AI 游戏创作工具用户研究

大学生 AI 游戏创作工具用户研究

研究报告

量化图表版 | 生产端与消费端的实际使用、协作过程与评价条件

编辑：左涵俊

2026 年 7 月

目录

第一章 研究背景

1.1 AI 已进入游戏生产和发布的具体环节

2026 年 GDC 游戏行业状况调查收集了 2300 多名游戏行业从业者的回答。调查中，36% 的受访者表示已在工作中使用生成式 AI；常见用途包括代码辅助（47%）和原型制作（35%）。与此同时，52% 的受访者认为生成式 AI 对游戏行业产生了负面影响。[1] 这组结果说明，AI 已经进入实际工作，但用户是否接受它，仍与作品质量、责任归属和创作控制有关。

产品功能和发行规则也在发生变化。Unity AI 已能够理解场景、检查 GameObject 与组件、协助执行编辑器操作，并核对修改后的运行结果；Steam 的内容问卷则要求开发者说明作品中预先生成和运行时生成的 AI 内容。[2][3] AI 因而不再只出现在一次问答或单张素材生成中，它已经进入开发、检查、发布和玩家接触作品的多个环节。

1.2 研究从哪里出发

公开资料能够说明产品提供了哪些功能，却很难回答这些功能进入真实项目后发生了什么。专业工具可能加快编码和调试，也可能在项目理解、代码维护和人工检查上带来新的要求；一站式平台能够快速生成可玩的初稿，但作品能否继续修改、稳定运行并完成发布，仍取决于生成之后的过程。作品进入玩家视野后，玩家是否注意到 AI、是否在意，以及他们最终按照什么标准评价作品，也需要通过实际体验来观察。

大学生常同时处在学习、课程项目、Game Jam、社团协作、独立创作和游戏体验等情境中。任务周期、技术基础和项目目标并不相同，因而能够呈现 AI 在不同条件下怎样帮助推进任务、又在哪里遇到问题。本研究据此把关注点放在大学生的具体使用和体验，而不是只整理产品功能。

1.3 核心概念

专业 AI 工具：进入已有专业工作流程、协助完成某一类具体任务的 AI 工具，可用于程序、美术、音乐音频和文本策划等环节。本报告只深入考察其中的程序开发。

一站式 AI 游戏创作平台：用户主要通过自然语言提出创意，并在同一平台中继续生成、修改、测试和发布作品的产品。

AI 参与作品与玩家感知：AI 在开发或内容生成中参与过的游戏作品。消费端关注玩家是否注意到这种参与、依据什么作出判断，以及判断怎样影响评价；这里记录的是玩家的主观感受，不把它当作准确识别结果。

1.4 本研究的定位与范围

本研究聚焦生产端与消费端两个方面。生产端包含两条创作路径：第一条以 Codex 为核心代表，考察专业 AI 编程工具如何进入大学生游戏开发的程序工作流，包括需求拆解、编码、调试、人工检查与项目接入；第二条以 TapTap 制造为核心代表，考察一站式 AI 游戏创作平台如何支持大学生用户从创意生成走向修改、可玩或发布。消费端则关注实际体验过一站式 AI 平台作品的大学生玩家，观察他们对 AI 使用的感受、作品评价与进一步支持意愿。

Codex 与 TapTap 制造分别呈现两种差异明显的产品路径。Codex 能够进入已有项目环境，读取和修改项目文件，并配合命令执行、测试与人工检查，适合观察 AI 如何嵌入专业程序工作流，以及用户如何与 AI 分工、验证和接管。TapTap 制造则以自然语言创作为主要入口，将代码、美术、音乐、试玩和发布等环节集中在同一流程中，适合观察用户如何在不先搭建完整专业开发流程的情况下，把创意持续推进为可玩的作品。前者侧重增强已有开发流程，后者侧重降低完整作品的制作门槛，这一差异正是本研究选择二者的原因。[4][5]

本报告是一项围绕大学生实际使用过程展开的用户研究，重点呈现人的诉求、人与 AI 的协作过程、创作推进中的顺利与受阻环节，以及玩家形成评价的条件。产品之间的优劣、行业整体采用情况和长期使用表现，需要结合更大范围的数据与持续观察另行研究。

第一章资料来源（访问日期：2026 年 7 月 28 日）

[1] GDC Festival of Gaming, “2026 State of the Game Industry”

[2] Unity, “Unity AI”

[3] Valve, “Steamworks Documentation: Content Survey”

[4] OpenAI, “Introducing Codex”

[5] TapTap，《“TapTap 制造”重磅发布：一款让想象力直接变成游戏的 AI 智能体》

第二章 研究目标与核心问题

2.1 研究目标

本研究旨在了解大学生在游戏开发、平台创作和作品体验中希望借助 AI 解决什么问题，并考察任务推进、作品评价及后续选择受到哪些条件影响。

2.2 核心问题

2.2.1 生产端——专业 AI 编程工具

以 Codex 为核心代表，大学生使用专业 AI 编程工具的核心诉求是什么？这类工具在不同任务中，主要是提升已有能力的效率，还是补足自身尚不具备或不够熟练的能力？在程序工作流中，人与 AI 如何协作，哪些环节进展顺利，哪些环节容易受阻？

2.2.2 生产端——一站式 AI 游戏创作平台

以 TapTap 制造为核心代表，大学生使用一站式 AI 游戏创作平台的核心诉求是什么？从想法生成到可玩或发布，用户如何修改、测试并推进作品，哪些环节容易受阻？

2.2.3 消费端——作品体验

大学生玩家体验 AI 参与创作的游戏时，对游戏中的 AI 使用有怎样的感受？他们是否能够感知 AI 参与、是否在意，并依据什么评价作品？哪些条件会影响他们继续游玩、关注、推荐、提供反馈或付费的意愿？

2.3 研究对象与主要证据

研究方向	研究范围与结论边界	主要数据来源
生产端：专业 AI 工具	聚焦专业 AI 工具中的程序开发路径，以 Codex 为核心代表，并以具有 Codex 使用经历的大学生为核心样本；其他程序开发工具的使用材料仅用于补充不同工具下的实际做法	专业工具相关问卷数据与开发者访谈
生产端：一站式创作平台	聚焦从自然语言生成创意到形成可玩、可继续修改或可发布作品的过程，以 TapTap 制造为核心代表；其他平台材料用于补充行业背景和不同产品的做法	相关问卷数据与开放回答
消费端：AI 参与作品体验	聚焦大学生玩家对 AI 参与游戏作品的实际体验，考察其主观感知、态度和作品评价，不将玩家的主观判断理解为对 AI 参与情况的准确识别	消费端问卷数据、作品体验记录与相关访谈

第三章 研究材料与样本

本报告的实证分析主要依据问卷、开放回答和5名大学生的深度访谈。问卷共导出117份记录，排除1份题目结构不兼容的历史记录和1份完全重复的提交后，保留115份可用记录。其中69份来自大学生，作为本报告的主要定量样本。

样本	人数	对应内容
大学生问卷样本	69	本报告的主要定量样本

样本	人数	对应内容
Codex主要使用者	15	主要使用Codex并完成相关问题，用于第四章
一站式平台创作样本	6	经使用记录复核后纳入，用于第五章
相关作品实际体验者	6	经作品与体验记录复核后纳入，用于第六章
深度访谈对象	5	补充程序开发和作品体验中的具体过程与原因

15人、6人和6人分别依据不同的使用或体验经历形成。其中，一站式平台创作样本包含TapTap制造及其他平台的使用记录，并非全部来自TapTap制造。因此，第五章以TapTap制造为核心案例，但问卷结果用于观察一站式平台创作中的共同情况，不作为TapTap制造的产品专属结论。目前访谈也尚未覆盖一站式平台创作者。

由于两个核心分组均只有6人，后文只呈现具体人数、回答差异和实际案例，不使用百分比推断更大范围的大学生群体。

第四章 生产端发现：以Codex为代表的大学生游戏程序构建实践

本章发现，Codex最集中的作用评价是提高开发速度和减少重复劳动。程序协作经过任务设定、代码生成或修改、运行检查、反馈调整和项目接入。三个案例中的开发者分别设置任务范围，并通过运行、逻辑检查或试玩验收结果。

本章材料包括15名大学生用户的问卷回答和开发者A、B、C三份深度访谈，并以6名已毕业用户作为对照。

4.1 问卷发现：AI编程工具参与下的大学生游戏程序构建

4.1.1 核心样本及其项目情境

69份大学生有效问卷中，50名进入程序能力与项目题，36名完成专业AI编程工具深答，其中15名以Codex为主要工具并完整回答程序协作问题。以下人数和比例均以15人为分母。

其中12人（80.0%）提供了游戏或互动开发背景，使用Unity、Godot或Unreal Engine。相关项目包括48小时限时开发、比赛演示项目、个人长期项目、团队项目和白盒原型，覆盖短期交付与长期维护情境。

4.1.2 程序基础与当前任务完成能力

15名用户覆盖基础课程学习、小功能编写、独立模块开发和成熟项目经验等程序能力层级。11人（73.3%）认为自己至少能够独立完成小游戏或程序模块，4人处于基础课程学习或小功能编写阶段。

当前任务完成能力方面，3人（20.0%）选择基本完成档，5人（33.3%）选择效率或质量受限档，7人（46.7%）选择完成程度较低的两档。

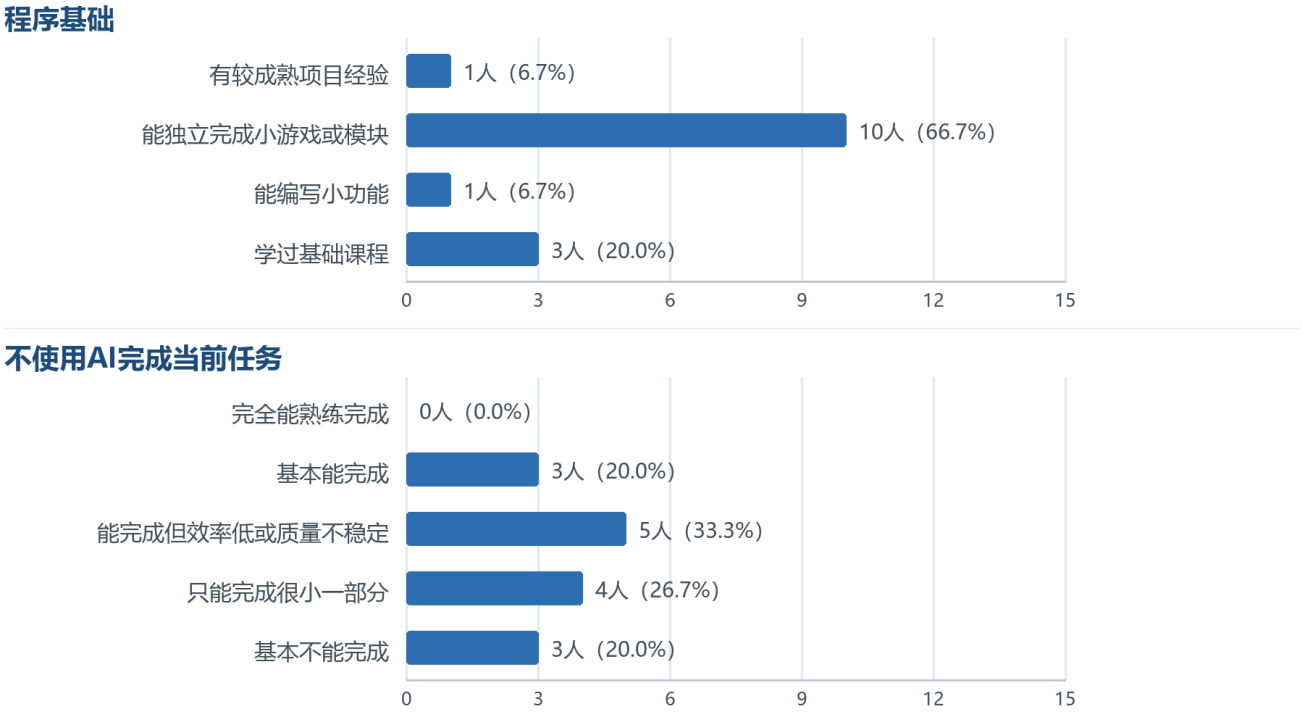

图4-1　大学生核心样本的程序基础与当前任务完成能力（N=15）

在10名认为自己能够独立完成小游戏或程序模块的用户中，4人在当前任务完成能力题中选择最低两档。5名代码质量低评价者中，4人具备独立模块开发经验，4人选择当前任务完成能力最低两档；中间评价组和高评价组分别为2人和1人。代码质量低评价也出现在具备独立模块开发经验的用户中，并较多集中于当前任务完成自评较低者。

4.1.3 AI工具进入的任务环节与实际作用

（1）从需求拆解延伸至已有项目修改

Codex主要进入需求拆解、代码生成、已有代码修改和调试等程序任务。15名用户中，13人（86.7%）用于代码生成，12人（80.0%）用于需求拆解，11人（73.3%）用于修改已有代码，10人（66.7%）用于调试程序错误。

项目阶段方面，核心功能开发为11人（73.3%），早期原型和性能或结构优化均为9人（60.0%），关卡或玩法迭代为8人（53.3%）。Codex的使用范围从早期原型延伸到核心功能、玩法迭代和结构优化。

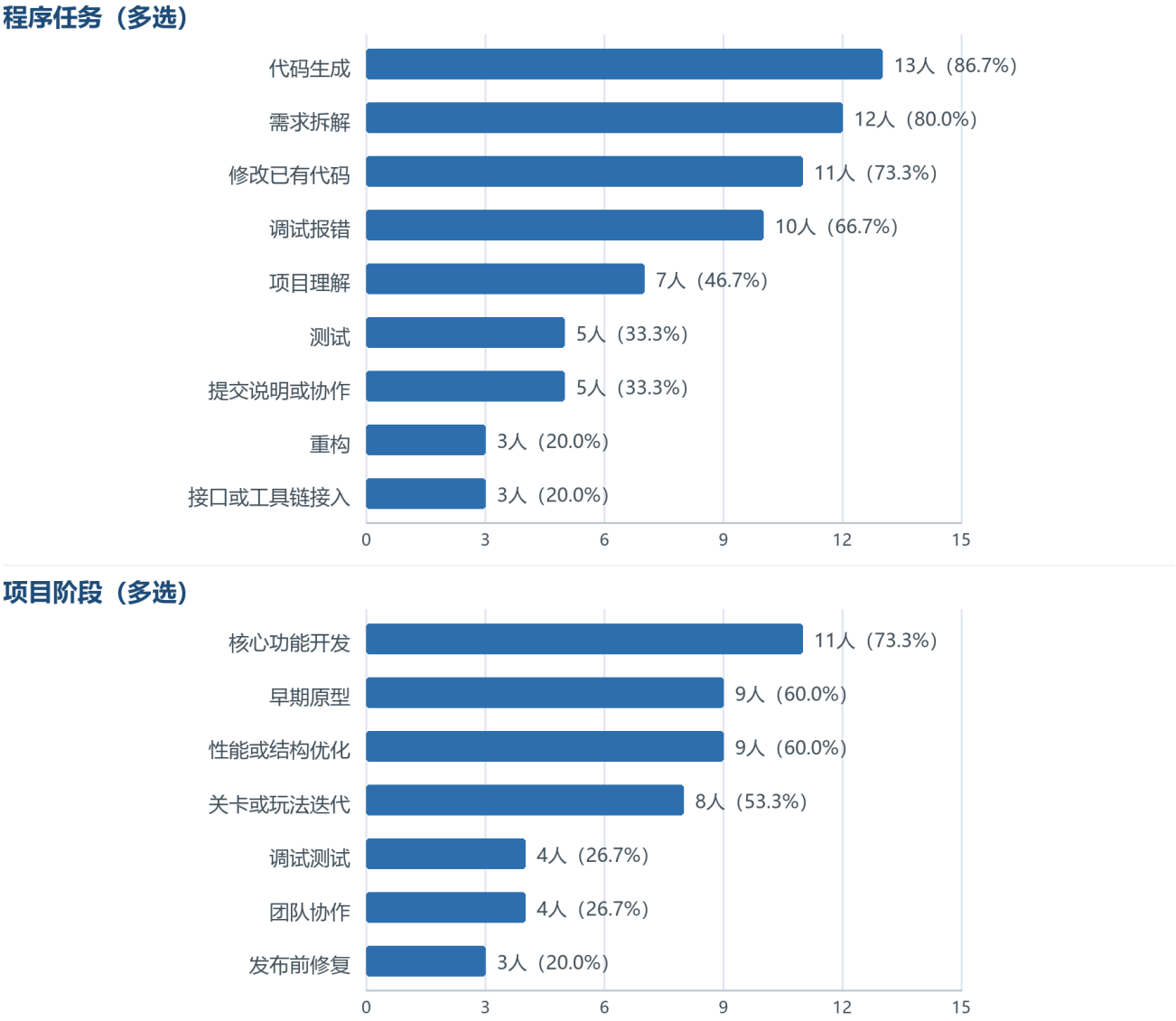

图4-2　Codex进入的程序任务与项目阶段（N=15，多选）

注：该题为多选题，各项占比之和超过100%。

样本dc9ec0cc在长期项目的地图编辑器中增加关卡编辑功能，过程如下：

“就比如我要给原本的地图编辑器加上一个新的关卡编辑器，我使用Codex，先让他只理解当前的架构并限制它过度设计，然后给出不同的相应方案，选择更改到我满意方案，最后让它直接修改，最后我自己测试能够在当前项目里运行并无Bug即可。”
——大学生问卷开放回答，样本dc9ec0cc

该用户先限定架构范围，再要求Codex提出方案和修改代码，最后通过运行测试验收结果。

（2）提速评价集中，能力补足获得较高评价，代码质量评价分化

提高开发速度和减少重复劳动是评价最集中的两项作用，两项均有14人（93.3%）给出4—5分；能力补足有11人（73.3%）给出4—5分。

样本8604f26f在暂时没有具体实现思路时，先使用Codex形成原型：

“实现某类暂时没有具体实现思路的需求，先把需求给AI写出原型，验证产出内容与需求无偏差后，会自己或AI辅助阅读源码理解逻辑链路，并向AI提出修改意见。”
——大学生问卷开放回答，样本8604f26f

面对暂时没有具体实现思路的需求，该用户先形成原型，再验证结果、阅读源码并提出修改意见。

用户对代码质量的评价分化：5人（33.3%）选择4—5分，5人（33.3%）选择3分，5人（33.3%）选择1—2分。

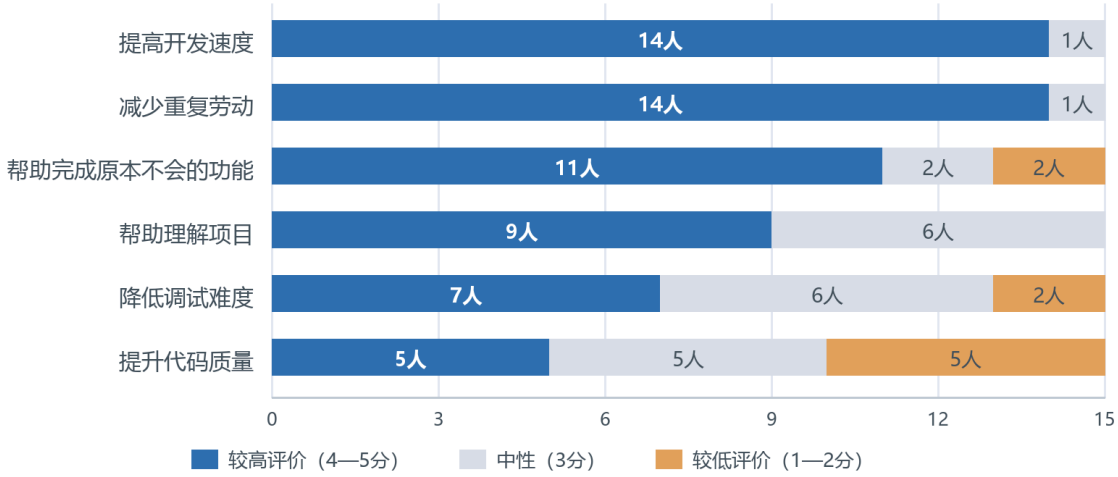

图4-3　大学生用户对Codex主要作用的评价（N=15）

5名代码质量低评价者均对提高开发速度给出4—5分，4人对减少重复劳动给出4—5分，4人的Codex产出仍有大部分进入项目。代码质量提升评价较低与速度认可、持续使用和项目接入同时出现。

开放回答呈现了不同评价对应的具体结果。样本8c9a9315使用Codex实现小地图，并继续调整位置、背景和方向指引：

“我要做小地图，使用Codex，AI给出了一个简略地图，我修改了地图位置、地图背景，还增加了地图的东西南北指引。最后效果很好。”
——大学生问卷开放回答，样本8c9a9315

样本9a4cfd90在界面生成任务中记录了布局和视觉问题：

“想要一个界面，我描述构想后AI生成前端组件重叠，视觉效果很差。”
——大学生问卷开放回答，样本9a4cfd90

另一名用户在实现软阴影、环境光遮蔽和体积雾时，先寻找参考资料并提供给Codex，并将参考材料作为效果调整依据。

低评价样本记录了界面组件重叠、视觉偏差、复杂图形需要参考材料和灯光调整后效果一般等问题；高评价样本8c9a9315在简略小地图基础上继续修改，最终达到预期。代码质量评价的差异集中在生成结果与任务要求的适配程度，以及调整后的最终效果。

样本44776c67对提升代码质量给出2分，其主要任务是需求拆解、项目理解和协作说明。开放回答记录了Codex帮助梳理行为分支并改善架构组织。该用户对需求整理和项目理解的评价较高，代码质量评价较低。

（3）人工判断贯穿生成结果进入项目的过程

阅读修改和运行测试是生成内容进入项目的主要步骤。14人（93.3%）会阅读后小幅修改，10人（66.7%）会运行测试；其他处理方式见图4-4，同一名用户会根据任务采用多种方式。

最终产出方面，11人（73.3%）表示大部分产出进入项目，2人（13.3%）表示修改后部分进入，2人（13.3%）表示少量进入。

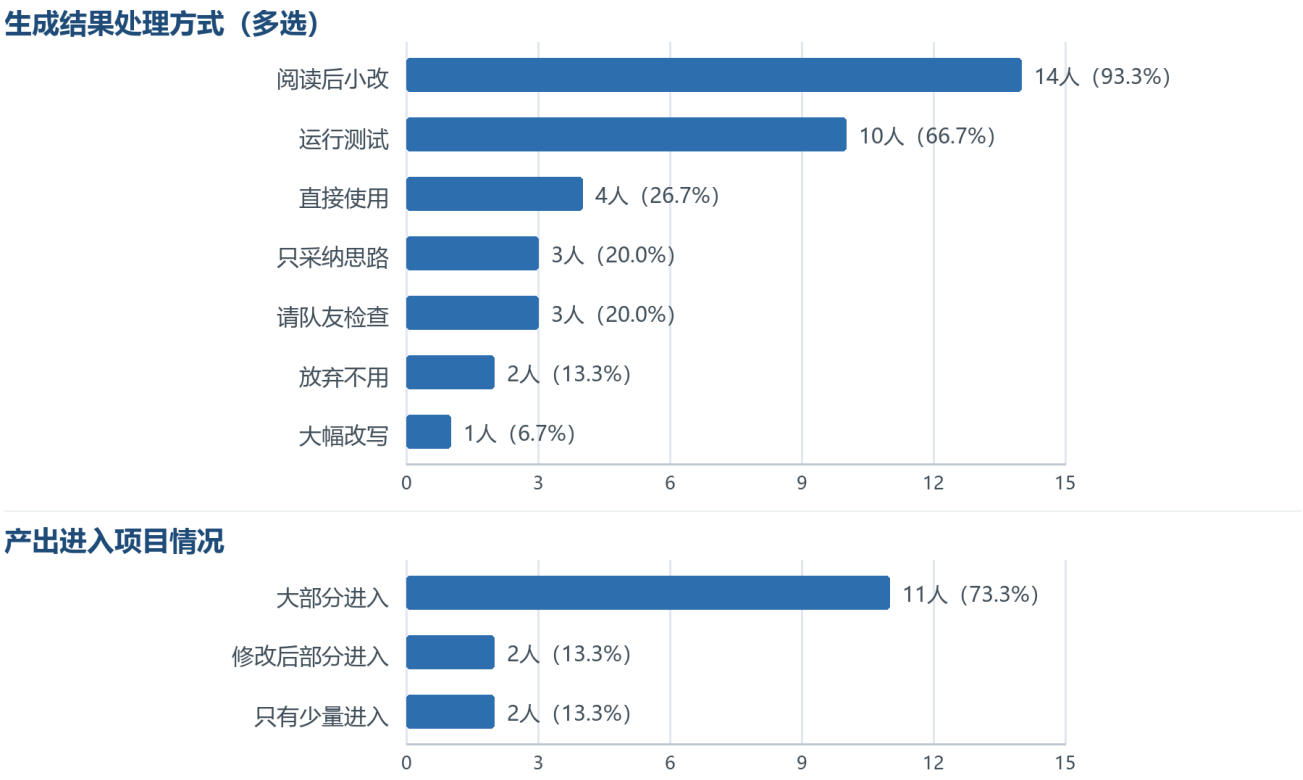

图4-4　生成结果的处理方式与项目接入情况（N=15）

对生成代码的判断与维护能力评价中，定位问题和解释代码均有11人（73.3%）给出4—5分，判断正确性和后续维护均为9人（60.0%），控制依赖风险为8人（53.3%）。

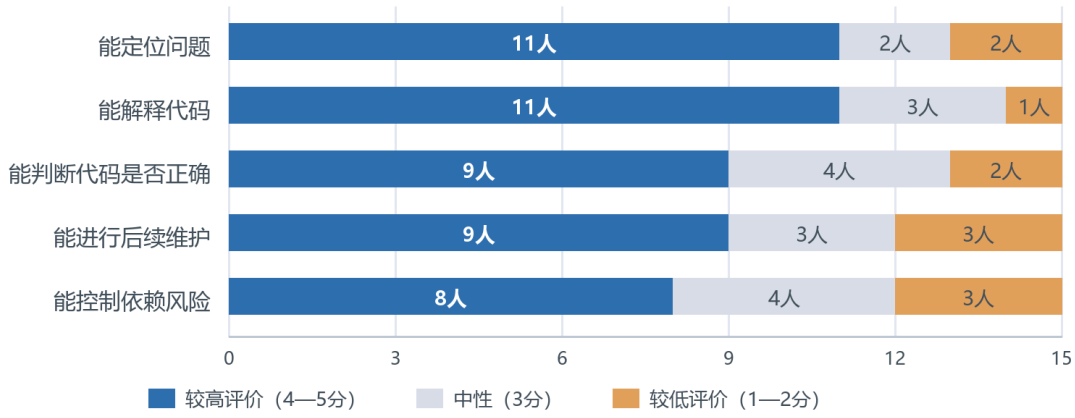

图4-5　用户对生成代码的判断与维护能力评价（N=15）

低评价组和高评价组各5人，均会阅读后小改。低评价组中4人运行测试、2人放弃部分结果、1人大幅改写；高评价组中3人运行测试、2人请队友检查。对判断正确性、后续维护和依赖风险给出4—5分者，低评价组分

别为1人、2人和1人，高评价组均为4人。低评价组还通过改写或放弃结果处理质量问题。

整体价值判断中，11人（73.3%）认为提高已有能力与补足不足对应不同任务，3人（20.0%）认为作用主要是提高已有能力的效率，1人（6.7%）评价整体价值有限。

唯一评价整体价值有限的样本e80e7f1a，对提高速度和提升代码质量均给出3分，对减少重复劳动给出5分；其当前任务完成能力为只能完成很小一部分，解释代码和后续维护分别为2分和1分，开放回答将结果概括为效果一般。在上述三项中，该样本只对减少重复劳动给出高分。

总体上，提高开发速度和减少重复劳动获得最集中评价。代码质量评价较低的用户仍认可速度，并通过阅读、测试、改写或筛选将结果接入项目。

4.2 从需求拆分到项目接入：三个案例中的程序配合过程

本节结合开发者A、B、C的项目，呈现任务提出、检查修正和项目接入过程。

开发者A参与48小时Unreal Engine限时开发，开发者B参与多个Unity短期演示项目，开发者C使用Godot开发限时作品和个人长期沙盒项目。

表4-6　三名开发者的项目情境与配合方式

案例	项目情境	Codex参与的主要任务	用户主要控制方式
开发者A	48小时Unreal Engine限时开发	钩锁物理约束、射线检测等核心玩法代码	拆分需求、分配对话、确认实现计划、检查逻辑并测试
开发者B	多个Unity短期演示项目	玩家控制、游戏对象创建、移动逻辑、参数公开和程序错误修改	描述功能、运行验证、继续修改或人工处理
开发者C	Godot限时开发与长期沙盒项目	抛锚玩法、物理手感、旧项目代码参考和架构理解	限定目录、提供参考项目、试玩反馈、按项目周期调整修改范围

4.2.1 用户先把创作目标整理成可执行任务

三名开发者都先把玩法想法整理成具体程序任务，再确定Codex的输入内容和修改范围。

开发者A的需求来自团队策划案。限时开发主题公布后，策划先形成基本方案，他再将程序工作拆开，分别交给Codex处理：

“等主题出来，我就等策划先把策划案大概写好，我这边把需求拆分一遍，然后分多个不同的对话喂给Codex。”
——开发者A访谈

开发者A把拆分后的任务放入不同对话，先查看实现方案，再让Codex修改代码。策划案提供玩法目标，开发者负责拆分任务和确认方案。

开发者B从单项功能出发，提出创建游戏对象、挂载脚本、实现移动逻辑和公开Unity属性面板变量等要求。他直接描述功能目标和可调参数，复杂需求则先经通用对话工具整理。

开发者C把已有项目作为参考材料。他要求Codex读取长期沙盒项目中的人物物理和“脐带物理”，再将连接方式迁移到Godot平台乱斗项目的抛锚玩法中，同时规定修改目录。

三名开发者使用Codex的直接动机分别是压缩限时开发中的实现时间、优先形成可运行演示，以及减少复制粘贴并同时推进多个功能。技术错误、要求遗漏和玩法偏差形成了后续检查和修改任务。

4.2.2 可运行结果需要经过持续测试和修正

生成内容进入项目之前，三名开发者都通过运行或试玩发现偏差，并把结果转化为下一轮修改要求。

开发者A参与48小时限时开发，完成速度是本次项目的优先目标。他表示，自行实现钩锁系统还需要查阅Unreal Engine文档并投入较多时间。确认方案后，Codex生成钩锁系统的大部分功能；首次运行发现射线检测通道问题后，开发者指出错误并要求修改。

开发者B同样把完成速度放在首位，短期Game Jam以形成可运行演示为目标。他将自己的开发方式概括为：

“Game Jam嘛，质量不是那么重要，反而先做出来是第一要务。”
——开发者B访谈

Codex生成玩家控制、对象逻辑和基础脚本后，开发者B根据运行情况继续询问AI、自己修改或寻求其他开发者帮助。

开发者B的运行和整合过程出现两类问题：多项要求同时输入时发生内容遗漏，部分生成方式偏离Unity常用组件并增加功能连接步骤。

开发者C的第一轮输出形成基础抛锚原型。试玩发现素材、锚节点、绳索连接和输入方向等偏差后，他补充具体要求，并继续调整力量、绳索长度、位移和空中控制，直至玩法手感符合预期。

4.2.3 不同项目中的修改范围与人工控制

三个案例中的开发者保留了不同的人工控制工作。

开发者A允许Codex承担钩锁系统和射线检测等核心代码，同时负责需求拆分、方案确认、逻辑检查和运行验收。

开发者B掌握项目框架和脚本功能，通过运行结果检查玩家移动、对象创建和数值调整，把控制重点放在原型运行和快速调整上。

开发者C根据项目周期和资源情况设置Codex的读取和修改范围。短期项目资源较少，他允许Codex创建文件并生成程序代码；长期项目已有持续积累的文件和资源，由他管理目录结构并限定AI的处理范围：

“长期项目文件夹不太会给AI动，短期项目资源少，AI随便建文件夹。”
——开发者C访谈

在长期项目中，开发者C让Codex整理字段、枚举和资源引用并生成配置说明，自己管理文件结构和资源修改。

三个案例的控制重点分别是需求与方案、功能与参数、目录与玩法手感。用户通过运行、逻辑检查或试玩完成验收，并据此决定生成内容如何进入项目。

4.3 问卷与案例共同支持的程序协作发现

以下三项发现分别概括工具价值、协作过程和人工控制。

4.3.1 核心诉求集中在快速形成可继续处理的结果

提高开发速度和减少重复劳动均有14人给出4—5分。5名代码质量低评价者也全部高度评价速度，其中4人的产出仍有大部分进入项目。开发者A压缩限时项目的实现时间，开发者B优先形成可运行演示，开发者C减少复制粘贴并同时推进多个功能。用户看重快速形成可继续处理的结果，并在生成后继续修正质量问题。

4.3.2 代码生成之后仍有持续检查和修改

程序协作包括目标和范围设定、代码生成或修改、运行检查、反馈调整和项目接入。生成后，14人会阅读并小幅修改，10人会运行测试。

开发者A检查射线检测逻辑，开发者B依据运行结果继续修改，开发者C通过试玩调整素材、输入和手感。三人都把检查结果转化为下一轮修改要求。

4.3.3 人工控制集中在任务范围、运行检查和项目接入

问卷中，定位问题和解释代码均有11人给出4—5分，判断正确性和后续维护均为9人，控制依赖风险为8人。

开发者A负责需求拆分和方案确认，开发者B检查运行结果和可调参数，开发者C根据短期、长期项目设置不同的目录和修改范围。

4.4 已毕业用户的对照观察与适用边界

本章结论依据15名大学生核心样本与开发者A、B、C三个案例形成。另有6名已毕业Codex用户完成相同问题，作为对照样本。

6名已毕业用户的当前任务完成能力均位于前三档；大学生核心样本中8人位于前三档、7人位于较低两档。

整体价值方面，已毕业用户在“主要提效”和“两类价值对应不同任务”中各有3人；大学生用户分别为3人和11人，另有1人评价整体价值有限。两组样本均以大部分产出进入项目为主。

表4-7　大学生与已毕业Codex用户的对照结果

比较内容	大学生用户（N=15）	已毕业用户（N=6）
不使用AI完成当前任务	3人基本能完成；5人能完成但效率或质量不稳定；7人只能完成少量或基本不能完成	1人完全能熟练完成；1人基本能完成；4人能完成但效率或质量不稳定
Codex整体价值	11人认为两类价值并存；3人认为主要提效；1人认为价值有限	3人认为两类价值并存；3人认为主要提效
产出进入项目	11人大部分进入；2人修改后部分进入；2人少量进入	4人大部分进入；2人修改后部分进入
常见人工处理	14人阅读后小改；10人运行测试；4人直接使用	4人阅读后小改；4人直接使用；2人运行测试

样本508ed5b3描述了从方案讨论到原型测试的过程：

“比如我要实现某个功能的基础原型架构，用Codex先计划和讨论方案可行性，AI会给出3个较为相近的方案，个人评估实现难度或者架构难度以及时间后选择方案原型然后测试，最后结果基本上都是满意的，细节肯定没法一次过，再小修小改一下就可以了。”
——已毕业问卷开放回答，样本508ed5b3

该用户先讨论方案，再评估难度和时间、测试原型并调整细节，与大学生案例中的过程相似。

第五章 生产端发现：以TapTap制造为案例锚点的一站式创作实践

本组记录显示，一站式平台获得的认可主要集中在原型形成、减少编码和快速验证，对调试、素材与逻辑整合的认可较弱。平台能够把创意推进为可以运行、试玩或测试的初步结果，但作品是否符合创作目标，仍需用户通过需求说明、方向选择、玩法测试和结果修改进行判断。后续创作问题集中在准确修改、玩法深化和表现统一。

本章分析6份大学生一站式平台记录。其中，taptap制造样本《炉釜猪》记录呈现可试玩作品的修改过程，一份造化工坊开放回答补充其他平台中的实现调整。一份非学生TapTap制造发布记录用于观察已发布作品记录中的创作工作。

5.1 一站式平台主要用于形成原型并验证玩法

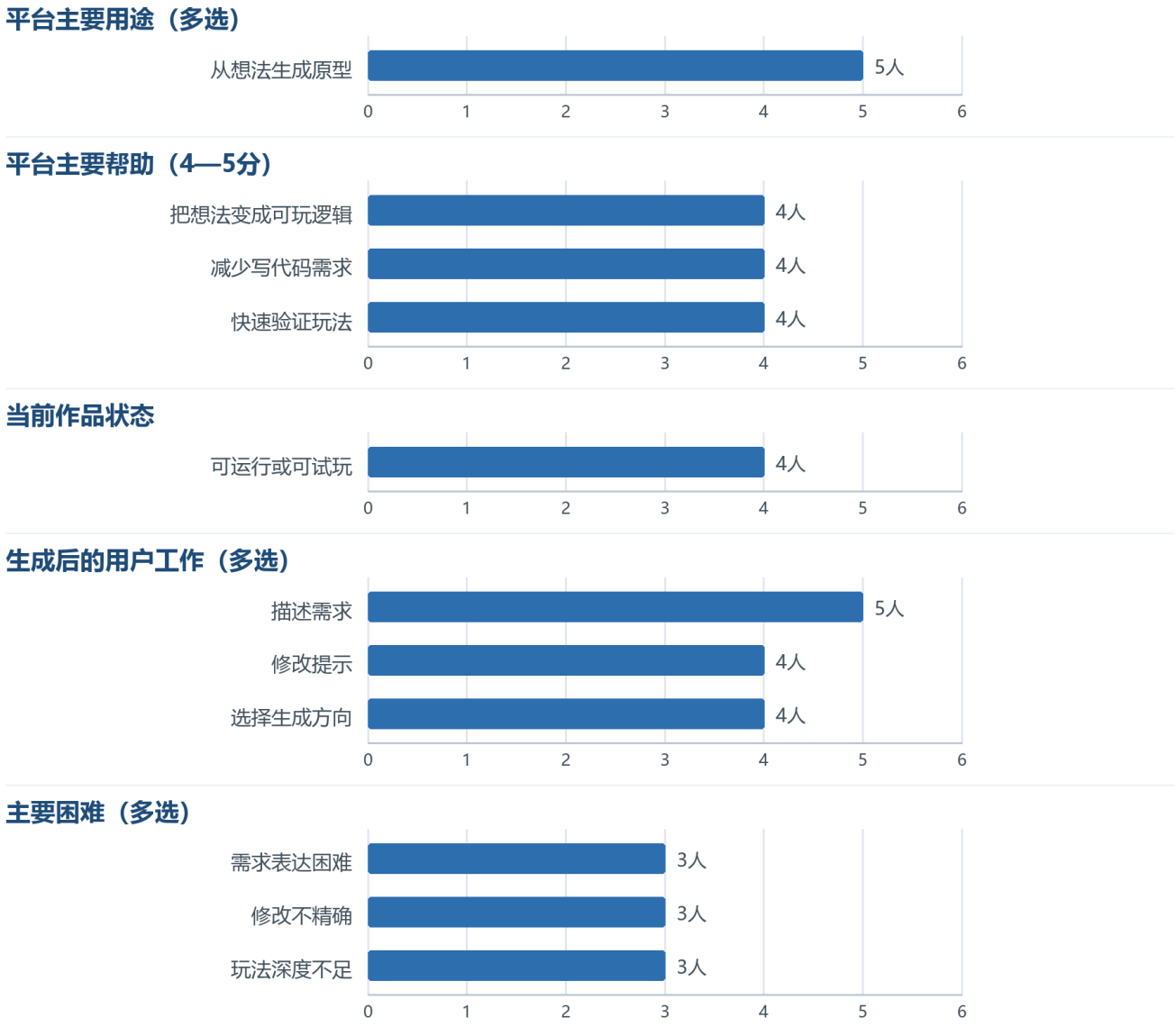

图5-1　一站式平台的主要用途、生成后工作与主要困难（N=6）

5.1.1 平台用途集中于形成可测试的初步结果

图5-1中，平台用途集中在从想法生成原型，较高评价集中在可玩逻辑形成、减少编码和快速验证。平台使用首先发生在原型形成和玩法验证阶段，用户获得的结果已经包含可供检查的作品形态和玩法逻辑。

作品状态包括已生成基础版本、可运行或可试玩以及继续完善。平台形成的初步结果已经可以查看、操作或测试，用户据此检查玩法表现。这些状态对应从基础版本生成到实际测试的不同阶段，创意已经转化为具体作品。

5.1.2 生成后的主要工作是说明需求、选择方向和修改结果

生成之后，用户工作集中在继续描述需求、修改提示和选择生成方向。用户把创作目标写成具体要求，并对生成结果作出保留、修改或放弃的选择。表中的需求说明、方向选择、结果筛选和玩法测试，覆盖了生成后的主要操作。

问卷记录的主要困难集中在需求表达、修改精度和玩法深度，分别对应目标说明、结果调整和玩法完善。这些工作分布在初步结果形成后的调整阶段：用户先说明希望实现的内容，再检查结果能否准确变化，并继续处理玩法逻辑。

5.2 生成结果在持续说明和测试中逐步接近创作目标

《炉釜猪》是一项使用TapTap制造完成的个人玩法实验，采用被动攻击式的类雷电玩法。记录提交时，玩法和系统已经形成，作品处于美术完善阶段。这份记录呈现了从整体方案输入到可试玩结果的调整过程。

5.2.1 首轮结果需要补充限制条件和目标效果

创作者输入整体方案后，taptap制造平台首轮生成的结果在玩法方向上明显偏离了预期——呈现更接近已有市场中的常见玩法，而非实验性玩法设想。创作者因此需要逐项说明哪些地方不是他想要的（排除项），以及实验性玩法的目标效果应该是什么样：

“……呈现出来的结果明显偏向市面上的玩法，必须得详细的跟他说明哪些地方是我不要的，我要做到什么样的效果，才能实现我的想法。”
——大学生问卷开放回答，样本31305bf4

5.2.2 多轮说明、筛选和测试贯穿可试玩结果形成过程

记录显示，《炉釜猪》经历7轮及以上调整，过程包含需求说明、玩法拆分、方向选择、结果筛选、提示修改和玩法测试。这些动作分别作用于方案输入、生成结果选择和实际玩法检查。

记录提交时，作品已达到可运行或可试玩状态，进入美术完善阶段。平台提供生成和修改结果，用户持续检查玩法、筛选结果并提出修改要求。可试玩结果由多轮生成、选择和测试共同构成。

5.3 平台支持创作启动，用户继续推进作品完善

5.3.1 平台把创意推进到实际测试阶段，但首轮结果可能偏离设想，方向调整仍需用户判断

一站式平台生成包含玩法逻辑的初步结果，使创意进入查看、操作和测试阶段。创作者围绕实际结果检查玩法方向，并继续修改作品。平台的直接作用体现在创意转化为可以实际检验的作品形态。

但首轮结果的方向可能偏离创作设想——例如本记录中生成的结果更接近已有常见玩法。创作者需要先判断偏差，再补充排除项和目标效果，才能把平台输出逐步推向实验性的玩法方向。

5.3.2 可试玩状态与后续修改问题可以同时存在

问卷同时记录了可运行或可试玩状态与修改不精确、玩法深度不足等困难。《炉釜猪》在可试玩状态下继续完善美术。可运行或可试玩之后，作品仍有准确修改、玩法深化和表现统一等工作，作品状态和后续工作在同一记录中同时出现。

5.3.3 后续完善需要准确修改、玩法测试和表现统一

平台需要继续支持局部修改、玩法测试、素材替换与统一，以及多轮调整。用户根据测试结果继续修改作品，作品方向和质量判断由创作者完成。

5.4 补充观察：可运行与已发布记录仍包含持续调整

5.4.1 实现方式调整后仍有玩法深化问题

一份造化工坊开放记录显示，创作者制作电子桌宠时，先考虑原生H5，随后因表现力不足转向Electron。作品已经形成可运行或可试玩结果，创作者同时认为玩法逻辑较弱。这份记录同时呈现实现方式调整和玩法深化问题。

5.4.2 已发布个案仍包含测试、素材处理和精确修改

一份非学生TapTap制造记录中的作品已经发布。该记录同时包含需求描述、玩法测试、素材替换、质量判断和修改不精确问题。已发布作品记录中仍可见测试、素材处理和精确修改等创作工作。

第六章 消费端发现：大学生玩家对AI参与作品的评价

本章回答大学生玩家如何感知AI参与、是否在意AI参与、依据什么评价作品，以及哪些条件影响继续游玩、推荐、关注、反馈和付费意愿。

问卷主线为6名相关作品实际体验者（N=6），其答卷均写明体验的平台或作品。访谈中，开发者B与该组重叠，玩家D为独立直接体验个案，开发者A、C、E用于补充观察。69份大学生问卷中，44人回答未体验原因题，这44人均回答了AI标注态度题；AI标注态度题共有63人有效作答。

6.1 问卷发现：AI参与认知、作品评价与支持意愿

6.1.1 事先信息和视觉内容构成AI参与认知

6名相关作品实际体验者中，5人在体验前已经明确知道作品有AI参与，1人在体验后得知。5人的AI参与认知在游玩前已经形成。

6份开放回答在描述AI参与感受时均提到画面、画风、图像或美术，其中1人还提到剧情语句模板化和人物对话缺少自然的情绪逻辑（样本95a19a61、695daf6e、8b3f7094、860630e2、0af2e92e、d23e0029）。事先信息和视觉、文本内容共同构成了这组体验者的AI参与认知。

6.1.2 玩法清晰较易获得认可，反馈与视听品质评价较低

玩法清晰的评价相对集中，目标反馈、美术质量和视觉统一性得到的较高评价较少。

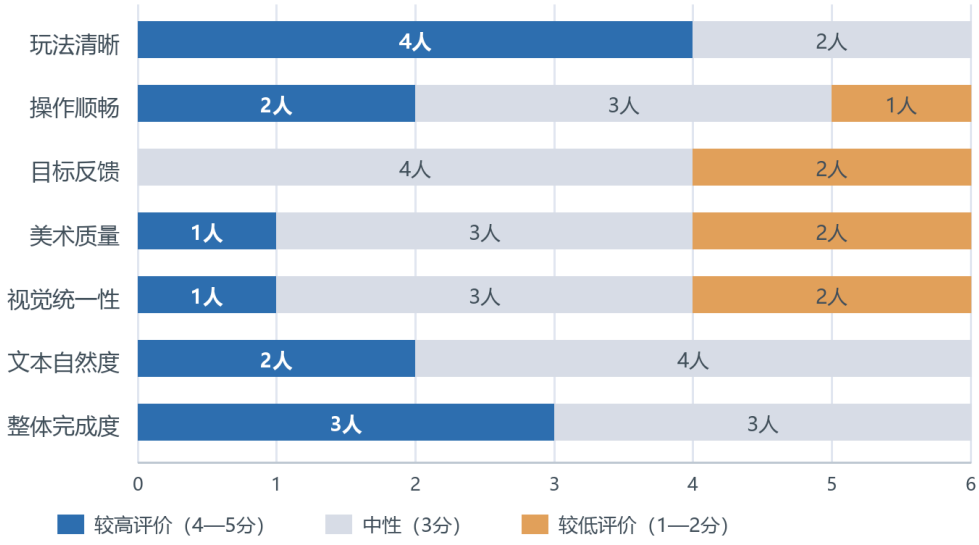

图6-1　相关作品实际体验者对主要体验维度的评价（N=6）

玩法清晰有4人给4—5分，低分为0；目标反馈无人给4—5分，美术质量和视觉统一性各有2人给1—2分。体验者能够理解基础玩法，主要问题集中在操作、目标反馈、美术细节和视觉统一。开放回答进一步记录了文本模板化、人物对话生硬和画面细节粗糙。量表与开放回答共同呈现“基础玩法成立、局部完成质量不足”的评价。一名体验者具体指出：

“部分剧情语句句式模板化，人物对话缺少自然的情绪逻辑。”
——大学生问卷开放回答，样本860630e2

开放回答记录了局部文本问题。量表中，文本自然度和整体完成度的6份回答全部位于3—5分。前者记录具体问题，后者记录整体评价。

6.1.3 AI参与评价呈现三种方向，支持意愿按行为分别形成

知道或感受到AI参与后，4人表示评价不变，1人更愿意尝试，1人略减分。6份回答呈现评价不变、更愿尝试和略减分三种方向。

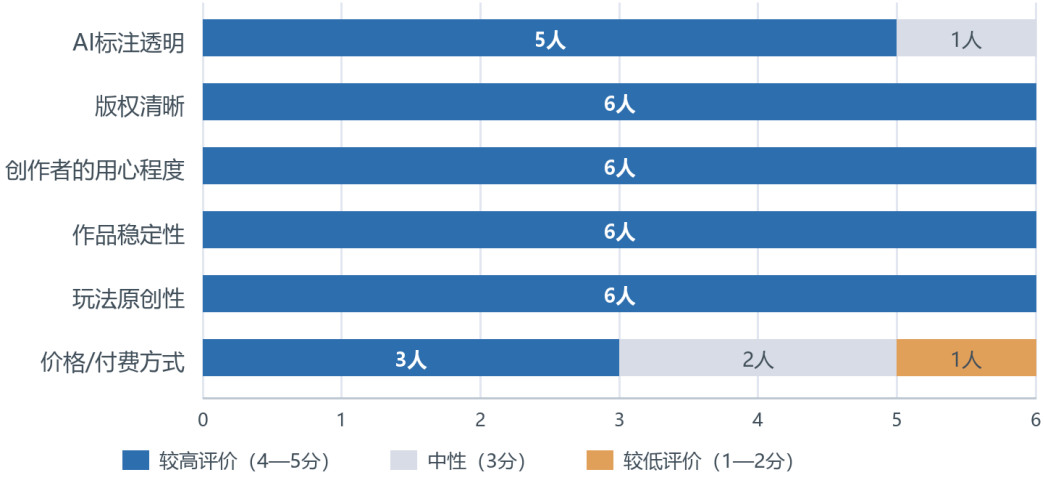

图6-2　影响信任与接受的条件（相关作品实际体验者，N=6）

版权清晰、创作者用心、作品稳定性和玩法原创性均由6人给4-5分，AI标注透明有5人给4-5分、1人给3分。价格与付费方式则有3人给4-5分、2人给3分、1人给1-2分。质量、版权和创作者投入相关条件的判断较为集中，价格判断因人而异。

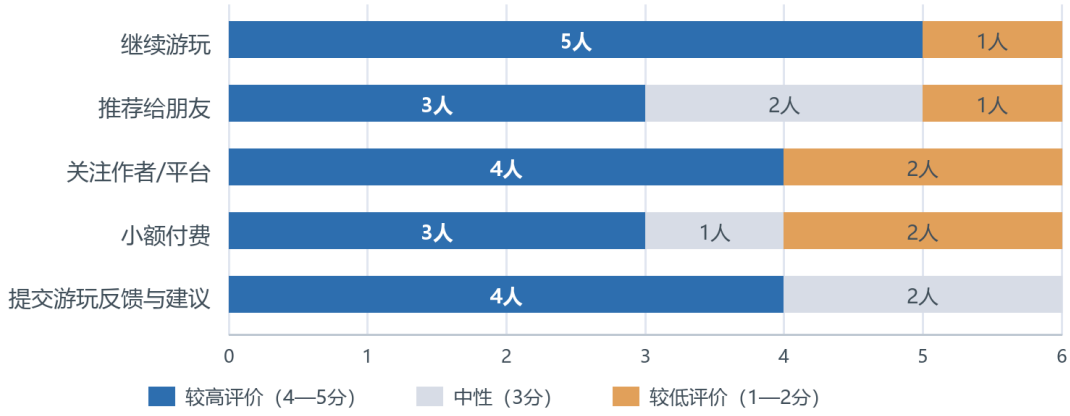

图6-3　对各类支持行为的意愿评价（相关作品实际体验者，N=6）

继续游玩有5人给4-5分，推荐和小额付费各有3人，关注和提交反馈各有4人。其中一名体验者对继续游玩给2分，对小额付费给4分，对提交反馈给5分。

具体到各个样本，样本0af2e92e认为纯AI美术容易被察觉，对AI参与略减分；其美术质量和视觉统一性均为2分，音效/音乐为1分，继续游玩和关注均为2分，但小额付费为4分、提交反馈为5分；样本695daf6e能够从画风感知AI参与，但对AI参与的评价不变；其耐玩性为1分，推荐为2分，关注和小额付费均为1分；样本860630e2指出画面粗糙和文本模板化，同时选择“更愿意尝试”，小额付费为5分。

部分答卷同时出现AI美术减分和视觉质量低分，部分答卷对AI参与评价不变，但在耐玩性、推荐和付费上给出低分；具体质量问题也会与尝试兴趣和付费意愿同时出现。

6.2 访谈发现：AI参与作品的价值判断

五名访谈对象分别具有开发者、玩家或双重经历。开发者B属于6.1的相关作品实际体验者（N=6），玩家D是独立直接体验个案，两者构成本节的主要体验材料；开发者A、C、E用于补充具体判断。

表6-4 访谈对象及其相关经历

访谈对象	相关经历	本章关注的判断
开发者A	48小时Game Jam开发与线下试玩	代码参与与美术、音效参与的感知差别
开发者B	Game Jam开发；AI角色扮演与酒馆产品体验	AI参与环节、作品评价与模型、价格判断
开发者C	Godot游戏开发与玩家体验	AI美术质量及其对付费意愿的影响
玩家D	线下完整体验TapTap制造活动作品	玩法、操作、完成度与创作者控制

访谈对象	相关经历	本章关注的判断
开发者E	Mod开发与玩家体验	代码与艺术内容的差别、游玩价值和创作主导

6.2.1 美术、文本和音乐比代码更容易被注意

具有开发者与玩家双重经历的开发者B认为，玩家更容易注意美术和文案，较少注意代码环节的AI参与。开发者A、E也把代码与美术、音效或音乐分开判断；开发者A认为，在作品质量相同的情况下，程序实现方式对玩家体验的影响较小。三名访谈对象都把美术、文案、音效或音乐放在代码之前讨论。玩家直接接触画面、文字、音乐和操作结果，代码实现过程较少进入评价。AI参与因此更多通过作品呈现被注意，美术和文本的完成质量也更容易影响整体判断。

玩家D通过TapTap制造专场参与体验，在游玩前已经知道作品有AI参与；体验时，他直接评价玩法、操作和视觉表现。该个案同时记录了活动招募、事先知情和游玩评价。

6.2.2 玩法成立是基础，生成内容仍需创作者筛选和统一

玩法是否成立先于AI参与方式。玩家D首先肯定多人拍照聚会游戏的玩法创意和操作理解，也指出键位仍需优化；开发者B同样把好玩、数据平衡和美术表现作为评价重点。

“游戏的本质还是好玩……应该是开发者主导AI的审美。”
——开发者B访谈

生成内容是否经过创作者筛选和统一，直接进入质量判断。玩家D强调主创需要确定美术基调和整体情绪，再约束、筛选和后处理生成素材；开发者C指出画面节奏平均、局部细节用力过度；开发者E强调人应保持艺术创作主导。

“如何呈现作品，这个是主创团队需要思考的问题。”
——玩家D访谈

这些判断把质量问题指向生成后的处理。主创需要确定美术基调，筛选素材，修正局部细节并统一风格；相关工作不足时，玩家直接面对画面节奏、局部细节和整体协调问题。生成效率提高以后，素材筛选与统一仍然决定最终呈现。

6.2.3 继续体验看玩法与流程，付费还看完成度、模型质量与价格

玩家D表示，玩法类型符合兴趣、流程完整可玩时，他愿意继续体验；付费则要看作品完成度和具体游戏质量。开发者B的付费判断集中在AI角色扮演产品的模型质量、内容开放性和价格结构：模型较弱而计费较高时，付费意愿会降低。

开发者C因明显的AI画面问题降低付费意愿，开发者E把游玩价值和人的艺术创作主导作为付费条件。继续体验首先看作品是否可玩并符合兴趣，付费还看完成度、模型质量、创作主导和价格匹配。

6.3 综合结论：作品质量是玩家评价AI参与作品的中心

外部桌面研究记录Steam市场反馈，本研究问卷和访谈记录大学生实际体验后的评价。三类核心问题及其依据如下。

桌面研究引用的Game Oracle对2025年1—10月Steam平台9879款游戏的研究显示，声明使用AI的游戏首月评论数量平均比未使用AI的游戏低52.6%；两类游戏的评论中位数为4条和7条，零评论比例接近20%和15.2%。在评论数不少于100条的游戏中，两类游戏的好评率中位数为84.6%和88.3%。这组外部数据记录Steam市场反馈，本研究N=6问卷记录大学生实际体验后的评价。

表6-5 第六章核心问题、前文主要依据与综合结论

核心问题	前文主要依据	综合结论
玩家如何认识AI参与	相关作品实际体验者中，5人事先知道、1人体验后知道；6份开放回答均提到画面、画风、图像或美术；访谈补充代码与美术、文本和音乐在注意程度上的差别	事先信息和视觉、文本、音乐内容共同构成AI参与认知。
市场反馈与体验后评价	外部桌面研究：AI作品的评论量、零评论比例和好评率等市场反馈较弱；本研究材料：6名体验者中4人评价不变、1人更愿尝试、1人略减分，图6-1呈现玩法、反馈、视觉和完成度评价	外部研究中的AI作品市场反馈较弱；本研究实际体验者的评价集中在玩法、反馈、视觉和完成度。
支持意愿如何形成	图6-2显示质量、版权和创作者投入条件较集中，价格为3人较高、2人中等、1人较低；图6-3显示五类意愿并不同步；访谈区分继续体验与付费条件	继续游玩、推荐、关注、反馈和付费分别形成；继续体验主要看玩法与流程，付费还看完成度、模型质量和价格匹配。

玩家从接触作品走向继续游玩、推荐和付费，要经过作品发现、实际质量判断和价值判断三个环节。AI参与呈现加分、不变和减分三种方向；阻碍后续支持的具体问题集中在反馈、美术与文本完成度、素材统一、模型质量和价格匹配。

6.3.1 玩家评价与质量筛选构成平台突破口

平台的产品天花板在作品质量。生成效率提高以后，创作者仍需完成玩法调整、目标反馈、美术统一、文本修改和整体打磨。平台的突破口是围绕玩家遇到的三类问题补充功能：作品发布、搜索和精选推荐帮助玩家找到作品；实际游玩评价、分项评分和AI参与说明帮助玩家判断作品；更新记录和创作者回应把具体问题交给创作者继续修改。评价功能负责发现和整理问题，作品质量由创作者通过实际修改提升。

6.4 补充观察：首次体验障碍与AI标注态度

以下结果均来自69份大学生问卷。未体验原因题有44人有效作答，这44人均回答了AI标注态度题；AI标注态度题共有63人有效作答，其中还包括另外19人。

6.4.1 知名度、发现渠道与质量担忧限制首次体验

在回答未体验原因的44名大学生中，17人（38.6%）没听说过相关作品，13人（29.5%）担心质量差，10人（22.7%）找不到作品。知名度、质量担忧和获取渠道是最集中的首次体验障碍。

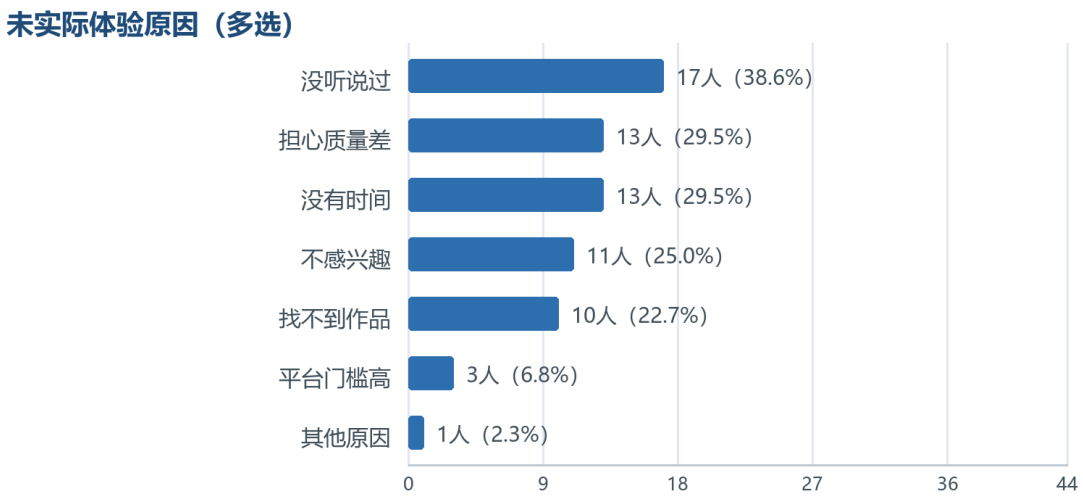

图6-4　未实际体验一站式AI平台作品的原因（N=44，多选）

注：该题为多选题，各项占比之和超过100%。

6.4.2 AI标注要求与体验后评价指向不同问题

在有效回答AI参与标注题的63名大学生中，30人（47.6%）认为必须明确标注，17人（27.0%）认为重要环节应当标注，合计47人（74.6%）希望作品至少说明关键AI参与情况。较广泛样本对信息透明有明确要求。

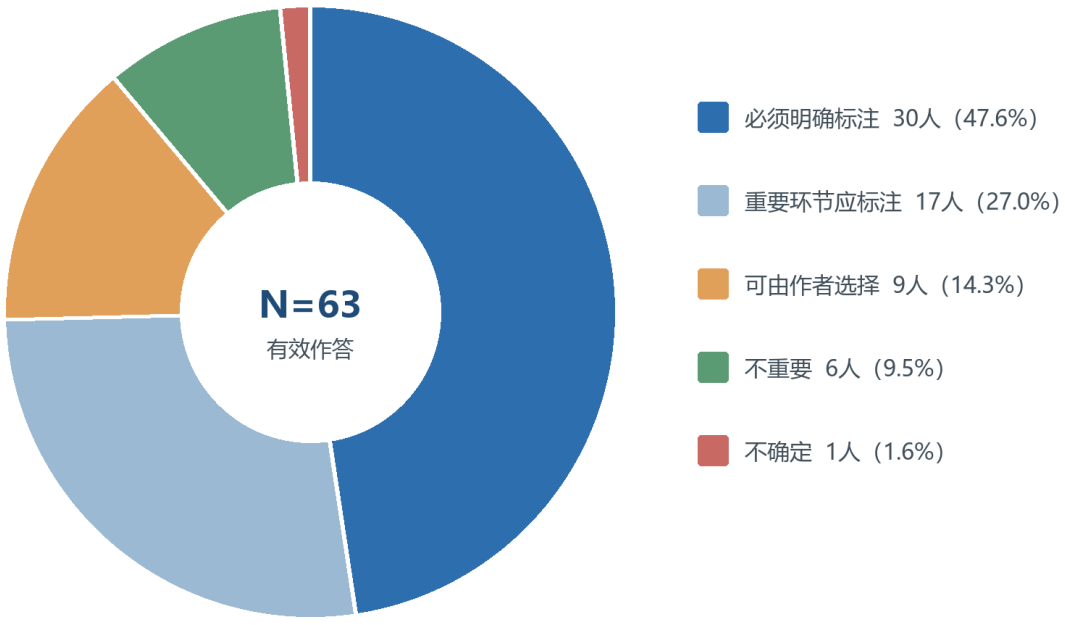

图6-5　大学生对AI参与标注的态度（N=63）

6名相关作品实际体验者中，5人对AI标注透明给4-5分；知道或感受到AI参与后，4人表示评价不变。N=6呈现实际体验后的评价，N=63呈现大学生对AI标注的态度。

AI标注用于交代作品在哪些环节使用AI，实际游玩评价用于呈现玩法、美术、文本和完成度。平台需要同时提供参与说明和作品评价。

第七章 综合结论

本章依据第四至第六章已有结论，分别回答大学生开发者使用Codex、一站式平台创作和相关作品实际体验三个核心问题，并归纳三部分材料共同指向的判断。

7.1 Codex在大学生游戏程序工作流中的使用结论

当前Codex核心样本使用Codex的主要诉求，是更快形成可继续检查和修改的程序结果，并减少重复劳动。提效与能力补足可以出现在不同任务中：开发者在熟悉任务中压缩实现时间，在不熟悉或暂时缺少思路的任务中先形成原型或方案，再继续判断和处理。

这一使用方式贯穿目标与范围设定、代码生成或修改、阅读和运行检查、反馈调整与项目接入。开发者先把创作目标整理为具体任务，并根据项目周期和已有结构限定处理范围；获得结果后，再通过阅读、运行、逻辑检查或试玩发现偏差，提出修改要求并决定哪些内容进入项目。

当前Codex核心样本在基础功能、原型形成和重复编写等环节推进较顺利，受阻环节主要涉及要求遗漏、实现方式偏离项目习惯、界面或效果不符合预期、运行错误，以及复杂任务需要补充参考材料。程序结果继续进入项目，仍需经过任务拆分、方案确认、运行验收和项目接入等人工处理。

7.2 以TapTap制造为核心锚点的一站式平台创作结论

以TapTap制造为案例锚点的一站式平台主要支持创作启动、原型形成和玩法验证。平台可以把想法推进为能够查看、操作和测试的初步结果，降低创作启动阶段的编码负担；作品进入可运行或可试玩状态后，仍需持续修改才能接近创作目标。

这组一站式平台创作记录呈现了从整体设想、首轮结果，到补充排除项和目标效果，再到多轮筛选、修改和玩法测试的过程。创作者需要根据实际结果继续说明需求、选择方向、保留或放弃生成内容，并通过测试判断玩法是否成立。创作推进依赖生成与人工选择、检查和修改的连续配合。

主要受阻条件集中在需求难以准确表达、局部修改不够精确、玩法深度不足，以及素材和逻辑难以统一。初步结果解决的是创意能否进入实际检验的问题，作品完成还需要继续处理玩法、操作反馈、素材和整体表现。

7.3 实际体验过一站式AI平台作品的大学生玩家评价

在经平台或作品线索、实际游玩深度和答卷一致性复核后纳入的6名大学生实际体验者中，AI参与本身没有形成一致的加分或减分。评价仍主要围绕玩法是否清晰、操作与目标反馈、视觉与文本表现以及整体完成质量形成。

这些实际体验者主要通过事先信息和作品直接呈现的内容理解AI参与，美术、画面和文本比程序实现方式更容易进入评价。玩家首先判断玩法能否成立，再结合操作、反馈、视听品质、文本自然度和整体统一评价作品。AI参与只有通过这些具体呈现进入体验，作品质量仍是评价的主要依据。

知道或感受到AI参与后，实际体验者对作品的评价方向并不一致，继续游玩、推荐、关注和提交反馈也分别形成选择。

7.4 三部分材料的共同判断

AI参与需要先形成可以检查、修改或体验的具体结果，才能进入人的判断。当前Codex核心样本关注程序结果能否检查、修改和接入项目；这组以TapTap制造为案例锚点的一站式平台创作记录关注想法能否转化为可测试作品；本次相关作品实际体验者依据作品呈现判断玩法、反馈、视觉、文本和完成质量。

三部分中人的工作处在不同位置，但共同指向结果形成之后的判断和处理。Codex使用者设定任务范围、检查运行并决定项目接入；一站式平台创作者说明需求、选择方向、筛选结果和测试玩法；实际体验者根据作品质量决定是否继续游玩、推荐、关注或反馈。初步结果提供了继续处理的对象，最终质量仍由具体任务中的检查、修改和选择形成。

前期预设据此作出三点修正：对Codex的判断不再在“提效”与“能力补足”之间二选一，而是确认二者随任务并存，核心是快速形成可继续处理的结果；对一站式平台的判断由“一体化流程直接推进作品完成”修正为“主要支持创作启动，完成仍需持续修改”；对消费端的判断由“AI标签形成稳定负面预期”修正为“反应方向不一，评价仍主要围绕作品质量形成”。

第八章 研究限制

8.1 样本来源与规模

本研究的样本主要从开发者群组和光核共创营内部招募，社交媒体渠道带来的样本很少，属于非概率样本；主要定量样本为69名大学生，其中Codex主要使用者15名，一站式平台创作样本和相关作品实际体验者各6名。这些结果不能推断大学生群体中的比例，也不能支持不同专业、年级、开发经验或使用背景之间的稳定比较。

8.2 一站式平台材料范围

第五章的6份一站式平台创作记录包括TapTap制造、造化工坊等多个平台，且没有一站式平台创作者访谈。这些材料不能形成TapTap制造的产品专属结论，不能比较不同平台的差异，也不能完整解释平台创作者的连续创作过程。

8.3 对照条件限制

消费端没有设置同一作品在“标注AI参与”和“不标注AI参与”条件下的对照，也没有设置“AI参与”和“非AI参与”的作品对照。现有评价不能用于判断AI标注或AI参与本身是否造成评价差异。

8.4 作品过程追踪限制

本研究没有连续追踪同一作品从创作、修改、发布到玩家评价的过程，生产端和消费端使用的也是不同样本。现有材料只能比较问题层面的共同点，不能建立创作选择、作品版本与玩家反馈之间的因果对应关系。

8.5 访谈材料的使用边界

5名访谈对象只用于解释需求拆分、检查修改、素材处理和作品评价等具体过程，不承担群体比例推断。即使多名访谈对象提出相近判断，也只能说明该现象在当前材料中重复出现。