其颜色。

重新诠释。 对某一概念的再创作，并在再创作过程中对其进行轻微修改。在这一过程中，某些特定属性可能会发生变化。例如，动画中可能会重新绘制现有绘画，以匹配指定风格。

复制。 忠实地再创作或重复使用现有素材。大多数属性可能保持不变。例如，将一个项目中的素材引入新制作中并加以重复使用。

2.1.2 组合。 创作过程中的想法可以通过组合得以推进。文献将概念组合 定义为“想象力将概念汇聚起来，从而在创作过程中产生新想法的一种心理活动” [39]。创造性理论通常聚焦于语言，并由此推导出六种组合类型 [82]。为简化概念并使组合更具

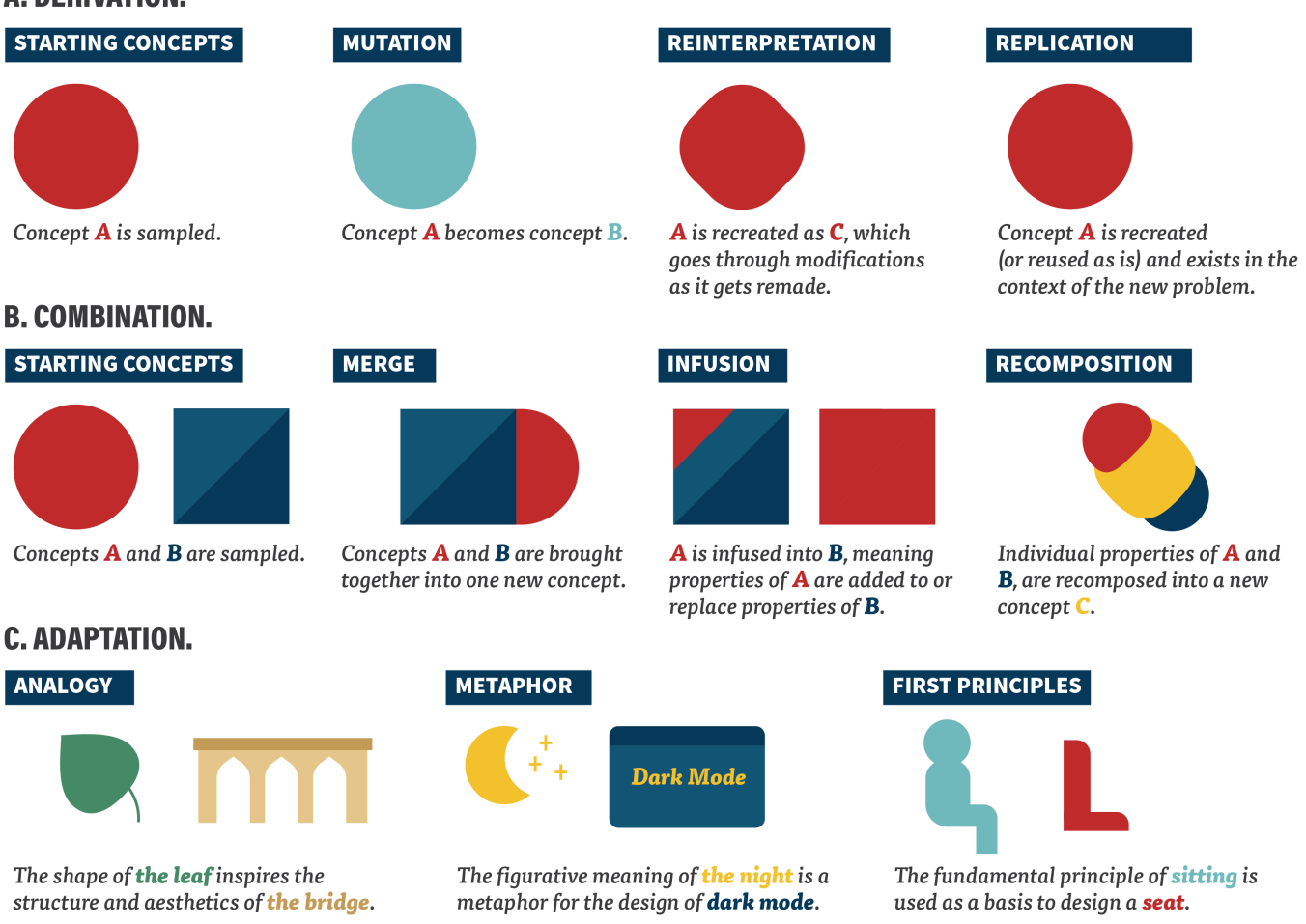

图 3：原型采样中的程序性操作，包括 (a) 衍生，即改变概念属性；(b) 组合，即将两个概念结合起来；以及 (c) 适应，即从其他指导原则中转移概念。

可操作性，我们提炼出三种组合类型，以拓展现有观点 [21, 33]:

合并。 将两个采样概念结合起来，置于创意问题的情境中。这两个概念共存。例如，为报纸文章做准备时，会将多个图像放在一起。

注入。 给定两个概念，将一个概念的属性转移到另一个概念上，从而改变其属性和/或添加新属性。例如，使用图像的色彩调色板来确定页面上字体排印的颜色。

重组。将两个概念拆解，并利用其属性创建一个新概念。例如，可以从两首不同的歌曲中提取不同的音乐采样，并将其用于创作一首新乐曲。请注意，重组意味着先解构源材料，再将其重新组合。

2.1.3 适应。创意作品往往能从将一个领域的概念引入另一个领域中受益 [82]。适应与衍生的区别在于源材料，以及它在创意问题的语境中如何被使用。

类比。 这是一种结构比较，将功能、逻辑或关系转移到某个领域。Christie 和 Gentner 将类比描述为识别不同想法之间的模式或脉络 [18]。例如，地铁地图将实际地点转换为一种抽象形式，只显示相关信息。

隐喻。 这是一种形象化的框定，将意义从一个领域转移到另一个领域 [82]。与组合不同，隐喻依赖于一个与当前主题没有直接联系的概念 [76]。人机交互/用户体验设计中的拟物化就是隐喻的一个例子，例如文件夹和文件系统。

第一性原理。 第一性原理是关于某种情境的基本事实和假设，无法再进一步简化，是对基本原理的回归 [21]。例如，思考椅子时的一个第一性原理是：椅子能支撑处于休息状态的人体，而不是说它是一个有四条腿的物体。

2.2 情境性

原型采样可以被视为一种情境化活动，在工作空间（无论是实体还是虚拟）中进行。从传统意义上说，采样通常与情绪板相关联，情绪板有助于对设计进行悖论化、对齐、抽象化和引导 [56]。创造力理论着重探讨创造力中的构思与思考，而设计理论将原型制作描述为一种实践，但这些领域很少讨论作品本身的处理过程。幸运的是，我们还可以从关于知识工作的既有研究中进一步理解创作过程，因为创意从业者是知识工作者中的重要子集 [10]。

知识工作者处理信息以生成新知识 [43]。与创作过程类似，知识工作被描述为反复无常且非决定性的 [11], , 个人在同一时间处理多种素材。从业者的活动范围可以从强调信息的高度非结构化任务，延伸到强调文档记录的结构化任务 [10]。事实上，正如 Oren 所阐述的，创造知识的关键任务 [74] 与创作过程直接相关，因为它涉及捕捉、组织、形式化和检索之间的相互作用。素材不断被使用和重新混合 [36]。

知识工作研究也强调与工作空间互动的重要性——根据任务的不同，信息会被整理（例如放入文件夹）或成堆归类 [10, 59], , 工作空间则会主动提供情境，帮助人们进入适合的思维状态，而布局如同鲜活的提醒，提示当前活动。素材的外观、尺寸和位置各不相同；它们与工具的距离以及整体摆放方式，都隐含着有关任务的线索，有助于让工作不中断。例如，打印文档可以加上批注，并放在键盘上，笔则放在文档上。当从业者离开时，这份文档便成为一条轨迹 ，记录当前状态：人们可以看出这是什么文档、有多少页、批注进展如何，而它放在键盘旁则表明目前仍在使用。实体文档的可操作性——翻阅页面——无需付出额外成本便能带来情境信息 [11], , 而当文档处于“热” 状态时，使用搜索等机制的必要性也会降低 [84]。这就是为什么 Logan 和 Smithers 认为，素材和原型能够唤起记忆，帮助人们不断推进工作，逐步找到解决方案。与此同时，与当前任务关联较小的信息会被存放起来，以减少杂乱 [36, 84],，这会降低其可见性——它们虽可搜索，却也容易被遗忘 [36], , 只有在特定需求[41] 和具体情境下才会被取用 [75]。了解信息管理方式，有助于阐明创意从业者如何构建信息搜索，并将相关信息随手置于可用之处 [66]。

因此，创意从业者采用变通方法，在使用数字工具时模拟实体空间，也就不足为奇了。Frich 等人 [31] 描述了在数字工作空间周围留出边距，以便保留相关素材和过往版本。这表明，原型采样正在各个领域得到积极应用。因此，人们往往希望创建能够在恰当的地点、恰当的时间 [13]访问和操作信息的工具。

知识工作揭示了信息在处理过程中所具有的重要性，也让我们得以体会信息碎片化带来的挑战。能够追踪处理过程中的轨迹 ，或许具有额外价值，而在素材积极数字化的过程中，这一点往往容易被忽视。尤其是在思考酝酿时，这一问题更为突出：人们离开工作一段时间后，往往会获得顿悟 [22, 82]。

3 相关工作

本研究建立在此前支持构思采样的系统，以及提供多种输入与输出交互方式的生成式人工智能界面之上。

3.1 素材作为灵感与样本

数字和实体情绪板让创作者能够探索设计空间、将想法可视化，并通过以空间方式排列和收集参考素材，与他人分享愿景 [44]。这些情绪板有助于梳理问题并统一美学或概念方向 [57], ；若将其视为会随项目不断演变的主动型制品，效果则会更好 [62]。近期研究将采样视为一种核心创作活动：创作者捕捉并收集素材，甚至对其进行整理和混搭。MoodCubes [40] 和 VRicolage [88] 便是此类系统的例子，它们突破了平面情绪板的局限，借助空间化和多模态采样，鼓励对参考资料进行重新组合。

生成式人工智能能够快速、有针对性地生成灵感素材 [78, 93] ，并支持在各种视觉领域中快速构思 [6, 72, 95]。GanCollage [93]将生成式人工智能融入情绪板，为生成的图像添加语义标签，以探索潜在视觉空间。这些交互侧重于探索变体，而非创作或编辑作品。其他系统则通过语义标注和搜索来增强画布情绪板，帮助用户更好地浏览和反思这些视觉组织结构可能较弱的空间 [44, 45, 78]。这些研究将找到和生成的作品视为参考资料，供之后在实际创作过程中使用。

3.2 使用生成式人工智能界面创作媒体

生成式人工智能系统在处理输入与输出之间的关系、生成结果的产生方式，以及这些交互的排列方式上各不相同。

基于提示词的界面。 ChatGPT [73]和 Midjourney [63] 等系统主要将自然语言提示作为一种易于使用的方法，借助人工智能生成输出结果。这些界面以线性形式呈现提示词和输出结果，每次只能进行一个探索线程。输出结果天然按时间顺序排列，导致先前的生成结果被埋在历史记录中。这些方法易于使用，但无法控制输出结果，也难以发现可执行的操作。

基于节点的界面。 在控制程度的另一端，是 ComfyUI [20] 这类系统，它们通过基于节点的界面显示模型的内部参数，并允许用户将模块串联起来，创建用于生成媒体的工作流。这类系统对最终生成结果的控制程度更高，但仍以一次执行一个工作流为主。随后，可以将最终输出结果传回，作为新工作流的输入，而中间步骤则会被清除。基于节点的界面优先呈现用于原子操作的可视化组件，每个节点都可以加载文件、运行操作或保存输出结果。

一些系统致力于让基于节点的编程变得更易上手，例如 FLORA [28], Runway 工作流 [79], 和 Figma Weave [27]。这些工具将媒体转化为节点，从而能够使用多张图像，并在处理过程中呈现输出结果。这些系统强调对象之间的链接，凸显正在进行的操作。其思维过程仍主要围绕基于图的工作流展开，每次执行工作流时，输出结果都会被清除并重新生成。事实上，FLORA 将自己的方法描述为“专为以系统思维而非仅关注输出结果的专业人士打造” [28]。

将直接操控引入人工智能生成。在强调提示词和强调操作这两个极端之间，一些系统提供直接操控，并支持将文本提示与其他模态相结合，例如素描[19, 52, 80, 103], 图像参考 [50, 90],区域遮罩 [23],可组合提示组件 [5], 以及其他用于完善提示词、反复迭代输出结果的方法。

3.3 虚拟画布

画布界面是二维平面，用户可以在其中导入、创建或修改内容；有些画布界面专为头脑风暴和协作而设计，例如 Mural [67] 或 Miro [65], ；另一些则更适合作为记笔记的白板，例如 MilaNote [64]。画布支持在空间中排列媒体，让素材同时可见，并支持分支式、非线性的探索 [44]。其他研究则探索了将属性（例如形状或颜色）作为可操作的一等对象 [100],、支持自由绘制和收集资源 [37]，或将媒体与工具结合并作为关注重点 [13, 92, 100]。

随着生成式人工智能的发展，面向画布的新型界面开始涌现。ImaginationVellum [60] 利用“生成式笔触”快速创建不同设计理念的变体，并借助文本的空间关系（例如标注风格）在不同概念之间进行调节。Adobe Firefly Boards [2]提供画布，用于整理和显示生成的内容。Firefly Boards 提供一种全局通用界面，类似于屏幕底部中央的基于提示的文本框，并可选择导入风格和结构参考，从而访问许多在线模型。

3.4 迈向原型采样的实际应用

生成式人工智能已开始重塑视觉领域中的创作工作流 [17, 55, 85, 104], , 加速探索并支持快速生成作品。生成式人工智能的速度和能力不断提升，使采样与原型制作日益融合，将思考与创作合并为联系更紧密的处理过程。Firefly Boards 等系统 [2], Paratrouper [50]和 ImaginationVellum [60]已开始朝这一融合方向发展。Firefly Boards 提供多种在线模型，其图像生成可使用一张风格图像和一份结构参考。Paratrouper [50] 支持组合与衍生，同时提供更有条理的探索方式，并隐藏参考素材。ImaginationVellum [60] 擅长将衍生和草图融入图像。然而，这些系统着重于特定方面，而未能将采样与原型制作全面整合为统一的空间实践。

借助 Atelier，我们通过一种以媒体为先的画布来实现原型采样，让用户能够使用多种本地开源模型和算法来处理并生成内容。交互以模块化小组件的形式呈现，将各项活动（绘制、绘画、描摹、修改、Animate）抽象化，使创作者能够专注于任务本身，而不是像使用 ComfyUI 等工具时那样，需要选择模型或编写工作流 [20]。与此同时，这些小组件会展示所有可用输入，素材——包括输入和输出结果——都共存于同一视觉空间中。因此，我们限制了基于节点范式的高度灵活性，同时仍提供丰富的参数供用户访问和调整。我们还会以对创作者更实用的方式转换术语，避免使用过于技术化的表述（例如，将“去噪百分比”改为“保留原图”），并隐藏需要更专业技术知识的功能，例如选择采样器和调度器。因此，其新颖之处在于其方法、交互方式和抽象化处理。

4 设计原理

为了构建一个实现视觉媒体原型采样的系统，我们结合了（1）以往工作的经验，（2）对创作者使用图像和视频生成工具时在线视频的定性观察，（3）创造力理论，以及（4）我们在持续制作系统原型并迭代改进的过程中对其进行测试的自身经验 [49, 69, 87, 107]。这些内容共同塑造了我们的设计原理，以应对通过原型采样支持思考与创作所面临的挑战与机遇。

R1. 融合思考与创作空间。要缩短采样与原型制作之间的循环，就需要一个既能鼓励开放式探索与创建，又能将素材汇集在一起的工作空间。素材应当能够被引入、生成和重新组合。因此，这类界面应以媒体为核心，并利用空间布局来表达含义和组织信息，因此需要采用类似画布的交互方式。这也能支持冷热区域 [84],，让创作者根据自身需求堆放信息。为使采样媒体与生成媒体得到同等使用，它们应当无法被区分。

R2. 按活动封装技术工作流程，同时提供控制权。使用多种生成式人工智能模型意味着，创作者通常需要思考应使用哪种模型，并确定相关的设置和参数。这一处理过程可能相当技术化，从而打断创作者的工作流程。需要确定应显示哪些参数，使创作者拥有能动性和控制权，同时又无需花费大量时间学习或调整参数。界面必须为模型显示适当层级的详细信息，并使用能够生成一致且连贯媒体的参数。

R3. 通过轨迹突出处理过程。使用非线性自由形式画布的一个自然结果是，项目不再具有按时间顺序排列的关系，也不再存在从视觉输入到输出的直接连接。因此，为每个素材保留丰富的溯源历史非常重要，并以一种让创作者能够轻松重新查看素材、理解其产生过程、了解哪些输入生成了它以及它后续产生了什么的方式显示这些信息，同时能够看到项目之间的这些连接。这些功能应帮助创作者重走自己的步数，理解过去的决策，并重新访问处理过程中的早期节点，以探索新的替代路径。

R4. 支持组织、集合与可探索性。由于用于思考与创作的画布体现了创作过程的具体呈现，创作者能够组织和探索信息至关重要。创作者必须能够在恰当的地点和时间获取素材。尽管创作者自然采用的空间布局或许能提供部分支持，他们也应能够搜索素材并创建相关集合，以便轻松访问重要内容。

5 Atelier

为将原型采样付诸实践，我们设计了 Atelier（图 1），让思考、创作与反思在同一空间中发生。Atelier 将画布视为一个动态的工作空间，媒体在其中作为一等对象，可以自由排列，并进行拆解、采样、生成、混合与重组。

5.1 概述

画布充当工作空间，类似艺术家工作室：素材被带入其中、加以排列和处理，从而创作出新的成果。我们支持通过拖放上传图像、视频、文本、音频和三维模型，也支持直接在画布中创建文本和草图。所有媒体素材都可以移动和分组，随后立即通过快捷操作进行转换，将其拆解为可重复使用的组件，使用户能够在进入更复杂的处理过程之前，对素材进行实验和调整。

由于画布支持空间布局，创作者可以自由组织信息，也可以进行临时安排，让意义随时间逐渐形成。收集和生成的素材相结合，能够激发思考，并持续促使人们围绕可能引发重新诠释的因素进行交互参与 [26]。

在实体空间中，互动往往与发生互动的环境紧密相连 [25]。这一点在艺术家工作室中尤为明显：不同的“区域”承载着某项活动的意义，附近的素材与当前任务相关联，而素材也可能随时间迁移或循环回流。这启发我们将活动局部化：快捷操作可以直接在媒体上执行，而更大型的操作则可以在局部工作站中完成。我们将这些工作站称为画架，以提供细致的控制和结构。所有创建内容都会直接带入画布，并被视为可与其他内容共存的潜在新素材。

以往的研究常探讨应用程序如何将工作与特定时间绑定（例如，[13]）。因此，工作空间的状态反映了截至当时的处理过程，其历史记录蕴含着重要信息。这促使我们探索组织、意义建构和溯源的方法。

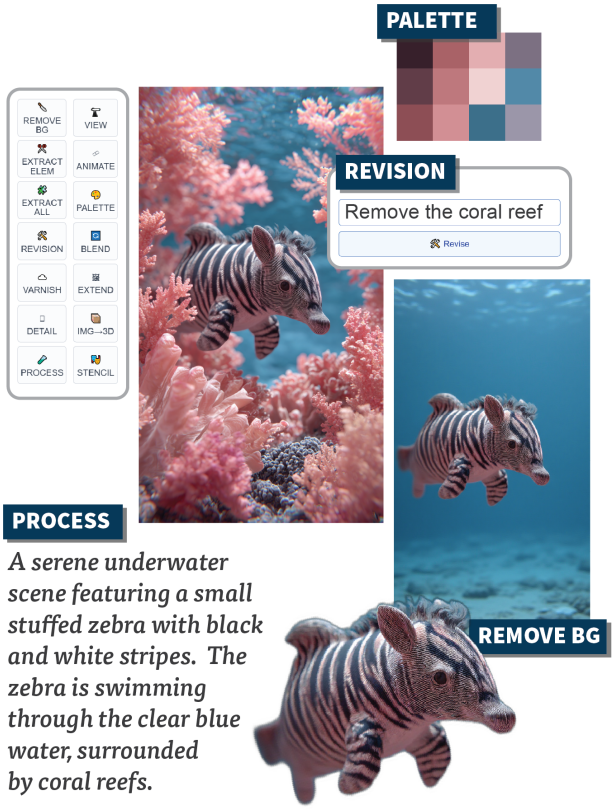

图 4：快捷操作。Atelier 中的图像展示了常用的图像处理功能。示例包括：修订，可对图像进行编辑；处理过程，可分析图像、生成 ControlNet 预处理器并为图像添加说明；以及移除背景。

5.2 快捷操作

可立即对媒体执行的操作是指几乎不需要额外输入的操作，例如根据文本生成**快速草图**，即将文本内容作为提示词来生成图像。对于图像，创作者可以使用移除背景（保留前景对象，并将背景设为透明）、提取元素（提取“蓝色的花朵”等指定元素）、调色板（创建调色板）和模板（生成深度掩码、线稿、姿势等结构控制图像，用于控制后续生成结果），迅速将素材拆解为更易复用的组成部分（图 4）。创作者还可以使用修订（进行单次小幅修订，例如“给兔子加一撇小胡子”）、放大（将图像放大至更高分辨率并增加细节）和混合（统一风格和灯光），对图像进行优化和微调。最后，他们可以使用扩展（将图像延伸到当前帧之外）来扩展图像，并使用视图（生成不同视角的视图）、快速动画（类似于快速草图，无需额外提示词即可根据视频生成 5 秒视频）和雕刻（将图像转换为三维

模型）。三维模型可以进行操控和旋转，并捕捉为新的图像，作为角色和透视参考。

5.3 画架：空间模块化工作站

画架是汇集工作空间素材并用于生成新媒体的工作站。与工作室工作流类似，人们会将素材实体地汇集在某个区域周围，以便使用；在 Atelier 中，素材会在使用前被移至画架附近。每个画架都围绕特定目标封装一种生成行为，并显示不同的控件；这些控件是后端不同 ComfyUI 工作流的抽象。生成结果不会破坏原始内容：输入内容会被保留，可重复使用并不断迭代。画架旨在作为溯源标记，其存在及周围的素材表明该区域蕴含着复杂的创作意图。

5.4 用于准备素材的画架

我们提供两种轻量级画架，帮助用户构思和起草概念，再进一步进行更复杂的生成。它们提供了汇集、整理和外化想法的空间，类似于先绘制缩略草图，再着手完成最终作品。

拼贴 画架（图 5）是一个可以自由排列多张图像并将它们“粘合”在一起，从而创建可重复使用的新合成图像的区域。它可用于勾勒场景中角色和物体的布局。例如，创作者可以从一张图像中提取角色，将另一次生成的图像用作背景，再将两者拼贴到一起，同时调整大小和位置。**草图**画架支持徒手绘图，可用于引导后续的图像或视频生成。这些用于前期准备、能够输出图像的画架，使草图和拼贴作品可以反复使用，并支持随时间进行迭代调整。

5.5 用于生成素材的画架

图像画架（Draw、绘画、描摹、修改）包含用于生成新图像的工作流，并通过抽象化的细粒度控制，融合文本提示、图像参考、参数滑块和图像蒙版。Animate 画架支持以图像作为起始帧或结束帧来生成视频。

绘制。 Draw 画架（文本转图像）接收文本作为创作者想要生成内容的正面提示 。可选输入包括：预设风格选项 （写实主义、Dreamlight、动漫、复古动漫、动画风格、三维、像素艺术）；用于设置最终图像细节量的细节 滑块；用于设置模型遵循提示词程度的遵循度 滑块；一张起始图像；以及用于设置原始图像保留百分比的保留 滑块。使用起始图像 类似于引入一幅已有的绘画并在其上作画，原有特征可能会显现出来，但最终生成的是一幅新的绘画 [48]。

绘画。 Paint 画架（图 6）扩展了绘制的功能，使用户能够在文生图的基础上添加最多 3 个图像参考，并调整其影响强度。此外，还可以绘制参考蒙版，指定应用这些参考资料的区域。也可以从画布上的任意图像中选择一张结构图像。这些结构图像用于控制

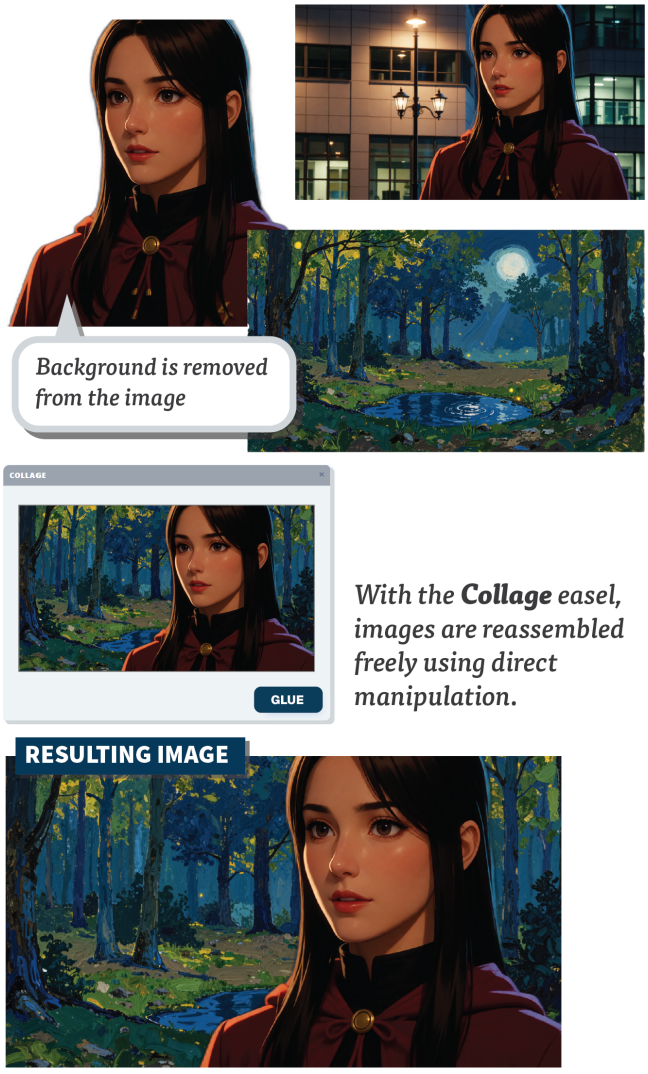

图 5：Collage 画架可将图像组合成一幅新图像。此示例中，一个角色从背景中提取出来，并置入一个新场景。

布局（深度、线条、涂鸦姿势），并可调整强度。额外的提示框支持负面提示，创作者可以在正面提示之外，指定不希望出现的内容。风格选项、细节、遵循度和保留 滑块与绘制中的相同。

描摹。Trace 画架支持基于现有图像生成新图像，并有选择地重新绘制其内容和结构。通过使用两个提示词，一个描述当前图像，另一个描述期望图像，系统可以解构输入图像。与绘制和描摹类似，系统提供了用于调节提示词遵循度、风格选项、细节和保留原始图像的滑块。额外的重新描摹范围范围滑块决定在解构过程中的哪个阶段进行重新绘制。画架还可选择使用输入图像中的结构图像、风格选项、细节，

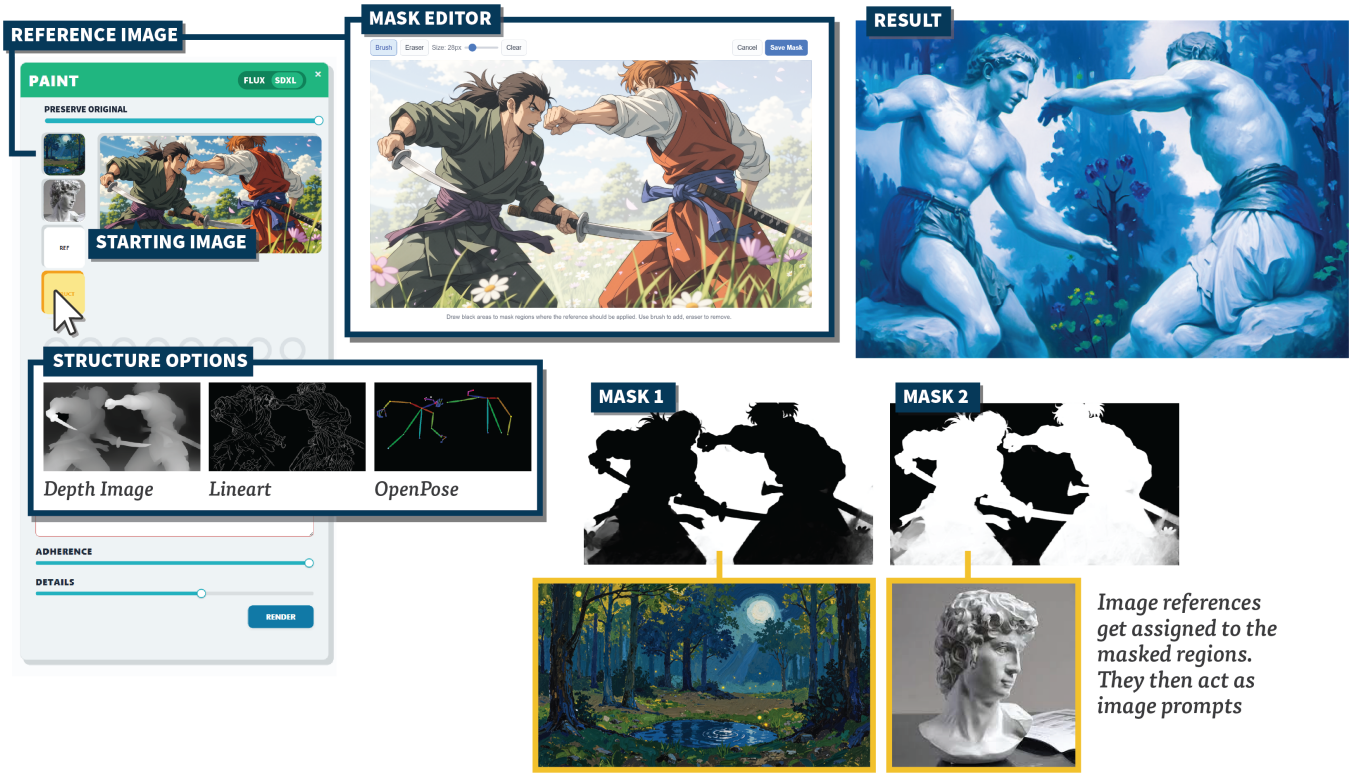

图 6：Paint 画架的说明，展示了如何添加参考图像、起始图像和结构图像。参考图像框允许设置强度，还可以打开蒙版编辑器，以确定该参考图像具体影响的区域。

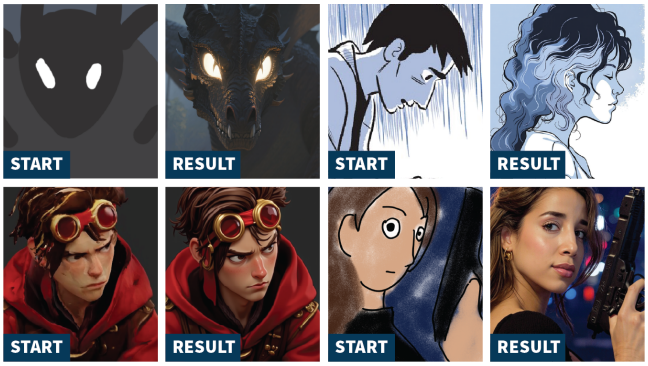

图 7：使用 FlowEdit（描摹）创建的示例图像。

保留和遵循度滑块。描摹操作尤其适用于彩色绘图、粗略的三维模型或图像风格重塑（图 7）。

修改。 修改画架支持应用广泛的编辑操作。为了帮助用户进一步发现这些功能，我们提取了一组电影化编辑。用户可以在不造成重大失真的情况下更改图像的宽高比。Atelier 提供了一组用于调整色温、光线方向和光照强度的重新布光 预设，以及一组摄像机角度和风格建议。点击这些建议后，它们会成为可叠加的提示词胶囊，即附加到提示词中的预设文本字符串；创作者还可以在提示词中指定自己的自定义修改。与其他绘制画架类似，这里也提供遵循度 和细节 滑块。

Animate。 动画画架通过支持为首帧 和/或末帧指定图像来创建视频，输出视频将在正向 和负面提示的引导下，在两个帧之间生成平滑过渡。两个帧都是可选的，因此可以指定起始帧 而不指定结束帧 ，反之亦然；也可以不提供任何帧，直接进行文本到视频生成。用户还可以从一组预定义的运动中选择镜头运动，创作者可通过提示词胶囊 将其添加到提示词中；这些提示词胶囊还可作为创作者构建自定义运动提示词的示例。生成的视频会直接添加到画布中，与其他素材并列显示，并可分解为静帧。

5.6 溯源、组织与意义建构

为了让非线性、非按时间顺序排列的自由形式画布具有结构，我们提供了直接融入画布的功能，以支持溯源和组织。Atelier 中的溯源如同一条不断延伸的轨迹，每个素材都带有其溯源信息，例如它是导入的还是生成的、生成它所用的画架类型，以及所使用的参数和素材（图8）。

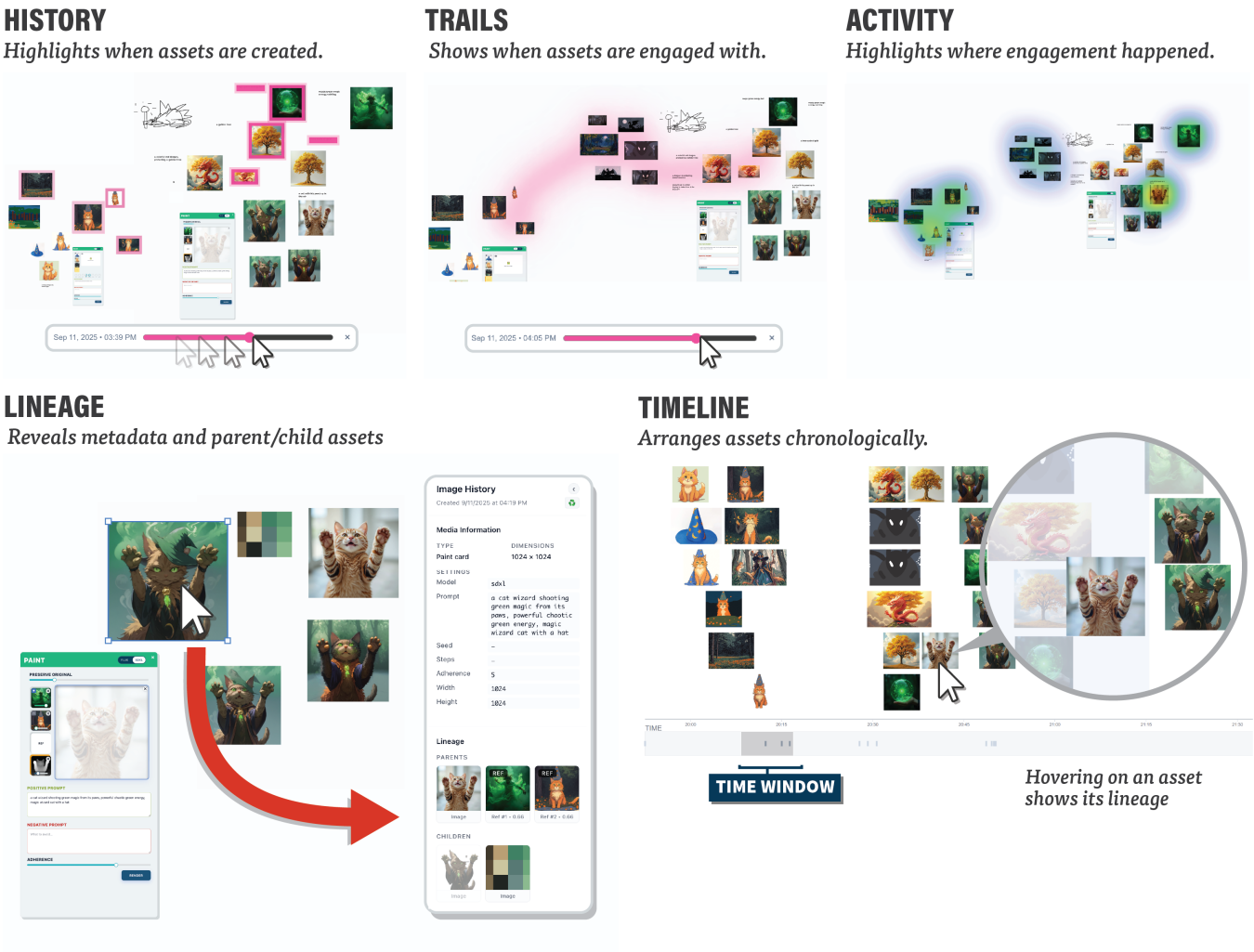

图8：Atelier 中的溯源示例。图中展示了历史记录、轨迹、活动、血缘关系和时间线。

溯源信息。 每当选中一个素材，沿革面板就会显示所有相关信息，包括生成它的父级资产，以及由它衍生出的所有子级资产，并在画布中高亮显示这些素材，方便快速跳转。已删除的图像不会显示在画布中，但仍会记录在数据溯源图中，并且可以恢复为活动资产。每个由画架生成的素材都带有一个重新创建 按钮，创作者可以借此显示或重新实例化生成图像或视频的画架，沿用完全相同的参数并继续迭代。对于给定的图像，可以通过沿革面板递归查看其父级，追溯生成它所用的所有参考资料和参数。任何时候，创作者都可以选择重新创建画架，从而以非破坏性的方式从过去的某个分支点继续创作。

资产强调。 输出结果的重要性可能各不相同，因为有些输出结果可能比其他输出结果更有用，或更贴近创作者的意图。因此，我们增加了降低素材强调程度的选项，该选项会降低素材的不透明度。

历史记录。 查看历史记录时，滑块会按时间顺序突出显示画布上的媒体，并标注时间戳；每次显示 5 个元素的滑动窗口，展现创建过程随时间和空间的演进。

轨迹。 另一种呈现交互随时间变化的方式是通过轨迹，它以扩散轨迹的形式展示画布上特定时间发生交互参与的位置，汇总历史记录中呈现的信息，概括整体路径，而非单个元素。

活动。 我们还可以通过叠加热力图，以更直观的方式呈现交互参与度。我们会记录素材的创建时间、上次被选中的时间，以及素材被点击或使用的总次数，以此作为创作者花费在其上的时间的参考。此叠加层使用点击次数来显示

活动区域，并突出显示创作者花费时间最多的区域和素材。

时间线。 整体活动随时间的变化可通过时间线查看。时间线以空间布局的方式将素材排列在视觉时间线上，悬停在某个素材上时，会突出显示其父项和子项。

组织。 每张图像都会自动添加人工智能生成的描述标签。借助描述、提示词和参数文本进行画布范围搜索，即可通过关键词查找媒体。画布还支持将素材分组并自动排列成网格。这有助于近距离浏览大量素材，也能将它们归为一个整体选择。为了支持长期冷存储，我们提供**收藏集**。重复出现的角色、场景和其他媒体组可以保存到收藏集中，并添加标签。之后，创作者可以在画布的任何其他位置从这些收藏集中调取副本，快速重复使用已保存的媒体。即使在不同的画布页面上工作，这些收藏集也会贯穿整个项目。创作者可以将想要保存的图像和视频添加到**展览**图库，为其添加说明并重新排列，无论它们是电影分镜板中的帧，还是独立项目的最终素材。这为当前项目中特定的重要素材提供了存放空间。这些功能有助于回顾创建过程、追踪媒体之间超越历史记录的联系，并以此作为浏览画布和查找相关信息的方式。

6 使用场景

为了说明 Atelier 的一些功能如何协同运作，我们为 Atelier 展示一个使用场景。Fiona 是一位电影制作人，她想构思一个关于战士屠龙的奇幻电影短片概念。

Fiona 打开 Atelier，导入龙和蜥蜴的图像、史诗英雄镜头以及能体现她想要打造的世界美学的奇幻场景（图9）。她看着画布上排列的这些图像，它们就像一块情绪板，让她感受到故事的氛围。



图9：初始画布，展示通过拖放添加的灵感图像。

Fiona 想先确定主角。她写下几段描述，并使用 **快速草图** 快速生成图像。她喜欢提示词“身穿斗篷的女性战士”生成的概念图，但不喜欢角色正面朝前的样子。她使用 **移除背景**，然后使用**雕刻**将她变成粗略的三维模型，并可以将战士调整到任意角度，捕捉不同视图的静帧。



图10：使用三维模型创建战士的不同视图。

她想看看战士置身场景中的样子，于是打开一个**拼贴**画架，将战士的图像摆放在灵感图像中的湖泊场景上，制作一个特写镜头。她拿着这张拼合后的图像，接着打开一个**描摹** 画架，将原始图像描述为“一名身穿红色斗篷的女子站在湖前”，并将目标图像描述为“一名身穿红色斗篷的女子，绘画风格”。生成的图像没有她想要的背景，但 Fiona 喜欢战士经过新风格化后的版本。她再次使用**移除背景** ，然后将图像保存到名为“战士”的新收藏集中。

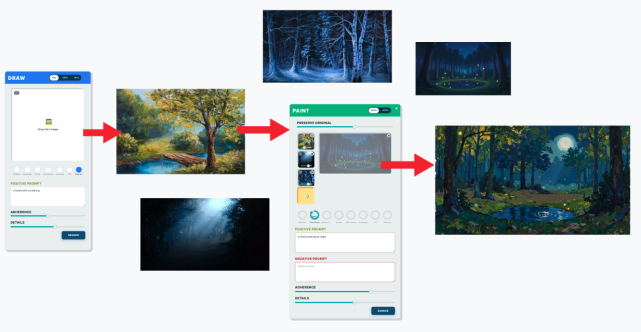

图 11：设计场景。生成一片森林，然后将其与其他森林图像一同用作风格参考。

她现在开始专注于场景的环境：一片森林。她继续探索绘画风格，使用 **绘制** 画架，并输入“一片有林间空地的森林”（图 11）。她喜欢生成结果，但想把时间设定在夜晚，于是使用 **修订**。其中一个结果让她感到意外，也给了她新的灵感；她被那种阴森的氛围吸引，决定转而朝这个方向创作。她打开 **绘画** 画架，将阴森的森林和其他图像作为参考，并输入提示词“一片夜晚的林间空地”。她喜欢第一轮生成结果的绘画风格，于是将其作为后续生成的风格。她打开 **拼贴** 画架，将从收藏集中选取的战士图像特写与森林场景组合起来。Fiona 不确定这种组合效果如何，于是使用 **修订** 来



图 12：探索战士角色的摄像机角度和场面调度

尝试不同视图，并输入诸如“让女人面朝前方”和“展示女人腰部以上”的指示（图 12）。她喜欢其中一些缩小视图，于是 Fiona 再次使用 **移除背景**，提取出她想要角度的战士。现在，她的场景有了几个角度和缩放级别。接下来是龙。她回到早先导入的龙图像，

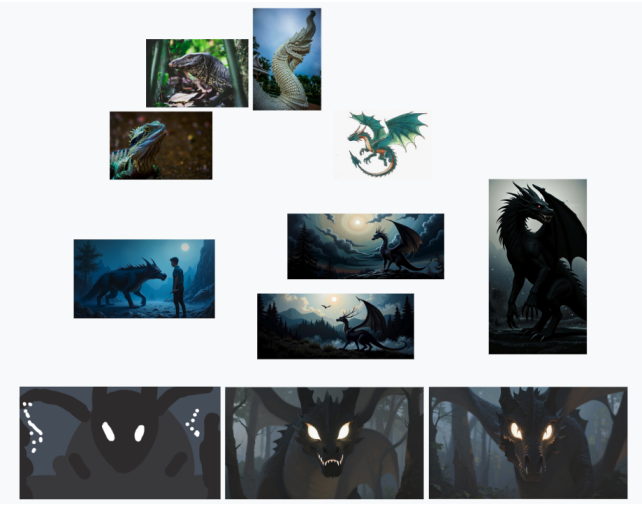

图 13：龙的角色生成探索。

蜥蜴方面，Fiona 尝试不同的画架和组合，生成龙的设计（图 13 ）。最终，她选定了一套灰暗的配色方案。这启发她将龙的场景构思得比其他场景更加阴暗。她在 **草图** 画架上使用这些灰色调绘制龙，然后使用 **描摹**，这次以自己的绘图作为输入。经过多次调整参数后，她得到了满意的设计。接着，她使用 **修改** 画架，生成各种摄像机角度和灯光效果的 16:9宽高比镜头。她最喜欢的作品被加入一个新的“龙”收藏集。

观察画布时，Fiona 想到了一个镜头序列：战士站在森林中，摄像机从她的脸部摇摄至背后，然后推进树林，直到龙现身。她选择了战士在森林中的图像、空无一人的森林、龙靠近的画面，以及最后一张龙的特写镜头图像，并将它们加入图库。回顾这个序列时，她意识到还缺少一个关键帧，用来呈现摄像机位于战士身后的视角，于是使用 **修订** 来创建这一帧。有了这些帧（图 14），她可以使用 **Animate** 画架，将这些图像作为起始图像和结束图像输入，从而创建一段连贯的视频。

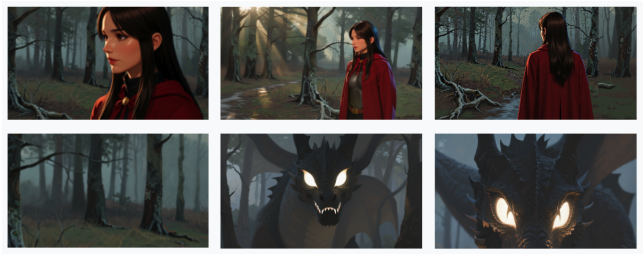

图 14：分镜板中的最终帧，展示了森林中的一名女性战士与一条现身的龙。

7 实现

Atelier 使用 tldraw [91] 无限画布SDK构建，内置文本和图像功能以及形状管理。画布上的每个项目都是一个形状，形状可以指向托管在其他位置的素材（图像、视频等）。我们将画架、生成逻辑、溯源和组织功能作为自定义 React组件，构建在画布之上，并使用 Cursor [7] 辅助实现。所有代码均由两位作者编辑和审阅。每项人工智能操作均通过自定义 ComfyUI 工作流（以 JSON 文件表示）执行。

溯源以有向无环图的形式运作，其中节点对应画布上的项目，而画布被视为画布历史记录的事实来源。图像、视频和三维模型在画布上可能有多个副本，每个副本都会作为一个新节点创建，并指向原始素材。这些副本节点保留了原始来源的血缘关系和元数据信息。父节点是用于创建当前子节点的任何媒体（参考图像、起始图像）。每个节点都会记录关键元数据（创建时间、最后交互时间、点击次数、生成参数，以及指向父节点和子节点的链接）。创建时间用于生成历史视图，点击次数用于生成热力图叠加层。时间线使用 D3[12] 构建。

7.1 工作流实现：通用策略

我们在配备 RTX 5090 和 RTX 4090 显卡的计算机上测试了 Atelier ComfyUI 服务器。这些工作流使用了多种本地模型，包括 Stable Diffusion XL [3], FLUX-dev 和 Flux Kontext [30],以及 Wan 2.2 [94]。我们的许多工作流都采用了三项关键策略，这也是它们区别于目前公开可用工作流的部分原因。前两项关乎如何

对图像进行采样，最后一项则关乎如何处理单个工作流中的多种排列组合。表 1 概述了所有工作流参数如何转换为画架界面。

7.1.1 自定义采样器与 *Lying Sigmas*。 Atelier 的工作流提供的一个新颖方面，是能够为图像添加细节。这是通过 Lying Sigmas 实现的，这是一种在扩散过程中注入噪声的技术[1]。我们使用 Detail Daemon 节点[2]。在 Detail Daemon 软件包中，Lying Sigmas 采样器使用一个用于细节的参数（不诚实因子）。我们选择 -0.05 到 0 之间的值。为了支持 Lying Sigmas，工作流需要将传统的 KSampler[3] 替换为更复杂的自定义采样器，后者支持操控引导器和西格玛值。

7.1.2 归一化注意力引导。 许多人工智能模型支持负向提示词。然而，FLUX 等 Transformer 架构不提供负向提示词，因为它们采用“弱引导”（无分类器引导值为 1.0） [38]。一种解决方法是使用归一化注意力引导 [16], , 从而能够将负向提示词纳入系统[4], ; 我们将其设置为 9.0（建议值为 5.0 到 11.0）。该算法还支持对 Wan 2.2 等为加速 LoRA 而放弃更高引导强度的模型使用负向提示词。归一化注意力引导可以连接到自定义采样器。

7.1.3 处理未激活的节点。每次出现新的提示词时，都会轮询 ComfyUI 工作流。对于某些需要多个输入的工作流（例如“绘画”），执行 API 有四种策略：(1) 根据可用输入以编程方式编辑 JSON 文件，重新连接节点（例如 Paratrouper [50]）；(2) 为每种可用输入组合创建工作流；(3) 将参数设为 0，使其不影响生成（某些功能不支持这种方式）；最后一种，也是我们采用的方法，是使用开关[5]。通过设置单个数字，开关可以动态地重新路由工作流中的节点，从而减少重新连接节点可能导致的错误。

7.2 小型工作流和预制工作流

许多快捷操作通过包含少量节点的 ComfyUI 工作流完成，或复用网上已有的节点，例如使用模板移除背景、提取元素（GroundingDINO[6]）、调色板、放大、扩展（扩图）和模板（控制图像）。其中一些操作是现有画架的精简版本，不包含额外设置（快速素描与绘画；修订、视图与修改、快速动画与动画制作）。雕刻使用混元3D封装工作流 [7]，将图像转换为三维内容。

7.3 预计算元数据

每当上传图像时，我们都会执行预处理步骤来存储元数据，避免重复计算。图像会使用 Florence2 [101], 生成描述，并为 ControlNet 进行预处理：

[1]https://github.com/muerrilla/sd-webui-detail-daemon/ [2] https://github.com/Jonseed/ComfyUI-Detail-Daemon [3] https://blenderneko.github.io/ComfyUI-docs/Core%20Nodes/Sampling/KSampler/ [4] https://github.com/ChenDarYen/ComfyUI-NAG [5] https://github.com/ltdrdata/ComfyUI-Impact-Pack [6]https://github.com/storyicon/comfyui_segment_anything [7] https://github.com/kijai/ComfyUI-Hunyuan3DWrapper

OpenPose [61],DepthAnything V2 [102],涂鸦和线稿。之后，这些元数据会用于其他功能，例如搜索，或为某些画架提供结构图像支持。

7.4 绘制与绘画工作流

Draw 画架对 Stable Diffusion XL 和 FLUX 使用相同的工作流，但不会显示参考图像、负面提示词等功能。不过，Draw 画架还通过我们的自定义工作流支持将 Wan 2.2 用作图像生成器。

7.4.1 Wan 2.2 模型中的文生图。Wan 2.2 文生视频模型 [94] 使用“专家混合”来创建视觉内容。生成视频时，“高噪声”检查点负责运动，而“低噪声”模型则据此生成清晰易懂的高质量视觉内容。生成图像时，我们使用“低噪声”模型来生成单帧。Wan 图像生成模型能够呈现逼真的姿势和表情，这一点是传统图像模型较难做到的。

我们使用 Wan Lightning LoRA，它最初旨在将推理步数从 20 步减少到 4 步，分类器自由引导值范围为 1.0 到 1.5。我们发现，在 20 步时使用 Lightning LoRA，能显著提升渲染质量。我们额外加入了一层强度为 11 的归一化注意力引导，以便使用负向提示词和 Lying Sigmas 来呈现细节。我们发现，与典型的图像模型相比，Lying Sigmas 的值在视频模型中的表现有所不同，似乎也会影响风格化。

7.4.2 使用参考图像、风格和结构生成 *FLUX* 结果。 绘画工作流通过加载 FLUX 模型和编码器，然后嵌入当前启用的 LoRA 来运行。我们将三张参考图像连接到 FLUX Redux[8]，并使用 AdvancedRefluxControl 节点[9] 进一步增强可控性。AdvancedRefluxControl 没有禁用状态，因此，为了避免使用开关或自动重新布线，我们修改了该节点的代码，并添加了一个返回 0 的条件。为了控制结构，我们将预处理元数据中的一张控制图像以固定的结束百分比显示为 ControlNet，并使用 ControlNetUnion [106]（姿势：0.9，深度：0.7，线稿：0.4）。我们还使用归一化注意力引导处理负向提示词，并使用 Lying Sigmas 呈现细节。生成过程通过 FLUX Turbo[4]以 8 个推理步数完成。

由于 FLUX Redux 会影响条件控制，图像参考会同时影响风格和乐曲创作，这可能与用于结构控制的 ControlNet 产生冲突。我们创建了一个替代版本的工作流，其中一个图像参考连接到字节跳动 USO [99] , 仅用于定义风格，同时保持与 Redux 和 ControlNet 的兼容性。

7.5 修改工作流

修改功能使用 Flux Kontext 编辑现有图像，例如更改电影摄影属性。我们添加了一组 LoRA，用于更改摄像机角度、重新布光和风格。此工作流的独特之处在于能够将图像重新构图为完全不同的宽高比。

[8]https://huggingface.co/black-forest-labs/FLUX.1-Redux-dev [9] https://github.com/kaibioinfo/ComfyUI_AdvancedRefluxControl

	参数	设置内容	操作依据		在 Easels 上的操作
GENERALIZED IMAGE/VIDEO WORKFLOWS	宽度	图像尺寸	摘要	根据模型的训练情况，会有理想尺寸（例如，*SDXL* 为 *1024x1024*）。	基于纵横比的固定尺寸：*1:1*、*3:2*、*4:3*、*16:9*、*2:3*、*9:16*、*3:4*。
	高度			如果分辨率过低，模型就难以在合适的位置渲染细节；如果分辨率过高，则可能变得缓慢或生成内容不连贯。	
	潜变量	推理的起点 推理			可设置起始图像。
	检查点	可渲染内容的模型 渲染	预设	这在 *Stable Diffusion* 模型中最为常见。我们预先选择了具有代表性的风格。	*SDXL* 的风格选择。
	提示词	对什么进行条件控制 进行渲染	显示	提示定义了可以绘制的内容。负向提示词可减少不需要的特征。	用于提示词撰写的文本框。 用于触发 *LoRA* 的关键词。
	引导 强度	遵循度与 *the* 提示词	抽象 + 显示	不同模型的最优值各不相同。 引导强度还可能取决于 步数。	名为“提示词 遵循度”的滑块，对应于 特定模型范围。
	Seed	定义初始噪声 用于生成	显示	固定随机种子相比之下能保持更高的一致性，而更改随机种子会带来变化。	仅接受以下内容的文本框 数字。
	采样器/ 调度器	去噪如何 变成图像	*Hide*	最佳采样器/调度器取决于模型 或作为多次实验的结果。	*N/A*
	推理 步数	去噪步数至 渲染图像	*Hide*	模型建议的推理步数为 采样步数。	隐藏。
	去噪	去噪多少 图像	抽象 + 显示	“去噪”是一个复杂的概念，应关注 是否要在图像上进行渲染。	一个名为“保留 原图”的滑块，用于起始图像。
	批量大小	要生成多少张图像 生成	*Hide*	始终只生成一个结果。可以通过新的随机种子值获得变体。 可以通过新的随机种子值获得。	*N/A*
	偏移	采样器如何处 随时间处理噪声	*Hide*	此值通常由模型预设。	*N/A*
	文本编码器	从 提示词到视觉内容	*Hide*	每个模型都有一组预定义的编码器。	*N/A*
	变量 自编码器	翻译自 从像素到潜在空间	*Hide*	每个模型都有一组预定义的变分自编码器。	*N/A*
LORAS	风格 LoRA 检查点	图像类型 和风格 可根据 训练数据集	预设 + 显示	我们精选了可混合使用的风格 *LoRA*（例如，写实主义、三维）。强度应可调节。 编辑模型可使用 *LoRA* 来改善 结果（例如，重新布光），视频模型则可以 使用 *LoRA* 来控制摄影机运动。	绘制：开关 绘画/描摹：径向滑块 修改：用于单项编辑的选项卡 ***Animate***：相机运动。
	风格 LoRA 强度				
	效率 LoRA	减少数量 的推理步数。	*Hide*	部分 *LoRA* 可以减少总步数。更快的渲染会使视频中的运动减少。	***Animate***：显示“质量”（低、中、高）。
DETAILS	不诚实 因子	要注入多少噪声 以添加细节。	抽象 + 显示	不诚实因子（*-0.05* 至 *0*）可以增加 细节，而不会损害结果。	一个滑块重命名为 “添加细节”
	开始/结束	推理步数，应用 噪声注入	*Hide*	会增加复杂度，但不会被直接察觉。	*N/A*
CONTROLNET	预处理 图像	结构组 成搭配使用	抽象 + 显示	我们选择一组具有代表性的预处理器 以便在导入媒体时进行计算。	控制图像选择 在结构参考中。
	检查点	如何处理 图像	*Hide*	*ControlNet Union* 支持多种类型的 预处理图像。	*N/A*
	开始/结束	应用 *ControlNet* 的推理步数 应用 *ControlNet*	固定	*ControlNet* 固定时，最好从 *0* 开始。结束值可以固定。	*N/A*
	强度	的强度 *ControlNet*	抽象	*ControlNet* 的强度决定了 很大程度上遵循结构。	用于调节强度的滑块
IMAGE REFERENCES	图像	参考图像	显示	用作参考的图像。	用于添加图像的框
	检查点	风格模型	*Hide*	我们使用 *Flux REDUX / IP-Adapter Plus*（*SDXL*）。	*N/A*
	CLIP视觉编码器	视觉模型	*Hide*	每个模型都有兼容的 *CLIP*视觉编码器	*N/A*
	注意力 Mask	确定区域 供参考内容影响	抽象 + 显示	仅限 *SDXL* 搭配 *IP-Adapter*，黑白 图像定义感兴趣区域。	用于设置的绘图工具 起始图像上的遮罩
	强度	参考强度	抽象	可以设置参考资料的影响程度	用于调节强度的滑块
	Type	参考资料的作用方式	*Hide*	我们将 *IP-Adapter* 预设为“强风格迁移”	*N/A*
	开始/结束	推理步数 应用参考	*Hide*	我们将参考资料应用于完整推理。	*N/A*
VID	起始帧	第一帧视频	显示	可选起始帧	用于添加图像的框
	结束帧	最后一帧视频	显示	可选结束帧	用于添加图像的方框

表 1：工作流参数转换为画架的依据。

7.6 Trace 工作流

我们使用两种工作流来支持描摹——一种适用于 FLUX，另一种适用于 Wan 2.2。这两种工作流均基于 GitHub 用户 Log(td)[10] 的工作流和自定义节点，并采用一种名为 FlowEdit [47]的技术。FlowEdit 通过逆向工程处理带有标签的起始图像，并根据新的提示词重建图像。这种方法出现于 FLUX Kontext 和 Qwen Edit 等编辑模型之前，因此能够以有趣的方式转换图像。FLUX 工作流基于 Log(td) 的 Fluxtapoz 节点包[11]。我们修改了示例工作流，添加了自定义 LoRA、ControlNet 和 USO 风格图像[99]。在我们的实现中，我们使用 20 个推理步数。

对于 Wan 2.2 图像生成中的 FlowEdit，我们创建了一种工作流，将文生图工作流与受 Zack Abrams[12]启发的工作流结合起来；后者修改了一个用于 Wan 2.1 图像转视频模型的混元视频 FlowEdit 工作流[13]。

8 Atelier 的扩展式首次使用探索

我们邀请创意专业人士使用该系统进行了 4 小时的首次使用探索（包括休息时间），并将其设计为开放式环节，以便留出时间适应学习曲线。我们的目标并非评估 Atelier 的表现，而是了解创意专业人士可能会如何将 Atelier 这样的工具纳入自己的实践。为此，我们采用了结合引导演示与观察的研究设计 [49]。具体而言，我们希望了解实践者可能如何使用 Atelier、会采用哪些不同策略、它如何融入现有的创作过程，以及其表现力（包括使用门槛和上限 [68]）等。

8.1 参与者

我们招募了 5 位创意专业人士（2 位女性、2 位男性、1 位性别不符合二元分类者），年龄为 29 至 31 岁。我们采用目的性采样招募方式 [8]，并通过电子邮件联系了多位直接联系人。参与者的专业领域涵盖媒体与娱乐行业的多个领域。每位参与者都曾在其创作实践中使用生成式人工智能。作为参与研究的答谢，参与者获得了等值 400 美元的礼品卡。该流程在研究开始前获得了机构伦理审查委员会的批准。由于我们的目标并非追求普遍适用性，而是呈现多样化的使用策略，因此每位参与者 4 小时深入而丰富的互动产生了大量信息丰富的数据。参与者

背景和人工智能经验总结见表2

8.2 流程

参与者收到了参加现场研究的电子邮件邀请，并获提供 Atelier 的视频图。参与者被要求提前想好一个项目，但无需进行大量准备。研究包含三项核心活动：前期研究访谈、引导式演示与自由探索，以及研究后调查。

[10]https://github.com/logtd [11]
https://github.com/logtd/ComfyUI-Fluxtapoz [12]
https://tinyurl.com/AbramsWanFlowEditWorkflow [13]
https://github.com/logtd/ComfyUI-HunyuanLoom

表2：研究参与者背景及其人工智能经验概要。

ID	背景	人工智能工具使用经验
P1	技术戏剧、场景设计、服装设计	当前：*Claude* 过去：*ChatGPT*、*Midjourney*
P2	电影摄影、灯光设计、摄影机操作、三维，虚拟制作	当前：*ChatGPT* 过去：*Runway*、*Midjourney*、*Google Veo3*
P3	数字产品设计，计算机科学、音乐制作	当前：*ChatGPT*、*Claude*、*Gemini*、*Nano Banana* 过去：*Midjourney*、*ComfyUI*、*Suno*
P4	动画、二维/三维、定格动画、电影，道具制作。	当前：*ComfyUI*。 过去：*ChatGPT*、*Claude*、*Leonardo*，*Midjourney*、*Suno*、*Udio*
P5	电影摄影、动画，视觉叙事、灯光设计。	当前：*ChatGPT* 过去：*Midjourney*

研究前：半结构化访谈（***15-20*** 分钟）。在参与者同意参加研究并接受屏幕录制后，我们进行了前期半结构化访谈，以了解参与者的背景、如何使用灵感素材、他们的人工智能使用经验和所有权感受等。前期研究访谈耗时 15-20 分钟。

˜引导演示与自由探索（***3.5*** 小时）。实验人员提供了一台计算机、显示器和外围设备。参与者通过网页浏览器使用 Atelier，并可全面使用实验人员的设备，包括网页浏览器和离线创意应用程序，以便在需要时使用熟悉的工具。

实验人员随时提供适度指导，以减少操作阻碍。参与者主导探索过程——只有在他们停止测试某项功能，或开始将已使用过的功能组合起来后，才会介绍新功能。实验人员会提出问题，或通过无提示的边说边想方式收集参与者的印象。

参与者首先了解了画布导航，随后观看了快捷操作的演示。介绍不同画架时，参与者会导入素材；鉴于“绘制”较为简单，因此从它开始。后续顺序则根据个人项目而定。最后，向参与者介绍了溯源、组织和意义建构工具。剩余时间里，参与者可以继续自由使用系统。我们鼓励参与者休息，大多数人全程休息了 2 至 3 次，其中包括中途一次较长的休息，时长为 15 至 20 分钟。

研究后调查（***10*** 分钟）。最后 10 分钟用于进行研究后调查，参与者在其中填写了人口统计信息，并回答了有关其创作实践的问题，还回答了一组配对李克特量表问题，用于对比他们当前使用人工智能工具与 Atelier 的情况。示例问题如下：“我对自己使用 *[Atelier | 我当前的人工智能工具]*”创作的内容有一种所有权感。完整细节请参见附录A。

表 3：关于当前人工智能工具与 Atelier 的配对问卷调查结果。第三列显示了每位参与者回答之间的差值。

问题	我当前的人工智能工具					Atelier					个人变化				
	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*	*Strongly Disagree*	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*	P1	P2	P3	P4	P5
所有权	0	0	2	3	0	0	0	1	2	2	1	1	0	1	0
能动性	0	2	2	1	0	0	0	0	4	1	1	2	1	2	1
早期思考	0	0	0	3	2	0	0	0	3	2	0	0	0	0	0
后期阶段思考	0	1	1	3	0	0	0	0	4	1	2	0	0	1	0
灵感	1	0	0	3	1	0	0	1	1	3	1	2	1	0	0
最终产品	1	0	3	0	1	0	0	1	3	1	2	0	1	3	-1
可塑性	0	4	0	1	0	0	0	0	1	4	1	3	2	3	3
整合至处理过程	0	1	2	1	1	0	0	0	3	2	0	2	0	1	2
整合至工具	0	2	1	1	1	0	0	2	1	2	0	3	0	0	1
端到端支持	1	2	1	0	1	0	0	2	1	2	0	3	2	4	2
反映意图	0	2	0	2	1	0	0	0	4	1	-1	2	1	2	0
愉悦感	0	0	3	1	1	0	0	0	1	4	0	2	2	1	1
Fun	0	2	2	1	0	0	0	0	1	4	1	0	2	1	2
专业感受	0	1	3	1	0	0	0	1	3	1	1	2	1	1	0
修复输出结果	0	3	1	1	0	0	0	2	2	1	1	1	1	2	1
富有表现力的-*ness*	0	1	3	1	0	0	0	1	3	1	2	2	1	0	0
被视为 *Tool*	0	0	0	3	2	0	0	0	1	4	0	1	1	1	-1
被视为协作者	3	1	0	1	0	3	1	0	0	1	0	0	1	0	0
赋能-*ment*	0	0	4	0	1	0	0	2	2	1	0	0	1	1	0
创造力	0	0	1	4	0	0	0	0	4	1	0	1	0	0	1

8.3 数据分析

经参与者同意，所有场次均进行了屏幕和音频录制，并在转录时删除了填充词。第一作者与第二作者共同分析了定性数据。第一作者对转录稿进行了开放式编码。负责开展实验的第二作者则审阅了编码和数据。通过反复讨论和主题聚类 [15]，作者们确定了一组主题。研究后调查中的李克特量表数据用于补充研究发现，并分析个人参与者情感倾向的变化。

9 结果

参与者都使用不同的功能和画架创建了画布（图 15）。P01 尝试将“一块由布料碎片制成的砖块”具象化，用于场景设计。P02 创建了一组电影场景：一名男子在赌场外弹奏吉他，并反复调整灯光、摄像机角度和摄影机运动。P03 在经过开放式探索后，重新拾起一个旧的应用程序构想：将股票图表呈现为山脉，并使用 Atelier 为这种视觉转换制作原型。P04 重现了一个旧动画项目，其中数字变形成各种形状，作为电影倒计时。P05 为一个电影概念探索镜头，随后为一个广告项目制作产品镜头，并反复调整灯光、框定和运动。

9.1 Atelier 是如何使用的?

最喜欢的画架。 参与者都介绍了自己最喜欢的画架。P1 和 P4 最喜欢绘画。P1 喜欢负向提示词和结构参考，因为它们让自己能够更好地掌控乐曲创作；而 P4 则指出，这提供了最大的灵活性和探索选择。P3 和 P5 最喜欢描摹，因为他们觉得可以转换起始图像：“我想这是因为我可以通过反复描摹的步数来进行控制。它是一种调整强度的机制，而且调整方式更可预测。我觉得通过这些反复描摹的步数，我可以成为 *[这个画架]*的高级用户……这让我得到更可预测的结果”(P3)。P2 则对 Animate 表现出最强的偏好：“能够在视频中看到它、看到它运动，让我有很多想法，也能进行创造性的反思……与一张图像相比，这带来了许多可能性，*[相较于]*一张图像 *[而言，是]*我的创作过程中的额外一步……它改变了一切”。

最大挑战。 P1 和 P2 表示，Trace 是最难上手的画架。尽管取得了一些成功，P2 仍不清楚它的工作方式：“我们在这里到底要做什么？因为我们有原始图像，也有最终图像……”。所有参与者起初都对必须描述原始图像、再描述最终图像感到困惑。对所有参与者来说，提示词撰写都是一大挑战。P1 和 P5 转而使用 Claude：P1 会请 Claude 帮忙改进提示词的措辞，而 P5 则用它来寻找恰当的词语来描述某件事，或将母语翻译成英语。P2 指出，使用生成式人工智能时，想要获得变体，往往唯一的办法就是“写得更好，给出更好的提示词……能做的事情非常有限”。

明确设计意图。 Atelier 为参与者提供了更多途径，让他们实现自己想要的视觉效果。其中一部分来自实验，P2 表示“喜[欢] 程序能做到如此精准 [地] 控制……绘制、绘画、修改这些选项给了你更多实验的空间”。P1 这样描述：“它让你能够掌控……更容易知道该如何想象……这更接近我想要的效果”。生成不同的图像往往有助于更清楚地把握设计意图：“看看其他探索成果也不错……现在我”

原型采样：通过画布驱动的视觉人工智能生成，实现采样与原型设计的自由形式融合



图 15：参与者使用 Atelier 创建的画布。P1 展示了一个用布料碎片制作砖块的概念；P2 展示了一个场景的镜头：一位老人坐在卡车里弹吉他；P03 创作了各种概念，包括一款将股票交易可视化为景观的应用程序；P4 重现了一项旧的动画练习，其中由不同素材制成的数字不断变形；P5 探索了电影的不同镜头构想，随后又探索了商业广告的镜头构想。

比起 *[先前的想法]*，我更喜欢这个想法”（*P5*）。不理想的结果也有助于进一步明确设计意图，因为参与者会更清楚自己不想要什么，从而缩小原本广阔的可能性空间：“对我来说，这会让我更清楚……我不想要拼接砖，我想要废弃砖。所以我会调整搭建方式，这样就能帮助我思考，更清楚自己不想要什么。这样就不容易迷失方向，也更快。”（P1）。

支持创作过程的不同阶段。研究后调查显示，相较于他们目前使用的人工智能工具，参与者认为自己可以将 Atelier 用于获取灵感、早期思考、后期思考以及制作最终产品（表 3）。不过，具体看法因从业者的主要媒介和目标而异，且参与者主要将其视为概念性工具。尽管 P1 过去曾在最终项目中使用人工智能图像，但她认为生成式人工智能仍属于概念范畴，因为它依然是数字形式。她认为自己的布景设计是一项需要亲自动手的实体活动：“我会用图像进行视觉沟通。提取概念和想法，融入我的兴趣，并进行交流，然后最终做出一个东西，让人们看了会说：‘对，这就是我们要搭建的东西。’”。P2 指出，Atelier 能帮助他们制作更逼真的提案：“我可以想象 *[自己]* 在每个阶段都用 *[它]* 进行预可视化……思考各种事情、进行头脑风暴……为提案材料包或资助申请准备参考视频或参考图像。我能想象自己用它向导演提案，比如说：‘这是我的构想，这是我对你这个场景的设想。这是我想要的动作和基调。’我可以把这些都说得非常具体，或许就能打动那位导演，让他聘用我，因为如此具体的呈现很难被超越；展示独特的东西，总比展示一部现成电影的图像更有说服力。”。P2 还表示，他们可能会用 Atelier 制作完整的视频，纯粹为了乐趣。P3 认为概念生成很有价值，同时也把自己的画架视为工作流，希望能获得 API 访问权限：“我很希望能创建这样一套工作流，最终将它导出，或将其作为 *API* 提供给我自己的应用程序使用。”。P4 和 P5 认为 Atelier 可用于制作分镜板或动态分镜；P5 尤其指出，它对“商业广告之类的内容……很有用，他们确实想看看最终效果。如果你能拿出这样的东西，在拍摄期间就能帮他们解决很多问题。”。

9.2 参与者如何组织空间?

画布布局。参与者在工作过程中各自形成了在画布上移动的方式。每位参与者在不同程度上都采用了从左向右的工作方式。P2 和 P3 以简单直观的方式按照时间顺序从左向右组织画布，P2 解释道：“我试着按照时间顺序来思考……按照西方世界从左向右的阅读方式移动，并尽量保持一致。”P2 指出，如果他们从事的是一个较长期的项目，就会为每个画架分别创建一个角落。P3 按时间顺序从左向右移动，并将变体垂直排列。P4 将输出结果放在右侧，将参考资料放在左侧，并紧邻画架；P4 指出：“如果我有更多时间来整理，就会把它们分得更开一些，形成更稀疏的簇，这样周围就有更多空间供我操作。”P1 则从中心开始向外移动，将偏好的作品放在最右侧，并放大这些图像以突出显示，而将舍弃的材料放在最左侧。P5 从左向右移动，探索新主题时，则会在当前工作的下方另起一行。大多数参与者（P2、P3、P4、P5）都将输出结果聚集在生成它们的画架周围；P4 提到，即使缩小画布：“我也能从视觉上分辨出是什么……这边有一小簇，那边也有一小簇……我能看出它们代表的是不同的内容。”参与者采用了不同的方法来管理大量生成结果。大多数人（P1-P4）保留了大部分生成结果，只删除与目标完全不符的输出结果，而 P5 则更倾向于大量删除画架和生成结果。几位参与者（P2、P3、P4）使用分组和打包工具来安排作品在画布上的空间布局，而 P1 和 P5 则更能接受元素彼此重叠。

作为空间锚点的画架。 画架在无限画布中充当锚点。P1 表示，“基于画布的创作空间如此广阔，很容易迷失其中。我认为画架能提供一定的结构，又不会限制你……画架的作用是……帮助你整理事物……这样你就能顺着正在发生的创作流程继续工作，同时也能进行整理，这样等你回来时，就不会被大量内容压得不知所措。”他们解释说，这样一来，隔天回来继续创作时，就能“有一个全新的工作空间，同时仍然能够建立联系并回顾之前的创作”。参与者很看重无限画布的空间，P2 这样描述：“任何画布式的创作几乎都像是无限的。它是一个多元宇宙。你可以创造一个充满想法和选项的多元宇宙……我喜欢这种没有尽头的感觉。”他们指出，即使组织方式更加随意，“仍然能在空间中看清处理过程。”参与者有时会把画架当作个性化工作空间，并根据自己的工作流进行配置。P3 这样描述画架：“我把这里看作我的小工厂区域……我的游乐场……这是我的空间……设置方式和参数有很多种。我一点也不想忘记自己是怎么设置的。我喜欢保留这种持久性。这几乎就像是我亲手打造了这个工具，因为我在里面添加了所有这些参数。”P2 想象着随着使用时间的延长，可以通过画架来组织空间：“如果我要更认真地使用它，我可能会把工具再分开一些。这是我的绘画区。这是我的修改区。”.

溯源工具。其他溯源工具，如时间线、历史记录、热力图、沿革面板和重新创建画架选项，可用于浏览作品，并理解创作过程。P2 称历史记录功能“改变了我的生活……为什么不是所有东西都有这个功能，比如说历史记录，能直观看到你做了什么的视觉历史记录”。P3 解释说“我喜欢保留自己的创作历史。这既让我觉得自己在不断进步，也能让我看到自己一路走来的历程；当我可能已经忘记旧想法时，还能回头看看”。“重新创建画架”功能可让用户回到特定的参数配置。P1 这样形容它：“就像有人把你的图像和笔记本一起送给你，笔记本还展示了你是如何创作出这幅图像的……你可以轻松回到绘画草图”。P4 认为溯源系统解决了参考资料管理方面的难题，并指出它“真的很有用。这样很好，因为它能保护你的处理过程。它简单易用……参考资料管理……你 *[可能]* 得用外部软件……但随后

你可以在这里的软件中管理这些资料，这很方便。而且你能确保不会丢失它们，因为它们保存在你的项目里”。对参与者而言，保留处理过程十分重要，既有助于管理参考资料，也有助于建立创意联系和进行反思。P3 这样描述：“我喜欢保留自己的创作历史。这既让我觉得自己在不断进步，也能让我看到自己一路走来的历程；当我可能已经忘记旧想法时，还能回头看看。我觉得这非常重要”。

9.3 Atelier 与其他人工智能工具有何不同?

使用 Atelier 时，参与者提到了他们此前使用人工智能图像或视频生成平台的经历，并将其与 Atelier 作比较。这些平台包括基于提示词的平台，如 ChatGPT 和 Midjourney，以及基于节点的平台，如 ComfyUI。

与基于提示词的界面对比。与基于提示词的界面相比，Atelier 的多模态特性通过画架的设计（例如，可用于填入图像参考的插槽）让人们得以了解模型的能力（表 3）。对参与者而言，基于语言的方法并未提供明确的方式来使用生成式人工智能进行创作：“*[人们]* 只能通过文字与计算机沟通，而文字无法充分表达 *[人们]* 脑海中的想法。工具本应让艺术家拥有一些自主权，但它却让 *[人们]* 不得不经过 *[语言]* 这一层……如果没有所有可用的工具，没有锤子，没有凿子，人们又怎么能把脑海中的想法化为实体呢……”(P4)。

与提示词驱动的交互相比，Atelier 转而让用户直接处理素材，从而将参与者与人工智能生成的关系从与模型对话转变为操控内容：“当我使用 *Midjourney* 或 *Runway* 时，我感觉自己是在与人工智能交互……我向它提出要求，然后接收到那张图像。在这里，我感觉自己拥有创作控制权。这是我的创作工作空间，而不是人工智能的工作空间”（P2）。P5 指出，与 Midjourney 相比，Atelier 提供了可供起步的选择：“像绘制、绘画、修改这样的选项……让人感觉更亲切。你有更多选择可以开始……当我尝试创作某样东西时，这对我来说是最令人望而生畏的部分”。这种转变赋予了参与者能动性；当生成结果失败时，他们不必盲目地重写提示，而是拥有“易于使用的调节控件”进行调整。

P2 认为，使用 Atelier 会让人将注意力集中到处理过程本身：“即使是与系统交互，这也是一个创作过程。而当我使用其他视频生成工具时，却感觉这并不具有创意”。P2 还指出，从鲜艳的色彩使系统显得更易于接近，到 Atelier 中命名约定的选择，许多细节都激发了创造性思维：“我觉得具有创意的是这里的工具。甚至包括你们为工具命名的方式，比如绘图、绘画、素描、拼贴……这些名称提到了各种艺术形式，促使我以一种更具创造性的方式从艺术角度思考，而不是撰写提示词。”

与基于节点的系统的比较。P3 和 P4 此前都有使用 ComfyUI 的经验。他们表示，Atelier 更易于理解，也更容易学习和使用。尽管基于节点的系统能够提供更精细的控制，但其学习曲线却构成了创建过程中的一道令人生畏的障碍，许多参与者认为，这款软件只是达成目的的手段，而如果“学习这款软件的难度和我将要创作的艺术作品一样大，那我就不想做了”（*P1*）。*P3* 回忆道，像 *ComfyUI* 这样的工具：“是非常令人望而生畏的工具……不同的输入和输出、细小的连接器，还要知道它们该连接到哪里以及如何运作……内容实在太多了……即使我来自软件开发领域，也使用过 Blender，我仍然没有完全适应节点……我试过 ComfyUI……但失败了，因为它的学习曲线太难了”。相比之下，*Atelier*：“至少让我能够以一种更适合新手的方式，尝试使用这些工作流”。*P4* 指出，*Atelier* “感觉更加私人化……也更具协作性……它更具视觉性，也更容易解释，即使是可能没有使用工具经验的人，也更容易理解。如果我在他们面前进行操作，那么他们通过视觉就已经能够理解一些东西；而使用 ComfyUI 时，你会看到大量文本和节点……最后才在末端得到一张图像。使用 Atelier时，里面只有图像。相比于一个完整的系统生成一张图像，更容易展示这个 画架正在生成这么多图像”。这使参与者能够理解这项工作，也进一步便于潜在的协作分享。*P1* 经常与计算机素养较低的合作者一起工作，他指出：“这里有一些东西让人们很容易使用它，你不需要让他们投入大量精力去理解它有什么用。你在操作的过程中就能得到结果”。

参与者能够更快获得理想结果，因此在创意探索过程中可以保持上下文连贯。P4指出，“要是在 *Comfy* 里做出这样的结果，*[他们]* 得花上久得多的时间”。所有生成结果都集中在一处，Atelier 因而提供了连续的视觉历史，P4称之为想法发展过程的“鲜活记忆”。此外，由于各次迭代在时间和空间上都彼此接近，他们可以通过缩小画布视图“在脑海中追溯整个过程”，并“把一切尽收眼底”。P4指出，如果使用 ComfyUI，他们就得把生成工作分成两天完成。这种连贯性促成了某些比较和判断方式，而这些方式在基于聊天的提示中很难实现。P3形容基于聊天的提示“僵化”且“实用主义”，倾向于产出单一结果，而非探索其他可能性：“这个工具基本上让我能够探索各种创意可能性的搜索空间……而我用过的其他人工智能工具并没有真正采用这种画布式方法……感觉更注重完成特定任务”。

9.4 Atelier 如何支持工艺与所有权?

前期研究访谈凸显了参与者对于工艺和所有权的一些基本看法。然而，使用 Atelier 也展现了反映人们工艺与所有权的新方式。具体而言，我们确定了意图、可塑性、集体输出，以及参与者将人工智能理解为一种媒介等主题。研究后调查结果（表 3）也凸显了与这些变化相关的要素，即以下类别：所有权、能动性、可塑性、反映意图、修复输出的能力、表现力、赋能和创造力。所有这些类别中的情感倾向大多 favor Atelier。

意图。 参与者感受到所有权的一种方式是通过意图。在预研中，P02 描述了他们的生成结果为何往往独一无二，因为他们“对自己的想法有非常明确的主见 *[他们]* 心里想的，并且

他们会尝试许多不同的提示词”；P03 则描述了他们过去创作人工智能漫画时，觉得自己是在“创造一种艺术形式”，因为他们“非常有意图地进行创作”。使用 Atelier 也展现了意图如何体现出来。P01 描述道：“这感觉像是我脑海中早已存在的东西，而人工智能只是……图像生成不过是一种工具……用来确保人们理解并想象出与我所想象的相同的事物”。P05 虽然不确定自己是否有所有权感，但觉得 Atelier 让他们能够通过实验来完善自己的设计意图：“这个东西已经包含了一些我有的想法，或者是一些我希望实现的想法的组合”。

可塑性。 在研究后调查中，参与者的可塑性提升最为显著。在前期研究访谈中，P01 描述了他们为何需要 “对*[*人工智能*]* 有很多输入，才能觉得它是 *[*自己的*]*”。P02 描述了提示词如何通过多次迭代形成：“我把它们打磨得像钻石一样……钻石一直都在那里，但还需要把它雕琢成可爱的形状。否则它就只是一块石头。” 这些观点凸显了每次生成都会带来参数调整，从而实现快速迭代。使用 Atelier 后，P02 对所有权的理解更加细腻：“在这里，我觉得自己是个创作者。我在创作、组合、素描、绘图……这关乎处理过程……提示词撰写感觉更偏技术性……你试图猜测公式、猜测人工智能的想法，并尝试打破那些公式”。P03 指出，画架体现了一组能够实现其设计意图的参数：“*[*在*]* 让 *[*这个概念*]* 准确无误这件事上……要实现这种技术可能性，整个流程中必须融入工艺。” P04 也表达了类似的情感倾向，因为在 ChatGPT 这类工具中，“你必须写一整段话来说明自己真正想要什么。但结果未必如你所愿”。 相比之下，“使用 *[Atelier]* 时，你可以将所有这些特征单独分离出来，这样即使修改其他内容，也能确保这些方面保持不变。”。

集体输出。 可塑性方面关注生成结果如何带来新的设置和迭代。也可以从生成输出结果的角度来看待这一处理过程。随着时间推移，输出结果集合能够反映设置的变化，以及生成过程朝特定方向发展的情况。P02 已经描述过如何通过迭代过程找到合适的提示词。不过，由于 Atelier 将所有输出结果都置于画布之中，P03 表示，看到不同的生成结果并排呈现，并反思它们的空间排列和涌现结果，让自己感受到工艺和所有权——他们的所有权感和工艺感源自见证这一历程：“尽管源材料来自一部非常著名的电影，但随着我们不断推进……挑选不同的变体和世界……我开始能够挑选并几乎对自己希望风格朝哪个方向发展形成看法……我认为，部分原因是我能更轻松地掌控并引导事物。但同时，看到这些变体以这种方式全部呈现出来，能够接受或拒绝它们，并最终从大量探索中选定一个创作方向……这也是创作过程的一部分，让我拥有了所有权”。P04 将使用 Atelier 的感受与故事板制作联系起来：“感觉更像是在制作分镜脚本和绘制缩略图”。

人工智能作为一种媒介。 谈到为何不觉得自己拥有人工智能生成结果的所有权时，P05 解释说，这是因为他们不了解生成式人工智能的工作原理，具体而言，他们表示：“有时这很有挑战性……我很容易就会感到绝望。比如，那不是我的想法，我就是不想付出 *[*这*]*份力”。P05 使用 Atelier 似乎带来了控制感和能动性，尤其是在使用 Trace 画架时；它使他们能够生成与参考素材相似、但更符合自己创作意图的镜头。尽管所有权感没有改变，P05 在实验中变得更加自信。这一点也反映在他们的研究后调查回答中：能动性和修复输出的能力均有所提升。P04 也谈到了理解人工智能作为一种媒介的重要性：“这就像是接受它本来的样子，并利用它能提供的东西，以某种方式塑造你想象中想要实现的内容……这与艺术家亲手创作时所做的事情相似。我的意思是，说到底，你想象的东西未必就是你最终创作出来的。从开始到完成某件事的过程中，有很多因素会发生变化”。虽然我们没有明确询问他们对人工智能作为一种媒介的理解，但研究后调查中的情感倾向回答（表 3）显示，若干可能与此相关的指标出现了情感倾向变化：可塑性、能动性、修正不理想输出的能力以及赋权感。

10 讨论

原型采样这一概念提供了一个观察创作过程的视角，以此为创作系统的设计提供参考。Atelier 展示了实现这一目标的一种方式，可用于验证这一概念 [98]。作为研究处理过程的一部分，使用 Atelier 引发了批判性反思：从将其作为原型采样的手段，到生成式人工智能具有非确定性所带来的挑战与机遇，再到其概念层面的影响、技术与概念上的局限，以及未来研究的潜力。

10.1 Atelier 作为原型采样的手段

在本文前文中，我们提出原型采样是一种平衡采样的策展性质及其提取顿悟的能力，与能够将想法具象化的原型制作行为的方法。这推动了 Atelier 中许多设计决策的形成。

10.1.1 融合思考与制作 . 媒介优先方法映射了现实世界 [25], ——在现实世界中，参考资料和灵感会作为整体被处理，因为创作者可能尚不了解某些事物为何有用 [44, 56, 57]。画架与快捷操作共同充当实现这一理念中“制作”部分的关键途径。许多参与者表示，该界面非常直观，使用起来仿佛这片空间可以被称为自己的空间。我们还发现，Atelier 直接支持原型采样中的程序性操作，尤其是衍生和组合方法。类比、隐喻和第一性原理等方面，是复杂得多的处理过程，需要实践者将其作为创造性问题解决的一部分自行设计。这可能意味着，创作者若直接与他人交流，或更多地了解周围的现实世界，便能从中获得更大的价值。重要的是，每位参与者都强调，在 Atelier 中工作让人感到轻松自在，或充满乐趣。这种玩乐感与诸如以下行为同时出现：

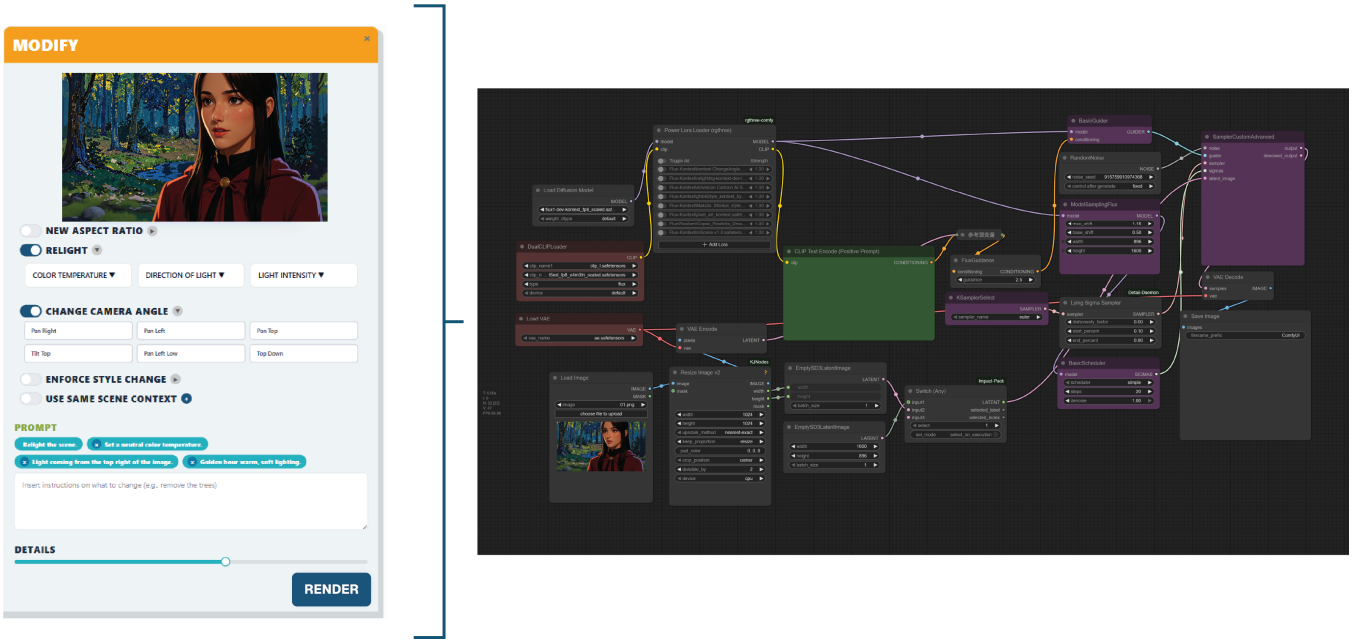

图16：修改画架的示例，以及它如何映射到对应的 ComfyUI 工作流。

自由地拓展、重新组合和测试想法，这表明 Atelier 支持一种开放且低风险的探索。

10.1.2 封装。使用当前的方法时，创作者必须在两种方案中做出选择。一种是使用专门的服务，例如 Midjourney [63] 或 Runway [79],，并采用其提供的界面和选项。这类服务通常较为易用，但各服务的功能范围不尽相同，因此往往需要在不同服务之间迁移工作。另一种方案是使用能够直接访问模型的系统，例如 ComfyUI，或基于 Gradio 的界面，例如 Automatic1111[14]。然而，正如一些参与者对 ComfyUI 或基于节点的画布感到沮丧所表明的那样，这些工具需要花费时间学习和训练。由此带来的挑战是，如何发挥基于节点的界面的潜力，并让它们更易于使用。这正是交互设计大有可为之处。基于此，我们尝试将绘制与上色映射到增强型文生图和图像生成图像工作流，将描摹映射到 FlowEdit 的使用，并将修改映射到 FLUX Kontext 等编辑模型。模型、检查点、LoRA 等的选择则放在后续处理，并预先打包。我们也考虑了预处理图像。早期构思阶段，我们打算将 ControlNet 所需的深度、姿势和线稿图像作为媒体放在画布中。但后来我们意识到，这些图像只会用在支持结构参考的画架中。我们还意识到，将这些图像放在画布上会使预处理素材泛滥，并要求用户理解它们的含义及用法。将它们作为画架结构区中的建议，可以降低这种复杂度，不过用户仍需具备相应的知识，才能了解这些图像是什么以及它们的作用。

画架如何映射到工作流的一个例子是修改画架（图16）。除了使用 FLUX Kontext-dev 和用于重新布光、摄像机变化及风格调整的 LoRA 外，我们还从不同的 LoRA 中提取了可能的关键词，并将它们添加为提示词胶囊。界面上的切换开关对应不同的 LoRA，或用于切换宽高比。

以下是构建这些抽象时需要考虑的诸多因素中的一些示例：我们希望降低使用门槛、提高能力上限 [68], 提供一组构建模块 [49], 并显示恰当的参数，使表现力尽可能契合任务需求 [71, 77]。构建这些抽象时，我们参考了不同在线社区的体验，包括 Reddit[15] 和 Banodoco[16],等社区的日常互动，以及已有记录的第一人称记述 [48]，还有我们自己学习和使用这些技术的经历。一位参与者（P3）表示，在研究期间使用 Atelier 后，他们觉得自己对 ComfyUI 有了更多了解。这表明，可以逐步引入更高级的概念，或解锁系统中更复杂的功能；此外，还可以进一步探索如何将创意系统作为学习人工智能概念的实践环境。

10.1.3 组织与溯源。从参与者的使用情况来看，内容往往聚集在画架周围。从这个意义上说，画架可以充当内容的锚点。与此同时，画布隐喻使内容散落各处、缺乏结构，甚至无法选择堆叠。结合有关知识工作、存储与搜索以及材料检索的讨论，[10, 11, 36, 84],我们发现，能够探索内容并创建更持久的收藏集十分重要。我们认为，与这些外化内容互动可以引发反思，[24, 83], 因为现在所有

[14]https://github.com/AUTOMATIC1111/stable-diffusion-webui

[15]https://www.reddit.com/r/StableDiffusion/ [16]https://banodoco.ai/

素材都能成为重新诠释的触发因素 [26]。这表明，以媒体为先的自由形式创建是可行的，但无限画布需要搭配同样强大的意义建构功能，参与者才不至于迷失。创意工作往往需要大量协作，参与者心里都想着要把作品展示给其他人，无论是客户还是合作者。这个空间非常适合协作或分享处理过程，因为参与者觉得，要理解自己的创作历程并不困难。

10.2 开放性与活动的一体两面

生成式人工智能面临的一项挑战是缺乏确定性。另一方面，这也可能是一项强大的优势，因为解决一个问题的方式有很多，因而存在许多阻力最小的路径。例如，使用 AI 图像生成时，角色的一致性仍然是个难题，通常需要训练自定义 LoRA。使用**修改** 画架，可以为角色生成一些不同的摄像机角度，并将其用作视频的起始帧。另一种方法是使用**Animate**生成视频，以起始帧或结束帧提示角色动作或镜头运动，然后从这些帧中提取姿势。解决问题的这些多种路径，加上移除背景并在拼贴中重新组合的能力，意味着我们可以实现角色的多种姿势和位置，并找到重复利用它们的新方法。在用户研究中，我们还发现了 Collage 画架的新颖用法。P1 创建了多个砖块副本，以思考如何制作门口；P2 则使用 Collage 创建合成镜头，之后可将其导入 Trace。当其他画架未能准确遵循用户意图（例如生成了错误的卡车）时，Collage 尤其有用。

这种开放性也可能带来问题，尤其是在尚未完全理解模型能力边界的情况下。“去噪”或“调度器”等术语对新手来说可能较难掌握，而且这些术语并不属于大多数创意领域。虽然今后还会出现许多需要学习的新术语，但究竟何时最适合使用 FlowEdit，而不是 Flux Kontext，该如何判断？例如，两者都可以改变图像风格，但 FlowEdit 能提供更多的创作自由。现实情况是，即使拥有一定经验，这也不是一个能够直接回答的问题；它需要反复失败，并从失败中学习。然而，即便已经了解如何使用这些技术，这一过程仍然为意外发现留下了空间。

10.3 概念性影响

虽然原型采样的主要关注点在于程序化行动和轨迹，但由这一活动产生的反思会直接影响处理过程本身。

转化。 Protosampling 的第一个概念性启示，指的是反思性实践的转化性质 [82]。评估截至某个时间点所收集或创作的内容，必然会重新界定创意问题。Ohlsson [70] 认为，这可以通过细化（通过添加信息改变表征）、重新编码（舍弃对问题的部分原始解释并加以修订），或约束放宽（改变对目标的不正确表征）来实现。

涌现。 第二个概念性启示是创作过程的涌现性质，Cross 将其描述为融入未来概念的未被识别属性 [21]。这凸显了意图驱动（与原计划保持一致）与涌现驱动（对外部影响、学习和灵感持开放态度）之间的相互作用 [32]。想法并非孤立存在[22]，，而是都属于一个相互关联、有方向且合乎理性的处理过程 [96, 97]，，在这个过程中，外化不断带来新的后续想法 [81]。

偏见与固着。 一个有趣的问题是，生成式人工智能的使用是否会加剧偏见或设计固着。所有模型都有不同的偏见与优化方向。在 Atelier 中，我们支持多种模型和活动，以便从业者尝试解决问题的不同方式。通过使用参考图像，模型会更多地受到参考内容的引导，而不是其基础训练内容的影响；使用个人风格时也是如此。根据我们使用本地人工智能生成超过 100,000 张图像的经验，我们还发现，不同模型和技术各自具有不同的“怪癖”。例如，Stable Diffusion 可能在渲染手指时存在问题，但它无法完全遵循提示词的能力，也使多种风格的混合成为可能。使用于描摹中的 FlowEdit，会增添从起始提示词到结束提示词之间应产生多大程度变形的细微控制。Wan 等模型在提示词遵循度和一致性方面表现更好，但当提示词含义模糊时，往往可能意外侵犯知识产权。

设计固着可能出现的一种情况是，某人投入更多精力进行迭代式优化，而非探索。在研究过程中，P1 经常提到沿着某条路径不断精修，并因此感到分心：“它会让人分心，因为它能生成这么多东西，以至于你会想，哦，这是我想要的吗？还是它只是让人兴奋？”。进行更多体力活动时，疲劳显然是一种限制；但使用人工智能生成结果时，这种情况可能长时间都不会被察觉。

10.4 本地模型及其影响

Atelier 的一项关键设计选择是依赖本地开源模型，而非云服务。虽然云服务有诸多好处，例如无需依赖个人硬件，但原型采样之所以能够实现，很大程度上得益于开源技术所提供的可控性与细腻度。开源模型允许用户访问所有必要的参数来控制模型，因此可以混合搭配不同的平台和扩展功能，例如添加细节、使用负面提示，或触发多种风格。举例来说，Paint 画架可以先用基础模型生成图像，再将其调整为强度为 30% 的写实主义风格、强度为 80% 的梦幻风格，以及强度为 60% 的复古动漫风格[17]。这种灵活性使用户能够组合不同模型来进行图像参考——使用 Flux Redux[30] 时，模型会根据强度设置自动从每项参考中提取元素；而使用 USO[99]时，则会明确分离出风格。在本地运行人工智能还有其他理由，包括提高能源效率、保护隐私、避免内容被困在封闭的生态系统中，以及能够自主选择要使用的模型。在

[17]强度无需加总至 100%，它们只表示单独微调有多少被嵌入主检查点中

在这一版本的系统中，我们预先选定了风格模型，但也可以设想让用户导入自己针对特定角色或风格微调的模型。由此，创作者将拥有广阔的可能性，真正发挥生成式人工智能在创作中的力量。

模型的灵活性也涉及伦理问题。AI 模型使用公开数据进行训练，既未提供选择加入或退出的选项，也对创意社区产生了显著的经济影响。其伦理影响错综复杂，目前仍在积极讨论中，因为在本地工作时，没有任何机制能够阻止不法分子创建恶意或不当内容，或使用特定风格的作品进行训练。

10.5 局限性

Atelier 当前的设计与实现存在三项主要局限。其中一项是技术层面的。尽管 ComfyUI 是功能强大的原型制作工具，但它并非为稳健性而设计。路径和模型通常被硬编码，存在内存泄漏；而且，由于系统很大程度上依赖社区贡献，更新可能导致系统无法正常运行。此外，我们是在消费级硬件上本地运行该系统。虽然运行速度可与商业级硬件相媲美，但无法将所有模型都保留在内存中。这意味着模型需要不断换入换出，既耗时，又带来大量冗余。这些局限并非 Atelier 交互模型本身固有的；随着模型服务基础设施日趋成熟，同样的工作流可以实现得更加稳健、易用。

第二个局限在于，Atelier 假设大部分工作都在系统内完成。虽然用户可以导入和导出许多媒体，但实际上，Atelier 也可能导致信息进一步分散在不同应用程序中。如果能将这一隐喻应用到整个操作系统，类似于其他系统中的做法 [13]，那么 Atelier 这样的系统或许会更有效。不过，研究参与者（P2、P4）表示，他们希望留在该工具中，即使是进行图像编辑这类可能更适合使用其他工具的操作，也希望避免切换应用程序。

Atelier 的最后一项局限在于概念层面。虽然我们主张让参考资料与生成的素材共存于同一个虚拟空间中，但现实是，要成功运用生成式人工智能，需要大量生成结果 [48]。这意味着生成内容中有很大一部分实际上都是废料。大量生成结果也会带来另一个问题：人们很容易忽视自己想要创作的内容，无论这是因为不得不迁就模型的局限，还是因为大量生成的内容让思维不堪重负，进而促使偶然的创意浮现。模型的数据和训练也带有一系列既定偏见，这些偏见也可能因此延续。像 Atelier 那样使用多模态输入有助于缓解这一问题，但解决方案仍有被这些偏见左右的风险，可能导致设计固着，或意外侵犯他人的知识产权。这些都是将人工智能作为一种媒介时必须面对的现实，要想更好地理解这些问题，还需要更多时间和研究。

10.6 未来工作

Atelier 的工作带来了一系列值得探索的研究机遇。首先是对视频的探索，因为迄今为止我们只探索了图像转视频。生成式视频还有许多其他创作方式，且具有更强的可控性[48], ；这些模态需要恰当的抽象，并且需要更深入地理解其中的交互机制。用于视频的画布式方法引发了对抽象的思考，也促使我们反思物理空间与动画运动基于时间的特性之间可能存在的相互作用。

另一个潜在方向是探索如何让更具技术稳健性的基础设施变得易于使用。例如，Atelier 目前支持少量 LoRA 和检查点，但若能使用专门训练的角色，就能大幅提升画架所支持的各类活动。此外，如何为画布增添更多智能，或以无需计算思维的方式支持工作流串联或嵌套，也是值得探讨的问题。

最后，对原型采样及其实际应用的探索，促使我们进一步研究如何更好地理解从业者如何将 Atelier 用于自己的创作过程，以及他们的想法和思路如何随时间演变。未来有趣的研究可以让从业者在一段时间内使用该系统，观察他们能创作出什么、会制定哪些策略，以及不同的人如何使用画布和所提供的工具。

11 结论

Sawyer 对创造性理论的汇编表明，顿悟无论看起来多么突然，都是一个系统性的处理过程，始终与从业者此前的工作、草图、笔记本等相联系 [82]。这凸显了整理收藏集、呈现想法并反思处理过程的重要性。将采样与原型制作紧密交织——原型采样——有助于更好地理解所收集的信息与想法呈现方式之间的关系。如今，生成式人工智能加速了这两项重要活动，并模糊了它们之间的界限，这一点尤为明显。通过推导并实践原型采样，以及理解信息在处理过程中的流动方式，我们得以将 Atelier 设计为一种开放式方法，用于融合思考与实践、将复杂工作流抽象为独立模块式画架，并梳理和组织创作成果。

Atelier 以原型采样为核心构建。Atelier 致力于让创作过程变得更加具体，并阐释在生成式人工智能的新时代，创意工作并非在于找到合适的提示词或使用最新的模型，而在于创作者所做出的每一个细微决定和采取的每一个审慎行动，以及合适的工具如何助力他们将想法变为现实。

参考资料

[1] 2005。 你如何进行设计？ https://www.dubberly.com/articles/how-do-you- design.html [2] Adobe。2025年。*Adobe Firefly*。https://firefly.adobe.com/ [3] Stability AI。2024年。*Stability AI*。https://stability.ai/ [4] 阿里巴巴。2024年。*Flux Turbo*。https://huggingface.co/alimama-creative/FLUX.1- Turbo-Alpha [5] Rifat Mehreen Amin、Oliver Hans Kühle、Daniel Buschek 和 Andreas Butz。2025年。面向创意写作的可组合提示工作空间：使用动态小组件进行探索与迭代。载于*CHI*计算机系统中的人为因素会议扩展摘要集论文集（*CHI EA '25*）。

计算机协会，纽约州纽约市，美国，第144篇文章，共11页。https://doi.org/10.1145/3706599.3720243。/。Victor Nikhil Antony 和 Chien-Ming Huang。2024年。ID.8：使用生成式人工智能共同创作视觉故事。《ACM交互智能系统汇刊》14，3，第20篇文章（2024年8月），页。https://doi.org/。/Anysphere。。Cursor。https://cursor.com/home [8] Ann Blandford、Dominic Furniss 和 Stephann Makri。2016年。定性人机交互研究：深入幕后。Morgan & Claypool出版社。[9] Margaret A Boden。2004年。创造性思维：神话与机制。Routledge。[10] Olha Bondarenko 和 Ruud Janssen。2005年。手边的文档：从纸质文档中学习，以改进数字技术。载于SIGCHI计算系统中的人为因素会议论文集。121–130。[11] Olha Bondarenko、Ruud Janssen 和 Samuël Driessen。2010年。个人文档管理系统的设计需求。《美国信息科学与技术学会期刊》61，3（2010年），468–482。[12] Michael Bostock、Vadim Ogievetsky 和 Jeffrey Heer。2011年。D³数据驱动文档。《IEEE可视化与计算机图形学汇刊》17，12（2011年），2301–2309。[13] Frederik Brudy、David Ledo、Michel Pahud、Nathalie Henry Riche、Christian Holz、Anand Waghmare、Hemant Bhaskar Surale、Marcus Peinado、Xiaokuan Zhang、Shannon Joyner、Badrish Chandramouli、Umar Farooq Minhas、Jonathan Goldstein、William Buxton 和 Ken Hinckley。2020年。SurfaceFleet：探索不受设备、应用程序、用户和时间限制的分布式交互。载于第33届ACM用户界面软件与技术年会论文集（美国虚拟活动）（UIST '20）。计算机协会，纽约州纽约市，美国，7–21。10.1145/3379337.3415874。[14] https://doi.org/。/。2010。Bill Buxton。。勾勒用户体验：正确进行设计，以及正确的设计。Morgan kaufmann。[15] Kathy Charmaz。2006年。构建扎根理论：定性分析实用指南。sage。[16] Dar-Yen Chen、Hmrishav Bandyopadhyay、Kai Zou 和 Yi-Zhe Song。2025年。归一化注意力引导：扩散模型的通用负向引导。arXiv预印本 arXiv:2505.21179（2025年）。[17] DaEun Choi、Sumin Hong、Jeongeon Park、John Joon Young Chung 和 Juho Kim。2024年。CreativeConnect：利用生成式人工智能支持平面设计构思中的参考素材重组。载于2024年CHI计算机系统中的人为因素会议论文集（美国夏威夷檀香山）（CHI '24）。计算机协会，纽约州纽约市，美国，第1055篇文章，共25页。https://doi.org/10.1145/3613904.3642794。/。[18] Stella Christie 和 Dedre Gentner。2014年。语言帮助儿童在经典类比任务中取得成功。认知科学38，2（2014年），383–397。[19] John Joon Young Chung 和 Eytan Adar。2023年。PromptPaint：通过类绘画媒介交互引导文本到图像生成。载于第36届ACM用户界面软件与技术年会论文集（美国加利福尼亚州旧金山）（UIST '23）。计算机协会，纽约州纽约市，美国，第6, 17页。10.1145/3586183.3606777。[20] https://doi.org/。/。2024。ComfyUI。。https://www.comfy.org/ [21] Nigel Cross。1997年。创造性设计的描述性模型：应用于一个实例。《设计研究》18，4（1997年），427–440。[22] Nigel Cross。2011年。理解设计思维。设计领域博士研究札记：来自米兰理工大学的贡献。FrancoAngeli（2011年），第19–37页。[23] Hai Dang、Frederik Brudy、George Fitzmaurice 和 Fraser Anderson。2023年。WorldSmith：使用生成式人工智能进行世界构建的迭代式和表达式提示。载于第36届ACM用户界面软件与技术年会论文集。1–17。[24] Alan Dix 和 Layda Gongora。2011年。外化与设计。载于第二届设计中的创造力与创新会议论文集。31–42。[25] Paul Dourish。2001年。行动所在之处：具身交互的基础。麻省理工学院出版社。[26] Claudia Eckert 和 Martin Stacey。2000年。灵感来源：一种设计语言。《设计研究》21，5（2000年），523–538。[27] Figma。2025年。介绍Figma Weave：Figma中下一代原生人工智能创作。https://www.figma.com/blog/welcome-weavy-to-figma/ [28] FLORA。2025年。FLORA。https://www.florafauna.ai/academy/quickstart [29] Christiane Floyd。1984年。一种系统化的原型制作观察。载于《原型制作方法》。Springer，1–18。[30] FLUX.1。2024年。FLUX.1。https://getimg.ai/models/flux [31] Jonas Frich、Michael Mose Biskjaer、Lindsay MacDonald Vermeulen、Christian Remy 和 Peter Dalsgaard。2019年。创意专业人士跨领域使用数字工具的策略。载于2019年创造力与认知会议论文集。210–221。[32] William Gaver、Peter Gall Krogh、Andy Boucher 和 David Chatting。2022年。涌现作为基于实践的设计研究特征。载于2022年ACM交互系统设计会议论文集。517–526。[33] John S Gero。1994年。创造性设计过程的计算模型。载于《人工智能与创造力：跨学科方法》。Springer，269–281。[34] Vinod Goel 和 Peter Pirolli。1992年。设计问题空间的结构。《认知科学》16卷3期（1992年），395–429。[35] Saul Greenberg、Sheelagh Carpendale、Nicolai Marquardt 和 Bill Buxton。2011年。《勾勒用户体验：练习册》。爱思唯尔。[36] Sarah Henderson 和 Ananth Srinivasan。2009年。个人数字文档结构的实证分析。载于《人机界面研讨会》。Springer，394–403。[37] Ken Hinckley、Shengdong Zhao、Raman Sarin、Patrick Baudisch、Edward Cutrell、Michael Shilman 和 Desney Tan。2007年。InkSeine：用于主动笔记记录的原位搜索。载于《SIGCHI计算系统中的人为因素会议论文集》（美国加利福尼亚州圣何塞）（CHI '07）。计算机协会，美国纽约州纽约，251–260。10.1145/1240624.1240666。[38] https://doi.org/。/. Jonathan Ho 和 Tim Salimans。无分类器扩散引导。arXiv预印本 arXiv:2207.12598（2022年）。[39] Ruey-Yun Horng、Ching-Wen Wang、Yung-Chieh Yen、Chia-Ying Lu 和 Chien-Tao Li。2021年。一种基于概念组合理论的想象力行为测量。《创造力研究期刊》33卷4期（2021年），376–387。[40] Alexander Ivanov、David Ledo、Tovi Grossman、George Fitzmaurice 和 Fraser Anderson。2022年。MoodCubes：用于收集、发现和构想灵感素材的沉浸式空间。载于《2022年ACM交互系统设计会议论文集》（虚拟活动，澳大利亚）（DIS '22）。计算机协会，美国纽约州纽约，189–203。https://doi.org/10.1145/3532106.3533565[41] William Jones。2010年。《让寻得之物不再遗失：个人信息管理的研究与实践》。Morgan Kaufmann。[42] James C Kaufman。2016年。《创造力入门》。Springer出版公司。[43] Alison Kidd。1994年。知识工作者留下的标记。载于《SIGCHI计算系统中的人为因素会议论文集》。186–191。[44] Janin Koch、Nicolas Taffin、Michel Beaudouin-Lafon、Markku Laine、Andrés Lucero 和 Wendy E. Mackay。2020年。ImageSense：支持多样化人机协作的智能协作构思工具。《ACM人机交互汇刊》4卷CSCW1期，文章45（2020年5月），27页。https://doi.org/10.1145/3392850。/. Janin Koch、Nicolas Taffin、Andrés Lucero 和 Wendy E. Mackay。2020年。[45] SemanticCollage：利用语义标签丰富数字情绪板设计。载于 DIS '20——《2020年交互系统设计会议》。ACM，407。https://doi.org/10.1145/3357236.3395494。/. Magda Kochanowska 和 Weronika Rochacka Gagliardi。2021年。[46] 双钻模型：追求简洁与灵活。载于《设计视角II：研究、教育与实践》。Springer，19–32。[47] Vladimir Kulikov、Matan Kleiner、Inbar Huberman-Spiegelglas 和 Tomer Michaeli。2024年。FlowEdit：使用预训练流模型进行无反演的基于文本的编辑。arXiv预印本 arXiv:2412.08629（2024年）。[48] David Ledo。2025年。生成式转描：关于生成式视频到视频实践的第一人称自传式探索。载于《2025年创造力与认知会议论文集》。931–948。[49] David Ledo、Steven Houben、Jo Vermeulen、Nicolai Marquardt、Lora Oehlberg 和 Saul Greenberg。2018年。人机交互工具包研究的评估策略。载于《2018年CHI计算机系统中的人为因素会议论文集》。1–17。[50] Joanne Leong、David Ledo、Thomas Driscoll、Tovi Grossman、George Fitzmaurice 和 Fraser Anderson。2025年。Paratrouper：使用生成式人工智能探索式创建角色阵容视觉图。载于《2025年CHI计算机系统中的人为因素会议论文集》。1–20。[51] Youn-Kyung Lim、Erik Stolterman 和 Josh Tenenberg。2008年。原型剖析：作为过滤器的原型，以及作为设计构想体现的原型。《ACM计算机与人类交互汇刊》（TOCHI）15卷2期（2008年），1–27。[52] Haichuan Lin、Yilin Ye、Jiazhi Xia 和 Wei Zeng。2025年。SketchFlex：通过基于区域的草图促进文生图生成中的空间语义连贯性。载于《2025年CHI计算机系统中的人为因素会议论文集》（CHI '25）。计算机协会，美国纽约州纽约，文章546, 19。10.1145/3706598.3713801 页。[53] https://doi.org/。/. Robert S Lockhart、Mary Lamon 和 Mary L Gick。1988。简单顿悟问题中的概念迁移。《记忆与认知》16卷1期（1988年），36–44。[54] Brian Logan 和 Tim Smithers。2013年。创造力与设计作为探索。载于《创造力建模与基于知识的创造性设计》。心理学出版社，139–175。[55] Yuwen Lu、Alan Leung、Amanda Swearngin、Jeffrey Nichols 和 Titus Barik。2025年。Misty：通过交互式概念融合进行用户界面原型设计。载于《2025年CHI计算机系统中的人为因素会议论文集》（CHI '25）。计算机协会，美国纽约州纽约，文章1108, 17。10.1145/3706598.3713924 页。https://doi.org/。/. Andrés Lucero。2012。[56] 框定、对齐、悖论化、抽象化与引导：设计情绪板如何发挥作用。载于《交互系统设计会议论文集》（DIS '12）。计算机协会，美国纽约州纽约，438–447。https://doi.org/10.1145/2317956.2318021

[57] Andrés Lucero。2012年。框定、对齐、悖论化、抽象化与引导：设计情绪板如何发挥作用。载于《交互系统设计会议论文集》（英国泰恩河畔纽卡斯尔）（DIS '12）。计算机协会，美国纽约州纽约，438–447。https://doi.org/10.1145/2317956.2318021。/. Matthew C Makel 和 Jonathan A Plucker。2014年。创造力不止于新颖性：重新思考将复制视为一种创造行为。（2014年）。[59] Thomas W Malone。1983年。人们如何整理自己的办公桌？对办公信息系统设计的启示。《ACM信息系统汇刊》（TOIS）1卷1期（1983年），99–112。[60] Nicolai Marquardt、Asta Roseway、Hugo Romat、Payod Panda、Michel Pahud、Gonzalo Ramos、Steven M Drucker、Andrew D Wilson、Ken Hinckley 和 Nathalie Riche。2025年。ImaginationVellum：融入空间提示、生成式笔触和构思历史的生成式人工智能构思画布。载于《第38届ACM用户界面软件与技术研讨会年会论文集》。1–19。[61] Ginés Hidalgo Martınez。2019年。OpenPose：全身姿态估计。博士论文。卡内基梅隆大学，美国宾夕法尼亚州匹兹堡。[62] Deana Mcdonagh 和 Ian Storer。2004年。情绪板作为设计催化剂与资源：研究一个鲜少受到关注的领域。《设计杂志》7卷3期（2004年），16–31。https://doi.org/10.2752/146069204789338424 arXiv: https://doi.org/10.2752/146069204789338424。/ Midjourney。Midjourney文档。2024。https://docs.midjourney.com/ [64] Milanote。2025年。Milanote。https://milanote.com/ [65] Miro。2025年。Miro。https://miro.com/ [66] Michele I Mobley、Lesli M Doares 和 Michael D Mumford。1992年。创造力的过程分析模型：组合与重组过程的证据。《创造力研究期刊》5卷2期（1992年），125–155。[67] Mural。2025年。FLORA。https://www.mural.co/use-case/design-canvas [68] Brad Myers、Scott E Hudson 和 Randy Pausch。2000年。用户界面软件工具的过去、现在与未来。《ACM计算机与人类交互汇刊》（TOCHI）7卷1期（2000年），3–28。[69] Carman Neustaedter 和 Phoebe Sengers。2012年。人机交互研究中的自传式设计：通过亲自使用来进行设计与学习。载于《交互系统设计会议论文集》。514–523。[70] Stellan Ohlsson。1992年。对顿悟及相关现象的信息处理解释。《思维心理学进展》（1992年），1–44。[71] Dan R Olsen Jr。2007年。评估用户界面系统研究。载于《第20届ACM用户界面软件与技术研讨会年会论文集》。251–258。[72] OpenAI。2024年。DALLE·3。https://openai.com/index/dall-e-3/ [73] OpenAI。2025年。ChatGPT。https://chatgpt.com/overview/ [74] Eyal Oren。2006年。信息管理与知识工作研究概述：对语义桌面的启示。SemDesk（2006年）。[75] Eyal Oren。2006年。信息管理与知识工作研究概述：对语义桌面的启示。SemDesk（2006年）。出版方：Citeseer。[76] Andrew Ortony。1979年。超越字面相似性。《心理学评论》86卷3期（1979年），161。[77] Srishti Palani、David Ledo、George Fitzmaurice 和 Fraser Anderson。2022年。“我不想觉得自己在20世纪60年代工厂里工作”：从业者对创造力支持工具采用的视角。载于《2022年CHI计算机系统中的人为因素会议论文集》。1–18。[78] Xiaohan Peng、Janin Koch 和 Wendy E Mackay。2024年。Designprompt：使用多模态交互与生成式人工智能进行设计探索。载于《2024年ACM交互系统设计会议论文集》。804–818。[79] RunwayML。2024年。Gen-3 Alpha。https://runwayml.com/ [80] Vishnu Sarukkai、Lu Yuan、Mia Tang、Maneesh Agrawala 和 Kayvon Fatahalian。2024年。块面与细节：为草图到图像生成提供支架。载于《第37届ACM用户界面软件与技术研讨会年会论文集》（美国宾夕法尼亚州匹兹堡）（UIST '24）。计算机协会，美国纽约州纽约，文章33，13页。https://doi.org/10.1145/3654777.3676444[81] R Keith Sawyer。2021年。创作过程的迭代与即兴本质。《创造力期刊》31卷（2021年），100002。[82] Robert Keith Sawyer 和 Danah Henriksen。2024年。《解释创造力：人类创新的科学》。牛津大学出版社。[83] Donald A Schön。2017年。《反思型实践者：专业人士如何在行动中思考》。Routledge。[84] Abigail J Sellen 和 Richard HR Harper。2003年。《无纸化办公室的神话》。麻省理工学院出版社。[85] Yulin Shen、Yifei Shen、Jiawen Cheng、Chutian Jiang、Mingming Fan 和 Zeyu Wang。2024年。神经画布：通过整合三维草图绘制与生成式人工智能支持舞台布景设计原型制作（CHI '24）。计算机协会，美国纽约州纽约，文章1056，18页。https://doi.org/10.1145/3613904.3642096。/. Edward E Smith 和 Daniel N Osherson。1984年。利用原型概念进行概念组合。《认知科学》8卷4期（1984年），337–361。[87] Pieter Jan Stappers 和 Elisa Giaccardi。2017年。通过设计进行研究。载于《人机交互百科全书》。交互设计基金会，1–94.[88] Evgeny Stemasov、David Ledo、George Fitzmaurice 和 Fraser Anderson。2023年。《沉浸式采样：探索媒体丰富的沉浸式空间中的采样，以开拓未来创作实践》。载于*2023*年*ACM*交互系统设计会议论文集。212–229。[89] Robert J Sternberg。2003年。创造力作为决策过程的发展。创造力与发展 1 (2003), 91–138。[90] Luming Tang、Nataniel Ruiz、Qinghao Chu、Yuanzhen Li、Aleksander Holynski、David E. Jacobs、Bharath Hariharan、Yael Pritch、Neal Wadhwa、Kfir Aberman 和 Michael Rubinstein。2024年。《RealFill：用于真实图像补全的参考驱动生成》。*ACM*图形学汇刊 43, 4, Article 135 (2024年7月), 12 页。https://doi.org/10.1145/3658237。/ tldraw。。[91] 2025。*tldraw*。https://tldraw.dev/quick-start [92] Jagoda Walny。2016年。《用草图思考：借助日常视觉内容进行信息可视化》。(2016)。[93] Qian Wan 和 Zhicong Lu。2023年。《GANCollage：由GAN驱动的数字情绪板，助力创造力支持中的构思》。载于*2023*年*ACM*交互系统设计会议论文集（美国宾夕法尼亚州匹兹堡）*(DIS '23)*。计算机协会，纽约，纽约州，美国，136–146。https://doi.org/10.1145/3563657.3596072。/。Team Wan、Ang Wang、Baole Ai、Bin Wen、Chaojie Mao、Chen-Wei Xie、Di Chen、Feiwu Yu、Haiming Zhao、Jianxiao Yang、Jianyuan Zeng、Jiayu Wang、Jingfeng Zhang、Jingren Zhou、Jinkai Wang、Jixuan Chen、Kai Zhu、Kang Zhao、Keyu Yan、Lianghua Huang、Mengyang Feng、Ningyi Zhang、Pandeng Li、Pingyu Wu、Ruihang Chu、Ruili Feng、Shiwei Zhang、Siyang Sun、Tao Fang、Tianxing Wang、Tianyi Gui、Tingyu Weng、Tong Shen、Wei Lin、Wei Wang、Wei Wang、Wenmeng Zhou、Wente Wang、Wenting Shen、Wenyuan Yu、Xianzhong Shi、Xiaoming Huang、Xin Xu、Yan Kou、Yangyu Lv、Yifei Li、Yijing Liu、Yiming Wang、Yingya Zhang、Yitong Huang、Yong Li、You Wu、Yu Liu、Yulin Pan、Yun Zheng、Yuntao Hong、Yupeng Shi、Yutong Feng、Zeyinzi Jiang、Zhen Han、Zhi-Fan Wu 和 Ziyu Liu。2025年。《Wan：开放且先进的大规模视频生成模型》。arXiv: 2503.20314 [cs.CV] 2503.20314。cs.CV [94] https://arxiv.org/abs/。Wen-Fan Wang、Chien-Ting Lu、Nil Ponsa i Campanyà、Bing-Yu Chen 和 Mike Y. Chen。2025年。《AIdeation：为概念设计师设计的人机协同构思系统》。载于*2025*年*CHI*计算机系统中的人为因素会议论文集（*CHI '25*）。计算机协会，纽约，纽约州，美国，文章 21，28 页。https://doi.org/10.1145/3706598.3714148[96] Robert Weisberg。1986年。创造力：天才及其他迷思。WH Freeman/Times Books/Henry Holt & Co. [97] Robert W Weisberg。1993年。《创造力：超越天才迷思》。*(*无标题*)*(1993)。[98] Mikael Wiberg 和 Erik Stolterman。2014年。什么造就了新颖的原型？交互设计研究中的知识贡献考量。载于第八届北欧人机交互会议论文集：有趣、快速、基础。531–540。[99] Shaojin Wu、Mengqi Huang、Yufeng Cheng、Wenxu Wu、Jiahe Tian、Yiming Luo、Fei Ding 和 Qian He。2025年。《USO：通过解耦与奖励学习实现统一风格与主体驱动生成》。arXiv:2508.18966 [cs.CV] https://arxiv.org/abs/2508.18966。[100] Haijun Xia、Bruno Araujo、Tovi Grossman 和 Daniel Wigdor。2016年。《面向对象的绘图》。载于*2016*年*CHI*计算机系统中的人为因素会议论文集。4610–4621。[101] Bin Xiao、Haiping Wu、Weijian Xu、Xiyang Dai、Houdong Hu、Yumao Lu、Michael Zeng、Ce Liu 和 Lu Yuan。2024年。《Florence-2：为各种视觉任务推进统一表征》。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。4818–4829。[102] Lihe Yang、Bingyi Kang、Zilong Huang、Zhen Zhao、Xiaogang Xu、Jiashi Feng 和 Hengshuang Zhao。2024年。《Depth Anything V2》。神经信息处理系统进展37 (2024), 21875–21911。[103] Chengzhi Zhang、Weijie Wang、Paul Pangaro、Nikolas Martelaro 和 Daragh Byrne。2023年。《利用设计草图作为输入的生成式图像人工智能：机遇与挑战》。载于第*15*届创造力与认知会议论文集（虚拟活动，美国）*(C&C '23)*。计算机协会，纽约，纽约州，美国，254–261 页。https://doi.org/10.1145/3591196.3596820。/。[104] 2024。Lei Zhang、Jin Pan、Jacob Gettig、Steve Oney 和 Anhong Guo。。《VRCopilot：在VR中使用生成式人工智能模型创作三维布局》。载于第*37*届*ACM*用户界面软件与技术年会论文集（美国宾夕法尼亚州匹兹堡）*(UIST '24)*。计算机协会，纽约，纽约州，美国，文章96, 13 页。https://doi.org/10.1145/3654777.3676451。/。[105] 2023。Lvmin Zhang、Anyi Rao 和 Maneesh Agrawala。。《为文生图扩散模型添加条件控制》。载于*IEEE/CVF*国际计算机视觉会议论文集。3836–3847。[106] Shihao Zhao、Dongdong Chen、Yen-Chun Chen、Jianmin Bao、Shaozhe Hao、Lu Yuan 和 Kwan-Yee K Wong。2024年。《Uni-ControlNet：为文生图扩散模型提供一体化控制》。神经信息处理系统进展36 (2024)。

[107] John Zimmerman、Jodi Forlizzi 和 Shelley Evenson。2007年。作为人机交互领域交互设计研究方法的通过设计进行研究。载于论文集 *SIGCHI*计算系统中的人为因素会议。493–502。

一份问卷

为了解参与者如何看待生成式系统在其创作工作流中的作用与特质，我们请他们使用五点李克特量表（强烈不同意 → 强烈同意）对一组陈述进行评分。下文列出了每个问卷条目的完整措辞，以及论文中使用的标签。

(1) **所有权**：我觉得自己对用 ____创建的内容拥有所有权。(2) **能动性**：使用 ____时，我感到自己拥有控制力和能动性。(3) **早期阶段思考**：____ 支持我的早期阶段思考。(4) **后期阶段思考**：____ 支持我的后期阶段思考。(5) **灵感**：我使用 ____ 来获取灵感。(6) **最终产品**：来自 ____ 的创建成果可能会出现在最终产品中。(7) **可塑性**：____ 给人的感觉很有可塑性。(8) **与流程的整合**：____ 能够顺畅地融入我的创作过程。(9) **与工具的整合**：____ 能够顺畅地与我的其他工具整合。(10) **端到端支持**：____ 支持端到端的工作流程。(11) **反映意图**：我确信 ____会反映我的设计意图。(12) **愉悦感**：我喜欢使用 ____。(13) **乐趣**：使用 ____ 很有乐趣。(14) **专业感**：____ 给人的感觉达到了专业级别。(15) **修复输出的能力**：当我使用 ____得到不符合预期的输出结果时，我觉得自己可以轻松修正它，以获得想要的结果。(16) **表现力**：____ 给人的感觉很有表现力。(17) **被视为工具**：我将 ____ 视为工具。(18) **被视为协作者**：我将 ____ 视为协作者。(19) **赋能**：____ 让我感到自己得到了赋能。(20) **创造力**：____ 让我感到自己富有创造力。