研究报告

AI 时代的视觉思维
及其教育与行业应用研究

Visual Thinking in the AI Era
and Its Educational and Professional Applications

中英文对照图文版
Bilingual illustrated edition

认知基础 研究证据 技术实现 行业应用 课程设计
Cognition Evidence Technology Applications Curriculum

研究更新日期 2026 年 10 月 17 日
Research cutoff 17 October 2026

参考三份课程与视觉思维附件 并结合英文学术研究及官方技术资料
Informed by three supplied documents English research and primary technical sources

摘要

视觉思维通过图像、空间、关系与可操作表征支持理解、推理和协作。生成式 AI 降低图文制作成本，也增加无依据结构和有说服力错误的风险。本报告结合认知理论、绘图学习与多重表征研究、视觉 AI 技术论文和官方文档，提出以问题为起点、以证据约束语义、以计算与测试验证的工作框架。研究证据支持主动建构与整合的重要性，但不支持把学生固定为视觉型或语言型后简单匹配教学。技术可行性不能直接等同于学习成效，AI 代画与 AI 反馈需要分别评价。报告提供酒店运营、营销、文旅金融、科研与设计案例，提出 AI 导论课程地图、完整课时、Skill 职责、评价实验与实施路线。主要结论是将视觉思维作为可训练的 AI 素养，使人能够解释图、修改图、核查图，并在新任务中独立判断。

Abstract

Visual thinking supports understanding, reasoning, and collaboration through images, spatial relations, and manipulable representations. Generative AI reduces production effort while increasing the risk of unsupported structures and persuasive errors. Drawing on cognitive theory, learning research, visual AI papers, and official documentation, this report proposes a question-led framework in which evidence constrains semantics, and computation and testing check outputs. The evidence supports active construction and integration but does not justify permanent learner categories followed by simple instructional matching. Technical feasibility is distinct from learning effectiveness, and direct AI generation should be evaluated separately from AI feedback. Applications cover hospitality, marketing, tourism finance, research, and design, alongside a curriculum map, lesson plan, skill responsibilities, experiments, and implementation. The central objective is trainable AI literacy in which people can explain, revise, and verify representations and exercise independent judgment in new tasks.

关键词 视觉思维 视觉素养 生成式 AI 图解推理 多重表征 AI 导论 人机协作

Keywords visual thinking visual literacy generative AI diagrammatic reasoning multiple representations introductory AI human AI collaboration

阅读指南

Reading guide

本报告按小节先中文后英文对应。图表采用共同编号，关键标签与图题对照。研究结果使用文献编号，合成情境在图注中标明，课程与架构建议属于本报告的综合设计。第 1 至 6 章讨论概念与证据，第 7 至 9 章讨论实现与协作，第 10 至 16 章讨论行业课程与治理，第 17 至 24 章提供模板延伸应用计算示例与行动建议。

Chinese and corresponding English appear within each subsection. Figures and tables share identifiers and bilingual labels. Numbered sources support research claims synthetic scenarios are identified and proposed designs remain recommendations. Chapters 1 to 6 address concepts and evidence 7 to 9 implementation and collaboration 10 to 16 industry education and governance and 17 to 24 templates further applications computation and actions.

目录

Contents

摘要 Abstract 2

阅读指南 Reading guide 3

目录 Contents 4

1 研究问题与主要判断 Research questions and principal conclusions 6

2 视觉思维的概念与理论谱系 Concepts and intellectual traditions of visual thinking 11

3 认知机制与个体差异 Cognitive mechanisms and individual differences 16

4 学习效果的证据与边界 Evidence and boundary conditions for learning 21

5 视觉表征的语法与设计 The grammar and design of visual representations 26

6 AI 时代的价值与能力重构 Value and capability development in the AI era 31

7 AI 实现的技术路线 Technical routes for AI supported visual thinking 36

8 可信的生成与验证架构 An architecture for trustworthy generation and validation 41

9 人与 AI 共同思考的工作方法 A working method for thinking with AI 46

10 酒店运营与服务设计案例 Hotel operations and service design cases 50

11 市场营销与文旅金融案例 Marketing and tourism finance cases 55

12 AI 导论课程的整体重构 Redesigning an introductory AI course 60

13 课堂实施与完整课时方案 Classroom implementation and a complete lesson design 65

14 Skill 与专家协作体系 Skills and expert role collaboration 70

15 评价实验与研究设计 Evaluation experiments and research design 75

16 风险治理与实施路线 Risk management and implementation 80

17 可复用任务与提示模板 Reusable tasks and prompting templates 85

18 术语与前沿研究议程 Terminology and a research agenda 90

19 科研阅读与知识发现应用 Applications in research reading and knowledge discovery 95

20 设计实践与组织协作应用 Design practice and organizational collaboration 100

21 部署 复现与日常使用手册 Deployment reproducibility and daily use 104

22 综合案例研讨与练习 Integrated case discussions and exercises 108

23 交互计算与可视推演 Interactive computation and visual simulation 112

24 综合讨论与优先行动 Integrated discussion and priority actions 116
参考文献与延伸阅读 References and further reading 120
附件来源与对应说明 Supplied materials and their role 125
图表索引 List of figures and tables 126

1 研究问题与主要判断
Research questions and principal conclusions

1 1 为什么在 AI 时代重新研究视觉思维
Why revisit visual thinking in the AI era

生成式人工智能使文字、图像、图表和演示材料的生产成本显著下降，但材料增加并不自然带来理解增加。面对一份自动生成的酒店经营报告，管理者仍要判断收入和利润是否使用相同口径，客流下降究竟是季节因素还是渠道变化，以及建议的动作由谁执行。面对一张漂亮的智能体架构图，学生仍要说明箭头传递什么，哪些节点读取真实资料，哪里需要人工审批。视觉思维的研究价值，就在于帮助人们把这些隐藏在流畅表达中的结构变得可以观察、比较、操作和质疑。

本报告把视觉思维界定为利用内部意象与外部视觉表征，选择信息、组织关系、推演变化、检验解释并协调行动的一组认知实践。它既包括头脑中的空间想象，也包括纸面草图、概念图、流程图、数据图、交互模型和多人共享的白板。视觉思维的核心单位可以是对象、位置、连接、顺序、尺度、边界和变化，而不是某一种特定软件或某一种画风。视觉表达是过程中的一个环节，理解和推理才是最终要评价的结果。

AI 带来的变化具有双向性。一方面，它能够把口述问题转成初步结构，把草图变成可编辑图形，把表格转成图表，把不同方案并排呈现，并支持读者以自然语言追问。另一方面，它也能快速制造缺少证据的关系、没有数据基础的曲线以及错误却具有说服力的流程。因此，AI 时代的视觉思维应同时发展建构能力和审查能力：既会提出可视化需求，也会识别一张图是否值得相信。这是本报告贯穿理论、技术与课程设计的基本立场。

Generative artificial intelligence has reduced the cost of producing text, pictures, charts, and presentations. An increase in available material, however, does not automatically produce an increase in understanding. A manager reading an automatically generated hotel report must still ask whether revenue and profit use consistent definitions, whether lower demand reflects seasonality or a change in distribution channels, and who is responsible for carrying out a proposed action. A student looking at an attractive agent architecture must still explain what each arrow carries, which component reads actual evidence, and where human authorization is required. Visual thinking matters because it can make structures hidden behind fluent expression available for inspection, comparison, manipulation, and criticism.

This report defines visual thinking as a family of cognitive practices that use internal imagery and external visual representations to select information, organize relationships, simulate change, test explanations, and coordinate action. It includes mental spatial imagery as well as sketches, concept maps, process diagrams, statistical charts, interactive models, and shared whiteboards. Its working elements may be objects, positions, connections, sequences, scales, boundaries, or changes. It is therefore broader than a particular software application or an attractive graphic style. Visual expression is one part of the process; understanding and reasoning are the outcomes that ultimately require evaluation.

AI changes this practice in two directions. It can transform a spoken concern into an initial structure, convert a sketch into an editable diagram, propose charts for a table, place alternative solutions side by side, and let

a reader ask follow-up questions in ordinary language. It can also produce unsupported relationships, curves with no empirical basis, and persuasive diagrams that encode incorrect procedures. Visual thinking in the AI era must consequently develop both construction and scrutiny. People need to specify useful representations and assess whether those representations deserve confidence. This position connects the theoretical, technical, educational, and professional sections of the report.

1 2 六项可以落实的主要判断
Six practical conclusions

第一，视觉思维适合作为跨任务能力来培养。课程可以要求学生在文字、图、表和计算之间转换，却不宜预先把他们划为不可改变的视觉型或语言型。第二，视觉材料应由任务决定。比较大小、解释因果、规划流程和探索空间，需要不同的表征。第三，AI 适合承担候选生成、格式转换和重复绘制，人应保留目标定义、证据判断和后果承担。第四，精确图表应优先由结构化数据和可执行规范产生；文生图适合情境插画与外观探索，不能替代数量关系的计算。

第五，教学中的成功标准应是学生是否能解释、修改和迁移一张图，而不只是教师是否做出了精美课件。第六，自动化系统必须保留来源、版本、验证记录和人工责任。图表越容易生成，这些看似附属的记录就越重要。上述判断是本报告结合研究证据与应用分析提出的综合建议，并非声称某一篇研究已经检验了全部体系。

这六项判断共同指向一个操作原则：先确定需要回答的问题，再选择表征，随后调用 AI 生成候选，最后通过数据、规则和人的解释验证候选。一个优秀的成品，应该允许读者追问其来源、拆开其组成部分、改变其关键假设，并说明在什么条件下它会失效。报告把这样的能力称为可检验的视觉推理，以区别于仅仅增加图像数量的做法。

Six practical conclusions organize the report. First, visual thinking should be cultivated as a capability that travels across tasks. A course can require movement between prose, diagrams, tables, and calculations without assigning learners to permanent visual or verbal categories. Second, the representation should follow the question. Comparing magnitudes, explaining causation, planning a procedure, and exploring space require different visual forms. Third, AI is useful for generating candidates, translating formats, and carrying out repeated drawing operations, while people retain responsibility for defining objectives, assessing evidence, and accepting consequences. Fourth, precise graphics should usually be generated from structured data and executable specifications. Image generation can support contextual illustration and appearance exploration, but it cannot replace calculation of quantitative relationships.

Fifth, educational success should be measured by whether learners can explain, revise, and transfer a representation, rather than only by whether the instructor has produced polished slides. Sixth, automated systems need sources, versions, validation records, and clearly assigned human responsibility. As graphics become easier to produce, these supporting records become more valuable. These conclusions are a synthesis of research and application analysis, rather than a claim that one study has tested the entire proposed system.

Together, the six conclusions imply an operational sequence: identify the question, choose a representation, use AI to propose candidates, and validate those candidates with data, rules, and human explanation. A useful final artifact should allow its audience to ask where it came from, inspect its components, alter an

important assumption, and identify the conditions under which it would fail. The report calls this capability inspectable visual reasoning. The phrase names a proposed design objective; it does not refer to a standardized psychological test or an established commercial product.

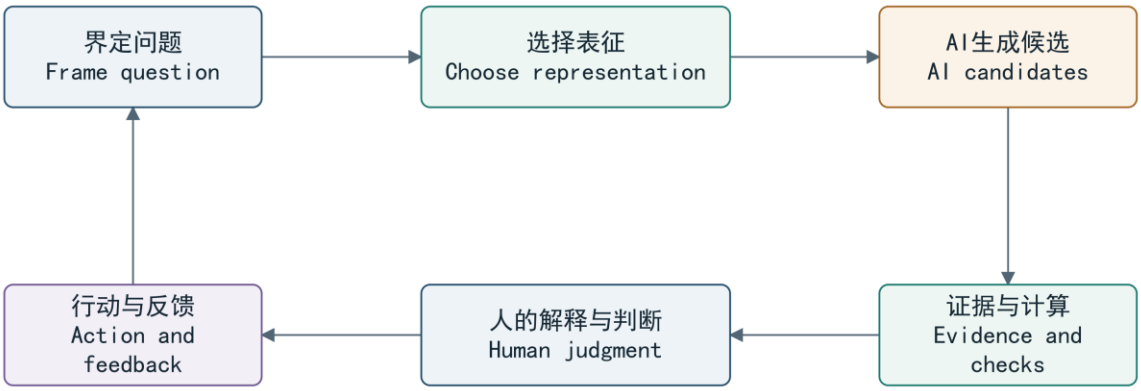


图 1 可检验的视觉推理闭环
Figure 1 A cycle of inspectable visual reasoning
本报告提出的概念与设计示意 Conceptual design proposed in this report

1 3 研究方法和证据的使用口径
Methods and the use of evidence

本研究采用问题导向的英文资料检索与批判性叙述综合。检索围绕视觉认知、图解推理、绘图学习、多重表征、可视分析、生成式 AI、多模态推理与人机协作展开，使用 Consensus 和 SciSpace 发现学术线索，再通过期刊、会议论文集、作者机构、官方技术文档等核对关键记录。检索更新至 2026 年 10 月 17 日。本报告不是具有穷尽检索和双人独立筛选程序的系统综述，不提供虚构的筛选流程图，也不把搜索引擎返回数量当作证据强度。

资料分为六类。综合研究与元分析用于判断较稳定的总体趋势；实验研究用于理解特定条件下的效果；认知理论用于解释机制；技术论文与研究原型用于判断实现的可行性；官方文档用于确认技术接口和规范；用户附件用于理解课程背景与已有设想。技术系统能够生成某种图，不代表它已经提高学生的迁移能力；调查中有人报告减少思考，也不等于研究已经证明长期能力下降。报告在涉及这些跨层推断时会明确说明其性质。

原始研究的效应量、样本量与模型成绩仅在必要位置引用。效应量不换算为“提高多少百分比”，不同任务、不同基线和不同年份的模型成绩不混在同一排名中。部分文献以摘要或正式记录为可核查依据，因此只支持该记录明确陈述的结论。英文技术资料的时效性较强，产品接口和支持格式可能变化；报告侧重可迁移的架构原则，并把特定能力视为需要在部署时重新验证的条件。

The study uses question-led searches of English-language materials followed by critical narrative synthesis. Search themes include visual cognition, diagrammatic reasoning, drawing to learn, multiple representations, visual analytics, generative AI, multimodal reasoning, and human–AI collaboration. Consensus and SciSpace were used to discover academic leads. Important records were then checked against journals, conference proceedings, author institutions, and official technical documentation. The search cutoff is 17 October 2026. This is not an exhaustive systematic review with independent dual screening. It therefore does not present an invented screening flowchart or interpret search-result counts as measures of evidential strength.

Six source categories serve different purposes. Reviews and meta-analyses inform relatively stable overall patterns. Experiments clarify effects under specified conditions. Cognitive theories supply explanations. Technical papers and prototypes establish particular forms of feasibility. Official documentation describes interfaces and specifications. The supplied documents explain the course context and the user's existing ideas. A system's ability to generate a chart does not establish that it improves student transfer. A survey respondent's report of reduced effort does not by itself demonstrate a long-term loss of ability. Such distinctions are maintained wherever the report moves between kinds of evidence.

Sample sizes, effect sizes, and benchmark results are reported only where they help a substantive judgment. Standardized effect sizes are not converted into percentage improvements. Scores from different tasks, baselines, or model generations are not combined into a single ranking. Some papers are supported by accessible abstracts or formal records rather than a full-text audit; the report uses them only for conclusions explicitly available in those records. Technical interfaces are also time-sensitive. The emphasis is therefore on transferable architecture and evaluation principles, with product-specific capabilities treated as conditions to recheck during deployment.

1 4 附件的贡献与需要修订之处
Contributions and revisions to the attachments

三份附件形成了一条清晰的发展路径：从葛兰汀的视觉思维启示，转向 AI 导论的整体设计，再转向 Skill 与专家协作的生产流程。这条路径的重要贡献，是把图从课件装饰提升为组织知识、拆解问题和设计行动的工具。报告保留课程地图、工具卡、工作流、学生快画与行业任务等思路，同时补上证据核查、表征语法、失败路径和学习评价。

附件中的若干判断需要降格为待验证设想。例如“结构图覆盖 90%需求”“十几分钟完成一套课时包”“全程无需干预”缺少适用范围、时间记录和质量基线；“所有概念都做卡片”“所有逻辑都用 Mermaid”也可能造成形式僵化。报告将其改写为可测试的设计选项。附件中的旧对话、旧篇幅要求和安装说明属于资料背景，不构成本次执行指令；本报告的任务是研究与论证，而不是安装技能或自动发布课程。

The three supplied documents describe a coherent progression: inspiration from Temple Grandin, redesign of an introductory AI course, and a production workflow involving skills and expert roles. Their principal contribution is to treat graphics as tools for organizing knowledge, decomposing problems, and designing action. The report retains course maps, capability cards, workflows, student sketching, and industry tasks, while adding evidence checks, representational conventions, failure paths, and measures of learning.

Several statements in the attachments are better treated as hypotheses to test. Claims that structural diagrams cover ninety percent of requirements, that a complete teaching package takes only minutes, or that no intervention is necessary do not specify a scope, timing record, or quality baseline. Universal rules requiring a card for every concept or a single diagram language for every relationship may also make design inflexible. They are therefore revised into testable options. Historical dialogue, earlier length requests, and installation notes remain background material. They do not become instructions for the present task, which is to investigate and develop a research report.

表 1 附件观点的保留与修订
Table 1 Retaining and revising attachment proposals

附件设想 Proposal	可保留价值 Value	修订要求 Revision
全课程视觉设计 Visual design across the course	知识与活动建立共同结构 Connect knowledge and activities	每幅图有问题与学习目标 Give each figure a question and objective
对象与空间两类 Object and spatial modes	提醒注意不同表征需求 Recognize representational differences	不作为固定学生分类 Do not assign permanent learner types
工具角色卡 Capability cards	说明输入输出与边界 Clarify inputs outputs and limits	复杂概念允许展开 Permit expansion for complex concepts
Mermaid 优先 Prefer Mermaid	支持可编辑结构图 Support editable structural diagrams	按语义选择渲染器 Choose a renderer from semantics
专家团生产 Expert role production	职责能够分开检查 Make responsibilities inspectable	增加独立审查与返工 Add independent review and rework
十分钟自动完成 Completion in minutes	提出效率优化方向 Suggest an efficiency objective	记录全部成本后再判断 Measure total cost before claiming gains

本报告的综合分析与设计建议 Report synthesis and design recommendations

2 视觉思维的概念与理论谱系
Concepts and intellectual traditions of visual thinking

2 1 从视觉知觉到视觉推理
From visual perception to visual reasoning

视觉思维并不是最近因 AI 而出现的概念。Arnheim 在 1969 年的《Visual Thinking》中强调知觉与思考之间的紧密联系，挑战把看见当成被动输入、把推理完全归于语言的划分。[1]这一理论传统提醒我们，分类、组织和抽象也可以发生在视觉活动之中。不过，哲学和心理学意义上的总体主张，不能直接转换成“任何课程加图都会更好”的教学结论。课程设计仍要说明图提供了什么信息，以及学习者进行了什么操作。

可以用三个层次理解视觉活动。第一层是辨认，例如识别酒店平面图中的出口。第二层是理解关系，例如判断前台、客房和消防通道之间的距离与连接。第三层是推理，例如在高峰时段改变人员配置后，预测排队和响应速度如何变化。许多 AI 系统擅长给图片命名，却未必能够稳定完成后两层任务。把图像识别、结构理解和情境推演分开，是评估 AI 视觉能力的必要起点。

外部图也不只是内部想法的复制。有时人只有把几个节点画出来，才发现分类标准不一致；只有把一条流程拉长，才看到重复审批；只有把两个时间尺度放在一起，才发现短期转化增长伴随长期复购下降。在这些情形中，图参与了想法的形成。研究视觉思维因而应关注“人如何与表征互动”，而不仅是“最后生成了什么图片”。

Visual thinking did not originate with contemporary AI. In Visual Thinking, first published in 1969, Rudolf Arnheim argued for a close connection between perception and thought, challenging a division in which seeing is passive input while reasoning belongs exclusively to language.[1] This tradition draws attention to classification, organization, and abstraction within visual activity. A broad theoretical argument, however, does not establish that adding images improves every course. Instructional design must still identify the information supplied by a representation and the operations performed by its learners.

Three levels help distinguish visual activities. Recognition identifies an exit symbol on a hotel plan. Relational understanding establishes how reception, rooms, and escape routes connect. Reasoning predicts how a change in staffing might affect queues and response times during a busy period. An AI system may label the visible objects convincingly while remaining unreliable on the latter two tasks. Separating image recognition, structural interpretation, and situated simulation is therefore a necessary starting point for evaluating visual capabilities.

An external representation is also more than a copy of a fully formed internal idea. Drawing several nodes may reveal inconsistent categories. Extending a process may expose repeated authorization. Placing two time scales together may show that short-term conversion growth accompanies a decline in repeat business. In such cases, the representation participates in the formation of the idea. Research should consequently examine how people interact with representations, as well as the appearance of the final image.

2 2 葛兰汀的启示及其适用范围

Grandin and the scope of her contribution

Grandin 在 2022 年的《Visual Thinking》中，以个人经验、职业实践和研究讨论强调思维方式的多样性，并区分偏重对象细节的视觉想象与偏重模式和空间关系的思维。[2]这对教育的启发，是扩大能够展示能力的任务形式，让制图、建模、观察、设计和动手解决问题获得充分位置。报告不把这本大众著作当作随机实验，也不以一位杰出人物的经历推断所有孤独症人士的能力、需求或成长路径。

附件把两种模式简化为“建造火车”和“让火车动起来”，这一比喻适合引发讨论，却不宜被当成严格分类标准。对象想象并不只处理静态外观，空间推理也不自动等于因果推理；理解动力传递需要专业知识、力学关系和经验约束。一个擅长画出设备的人可能同时擅长推演结构，一个偏爱语言的人也能通过训练掌握流程图。教育应用应把比喻转换成具体的任务要求，避免从偏好直接推断能力。

对中国教育的反思，也应从宏大定性转为可观察问题。某门课程是否只奖励长篇文字？学生是否能用模型和图说明理解？实验、设计与改错是否进入评价？这些问题可以通过作业、课堂记录和评分规则检验。将“成绩一般的学生可能存在未被看见的优势”作为开放假设有教育价值；直接断言大量低分学生都是视觉天才，或者把某种认知方式当作创新的唯一来源，都超出了现有依据。

In Visual Thinking, published in 2022, Temple Grandin combines lived experience, professional practice, and research discussion to advocate recognition of different ways of thinking. She distinguishes imagery concerned with detailed objects from thinking concerned with patterns and spatial relationships.[2] The educational implication is to widen the tasks through which competence can become visible, including drawing, modeling, observation, design, and practical problem solving. The book is not treated here as a randomized experiment, and one accomplished person's experience cannot determine the capabilities, support needs, or developmental paths of all autistic people.

The attachments summarize the two modes through the analogy of building a train and making a train operate. This can initiate discussion, but it should not function as a strict classification rule. Object imagery is not restricted to static appearance, and spatial reasoning is not automatically causal reasoning. Understanding power transmission requires domain knowledge, mechanical relationships, and experience. A person who draws an apparatus well may also reason effectively about its operation. A learner who prefers language can acquire diagrammatic competence through practice. Educational design should translate an analogy into observable task requirements rather than infer ability directly from preference.

Reflection on Chinese education also benefits from moving from general diagnosis to questions that can be examined. Does a course reward only extended prose? Can learners demonstrate understanding through models and diagrams? Do practical work, design, and error correction contribute to assessment? Assignments, classroom records, and marking criteria can answer such questions. The possibility that lower-scoring students possess overlooked strengths is a useful educational hypothesis. Claims that large numbers of them are necessarily visual geniuses, or that one cognitive mode is the exclusive source of innovation, go beyond the available basis.

2 3 心像 外部表征与视觉素养

Imagery external representation and visual literacy

心像指没有相应外部刺激时对视觉或空间内容的内部表征。外部表征则是可以共享和操作的图、物件、符号与界面。二者可能互相支持，但并不相同：一个人即使难以产生鲜明心像，也可能借助纸笔、坐标、关系图和计算工具完成出色推理。因此，课堂不应要求所有学生报告相同的“脑内画面”，而应允许多种到达理解的路径。

视觉素养强调阅读、解释、评价与制作视觉信息的能力。在 AI 环境中，它还包括识别生成痕迹、检查来源与判断呈现方式是否误导。可视化则通常指把数据或结构映射到视觉形式的设计与计算过程。视觉思维可以使用可视化，但它的范围更广：前者关注人在理解中做了什么，后者常关注信息被怎样编码。保持这些区分，有助于避免把软件操作熟练误认为思维能力提高。

视觉分析把交互可视化与分析过程结合，使人能够提出问题、调整筛选、比较结果并形成解释。图解推理关注图形结构如何支持推断。设计思维关注问题发现、原型与迭代，其过程经常运用视觉思维，但两者并非同义。知识图谱是一种具有类型化实体和关系的数据结构，也不能等同于任何带节点连线的知识地图。教学中应允许入门比喻，同时逐步建立这些边界。

Mental imagery concerns internal visual or spatial representation without corresponding immediate sensory input. External representations are diagrams, objects, symbols, and interfaces that can be shared and manipulated. They may support one another, but they are not identical. Someone who does not experience vivid imagery may still reason effectively with paper, coordinates, relational diagrams, and computational tools. A classroom should therefore avoid requiring every student to report the same inner picture. It should support several routes to understanding.

Visual literacy concerns the ability to read, interpret, evaluate, and create visual information. In an AI environment, it additionally includes examining provenance and judging whether presentation choices mislead. Visualization usually denotes the design or computation that maps information into visual form. Visual thinking can use visualization, but it asks a broader question about what a person does in order to understand. Visualization often emphasizes encoding, while visual thinking emphasizes cognitive activity. Maintaining the distinction prevents fluency with software from being mistaken for an improvement in reasoning.

Visual analytics combines interactive visualization with analysis so that people can ask questions, change filters, compare outcomes, and develop explanations. Diagrammatic reasoning concerns inference supported by graphical structure. Design thinking concerns framing problems, prototyping, and iteration; it frequently uses visual thinking but is not synonymous with it. A knowledge graph is a data structure with typed entities and relations, rather than any map containing connected nodes. Instruction may begin with accessible analogies while progressively clarifying these boundaries.

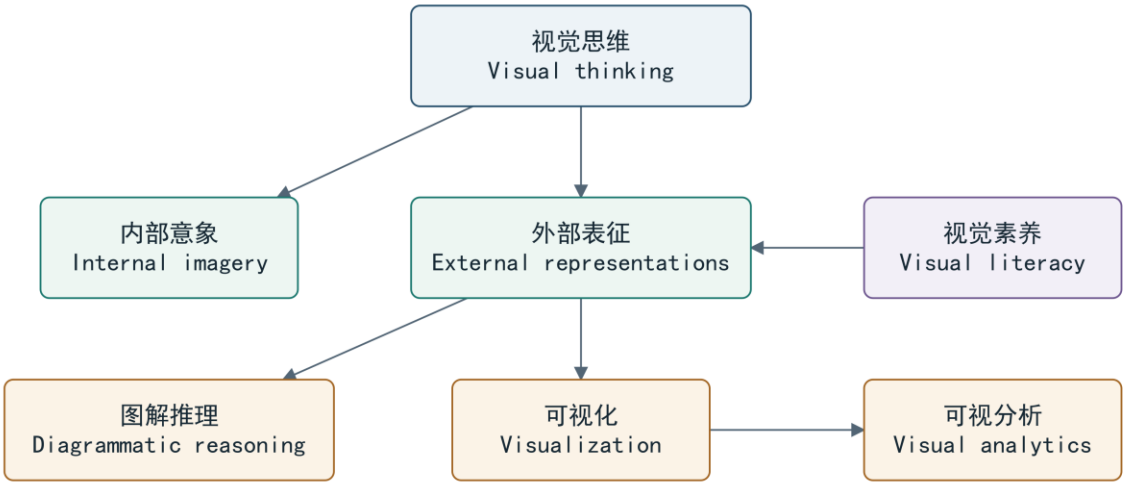


图 2 相关概念的边界与联系
Figure 2 Related concepts and their boundaries
本报告提出的概念与设计示意 Conceptual design proposed in this report

2 4 一项任务需要怎样的表征
Choosing representations for a task

表征的选择可以从问题动词开始。要列举，用清单；要比较，用对照表或共用尺度的图；要说明先后，用流程；要说明谁与谁有关，用关系网络；要检验原因，用带假设说明的因果图；要寻找空间冲突，用平面图；要观察参数变化，用交互模型。一个任务可能需要多种表征，但每一种都应承担明确职责。没有职责的图，通常只是在增加观看成本。

同一酒店问题可以由不同表征提出不同问题。问题树回答“可能有哪些因素”，时间序列回答“变化何时发生”，渠道分组图回答“变化集中在哪里”，流程图回答“信息和责任如何流动”，收益曲线回答“在某个假设下如何权衡”。这些图不能互相替代。尤其不能由一张原因清单直接宣布原因成立，也不能由一条拟合曲线直接决定真实价格。报告后续案例将反复展示表征之间的互相约束。

Representation selection can begin with the verb in the question. Listing may require a checklist. Comparison may require a table or charts on a common scale. Sequence calls for a process representation. Connections call for a network. Causal inquiry needs explicit assumptions. Spatial conflicts may need a plan. Parameter changes may require an interactive model. A task can use several forms, but each form should have a defined analytical responsibility. A graphic without such a responsibility usually adds viewing effort without adding understanding.

Different representations of the same hotel problem ask different questions. A problem tree identifies possible factors. A time series shows when a change occurred. A channel comparison shows where it was concentrated. A workflow shows how information and responsibility move. A revenue curve explores trade-

offs under stated assumptions. These forms are not interchangeable. A list of possible causes cannot establish causation, and a fitted curve cannot independently justify a real pricing decision. Later cases show how representations can constrain and check one another.

表 2 常见表征与问题匹配
Table 2 Matching representations to questions

表征 Form	适合的问题 Suitable question	主要局限 Main limitation
问题树 Problem tree	可能因素怎样分类 How might factors be organized	分支不能证明因果 Branches do not establish causation
概念图 Concept map	概念间有何明确关系 How concepts relate	连线必须命名 Edges need meanings
流程与泳道 Process and swimlane	谁在何条件下做什么 Who acts under which conditions	需要例外与权限 Exceptions and authority are necessary
数据图 Data chart	大小分布变化怎样比较 Compare magnitude distribution and change	口径与尺度决定解释 Definitions and scales constrain interpretation
空间图 Spatial plan	位置距离冲突在哪里 Locate distances and conflicts	比例与尺寸要可靠 Scale and dimensions need accuracy
交互模型 Interactive model	假设变化后结果怎样 What follows changed assumptions	模型内结果不是实证 Model behavior is not empirical confirmation

本报告的综合分析与设计建议 Report synthesis and design recommendations

3 认知机制与个体差异
Cognitive mechanisms and individual differences

3 1 把工作记忆中的关系放到外部
Externalizing relationships from working memory

当一个人需要同时记住酒店前台、客房、收益管理和客户服务之间的多个交接点时，认知负担不只来自节点数量，还来自关系的保持与更新。将关系放到外部，可以让人通过位置和标记重新找到信息，减少反复在记忆中搜索的需要。Larkin 与 Simon 的经典分析指出，信息相同的文字和图解，可能因为搜索与推断操作不同而具有不同的计算效率。[3]这一机制说明图可能怎样帮助推理，并不保证所有画出的图都更有效。

在课程中，这可以转化为一个小实验：先以文字叙述一段包含六次交接的客诉流程，让学生指出哪一步可能遗漏责任；随后提供节点与箭头相同信息的流程图，再比较他们寻找断点的过程。教师应观察学生怎样定位、回看和解释，而不是仅询问哪一种“更好看”。如果流程本来只有两个动作，文字可能已经足够；如果图上塞入二十种颜色，外化也可能增加搜索成本。优势来自信息组织与任务的配合。

外部表征还有保存思考历史的作用。一个假设被写在图边，就可以与后来的数据相对照；一个节点被标记为待核查，就不必靠记忆提醒团队。在 AI 协作中，这种可见的暂存区尤其重要，因为模型能够快速提出大量候选，而人需要区分哪些只是提议、哪些已经验证。视觉工作区应因此允许不确定、未完成和被否定的内容存在，而不是只容纳看似完整的最终答案。

Remembering the handoffs among reception, housekeeping, revenue management, and guest service creates a burden that concerns both the number of components and the need to maintain their relationships. Moving those relationships into an external display lets people recover information through position and marking, reducing repeated memory search. Larkin and Simon's classic analysis shows how informationally equivalent prose and diagrams can differ in computational efficiency because they support different search and inference operations.[3] This explains a possible advantage of diagrams without guaranteeing an advantage for every graphic.

A classroom exercise can make the mechanism observable. Students first read a description of a complaint procedure with six handoffs and identify a possible gap in responsibility. They then inspect a diagram containing the same information and explain how they locate the gap. The instructor observes looking, revisiting, and explaining rather than simply asking which version looks nicer. Two simple actions may need only a sentence. A diagram with twenty arbitrary colors may increase search effort. The benefit depends on the match between organization and task.

External representations also preserve a history of thinking. An assumption written beside a diagram can later be compared with data. A node marked for verification does not have to remain in someone's memory. This visible temporary workspace is particularly valuable when AI supplies many candidates quickly. People need to distinguish proposals from checked claims. The workspace should therefore accommodate uncertainty, incomplete work, and rejected explanations, as well as polished answers.

3 2 从选择和组织到整合
From selection and organization to integration

多媒体学习的认知理论强调，学习者需要选择相关信息、组织语言与图像内容，并把它们同已有知识整合，而不是被动吸收屏幕。[8]这对 AI 课件生成具有直接意义：系统不能只问“这页需要什么图片”，还应问“学生要把哪两类信息联系起来”。例如，讲 RAG 时，一张检索流程图与一段酒店政策原文之间必须建立对应，否则学生可能记住流程名称，却不知道证据为什么能够约束答案。

具体设计可以让学生给每一步贴上证据标记：问题中的日期条件被保留在哪里，检索返回的段落如何进入上下文，最终回答的哪一句来自哪一条政策。教师随后故意更换一条政策，要求学生指出图中的哪些部分必须更新。这种任务把整合变成可观察活动，也暴露“图已经生成，所以理解已经完成”的错误。AI 可以提供候选对应关系，但学生需要给出理由并纠正不成立的连接。

视觉与文字之间的整合还涉及时间安排。先展示完整图，再逐句读一遍，未必能够促使学习者主动处理。可以先让学生预测缺少的节点，再展示图；或者让他们在两张不同图之间寻找不一致。对基础较弱的学生，提供部分结构和有限选择可能更合适；对已经掌握概念的学生，则应逐步减少提示。支架的作用是帮助完成有价值的思考，而不是永久代替思考。

The cognitive theory of multimedia learning emphasizes selecting relevant information, organizing verbal and pictorial material, and integrating it with prior knowledge.[8] Learners are not passive recipients of whatever appears on a screen. For AI-generated instructional materials, the design question should therefore extend beyond which image belongs on a slide. It should identify which pieces of information the learner must connect. In an explanation of retrieval-augmented generation, a retrieval diagram needs a meaningful correspondence with an actual hotel-policy excerpt. Otherwise, students may remember component names without understanding how evidence constrains an answer.

One activity asks learners to attach evidence markers to the process. Where is the date condition in the question preserved? How does a retrieved passage enter the context? Which sentence in the response is supported by which policy? The instructor then replaces one policy and asks which elements must change. Integration becomes observable, and the activity exposes the mistaken assumption that a generated diagram means completed understanding. AI can suggest correspondences, but learners must justify and correct them.

Timing also matters. Showing a complete diagram and reading it aloud may leave students passive. Alternatives include predicting a missing node before the reveal or identifying contradictions between two representations. Learners with limited prior knowledge may need a partial structure and constrained choices. More advanced learners can work with reduced prompts. Scaffolding should enable valuable cognitive work and gradually release responsibility, rather than permanently perform the work on the learner's behalf.

3 3 操作图形也是一种认识活动
Manipulation as an epistemic activity

Kirsh 与 Maglio 关于认识性行动的讨论表明，一些看似额外的外部动作可以帮助人发现解决办法，而不只是执行已经确定的办法。[17]在视觉思维中，移动节点、改变分组、画辅助线和调整

参数，都可能改变问题可被看见的方式。这与把图看成最终展示材料的观念不同。一个可交互的图，既是答案的载体，也可以是产生问题的装置。

例如，学生把酒店服务流程中的投诉处理节点移到入住前，会立即发现该节点不符合时间条件；把客房清洁和入住登记放在同一泳道，又会引出职责混淆。此时拖动本身没有证明某种流程正确，却促成了新的解释需要。设计好的 AI 工作区应该在操作之后要求说明：“这个移动改变了什么关系？依据是什么？对其他节点有什么影响？”否则交互容易退化为随意摆放。

在数量任务中，参数操作可以把抽象公式转换为可观察变化，但也容易制造错觉。一个滑块移动后曲线立即变化，使用者可能以为已经验证了真实世界的规律。实际上，他只探索了预设模型。课程需要反复区分模型内推演和模型外检验：前者说明假设成立时会怎样，后者需要新的观察数据或实验。能够指出这一区别，是 AI 时代视觉思维成熟的重要表现。

Kirsh and Maglio's account of epistemic action shows why some apparently additional external actions help people discover solutions rather than merely execute decisions already made.[17] Moving nodes, regrouping items, drawing an auxiliary line, or adjusting a parameter can change what becomes visible in a problem. An interactive representation can therefore act as a device for generating questions as well as a medium for presenting answers.

Moving a complaint-handling node to a position before arrival may expose a temporal inconsistency. Placing housekeeping and check-in in one responsibility lane may reveal an ambiguity about ownership. The movement does not prove that a procedure is correct, but it creates a need for explanation. A well-designed AI workspace should ask what relationship changed, what justifies that change, and which other elements are affected. Without such questions, interaction can become arbitrary rearrangement.

Parameter manipulation similarly makes an abstract equation observable, but it can create a false sense of validation. A curve that changes immediately after a slider moves confirms the behavior of the programmed model, not the behavior of the world. Instruction should repeatedly distinguish exploration within assumptions from testing assumptions against new observations. The former asks what follows if a model is accepted; the latter asks whether it should be accepted. Recognizing this distinction is a central component of mature visual thinking with AI.

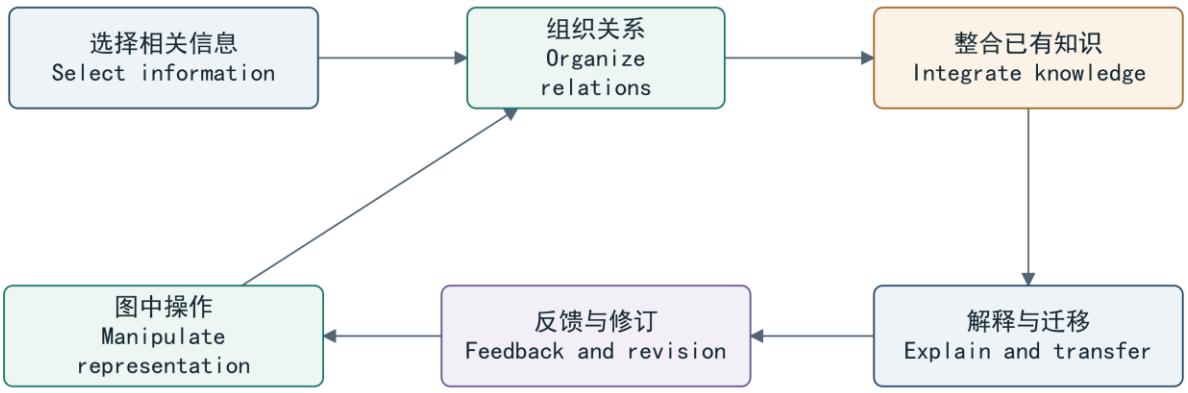


图 3 外部表征如何支持理解
Figure 3 How external representations can support understanding
依据认知与生成学习理论绘制 [3][8][17] 箭头为教学概念示意 Conceptual synthesis not a measured causal model

3 4 差异应该怎样进入教学

Bringing individual differences into instruction

对象、空间和语言加工的研究，支持在具体维度上讨论个体差异。[4][5]但偏好、能力和学习效果必须分别测量。喜欢图片不等于能够准确解释统计图；自认为擅长空间想象，也不等于熟悉一个酒店系统的业务规则。Pashler 等人的证据审查没有支持把测得的学习风格与教学形式匹配便能普遍提高学习的主张。[6]因此，课程可以承认差异，同时拒绝把差异固定为教学命运。

一种更稳妥的做法，是提供共同目标、多种进入方式和共同检验标准。学生可以先用口述、草图或清单提出方案，但最终都要说明关键关系、证据与限制。空间技能训练的元分析说明，相关能力具有可训练性，不能简单视作天赋标签。[7]报告据此建议把视觉表征能力纳入持续训练，而不是在开学时用问卷决定谁适合看图、谁只适合文字。

对需要支持的学习者，重点是具体障碍。例如颜色辨认、细小文字、复杂布局、工作记忆负荷和操作设备可能造成不同困难。适配应针对障碍提供替代文本、放大、逐步展开、键盘操作或额外时间。孤独症、注意力差异或心像差异都不能单独决定一个人应接受哪种图。教师需要与学习者共同观察任务表现，并在不降低核心目标的前提下调整进入方式。

Research on object, spatial, and verbal processing supports discussion of individual differences along specific dimensions.[4][5] Preferences, abilities, and learning outcomes nevertheless require separate measurement. Liking pictures does not establish competence in interpreting a statistical chart. Feeling confident in spatial imagery does not imply knowledge of hotel operating rules. Pashler and colleagues' evidence review did not establish a general learning benefit from matching instructional form to assessed learning style.[6] A course can recognize differences without turning them into fixed educational destinies.

A more defensible design offers a common objective, several entry routes, and shared criteria for success. Learners may begin by speaking, sketching, or listing, but all must eventually explain relationships, evidence, and limitations. A meta-analysis of spatial-skill training supports the view that these skills can develop rather than simply function as inherited labels.[7] Visual representational competence should therefore receive sustained practice, instead of a beginning-of-course questionnaire deciding who should use pictures and who should use prose.

Support should address concrete barriers. Color discrimination, small print, complicated layouts, working-memory demands, and interaction devices can create different problems. Appropriate responses may include alternative text, enlargement, progressive disclosure, keyboard access, or additional time. Autism, attention differences, or imagery differences alone cannot determine the right representation for an individual. The instructor and learner should examine actual task performance and adapt access while preserving the central learning objective.

表 3 个体差异与教学响应
Table 3 Responding to individual differences

观察维度 Dimension	应收集证据 Evidence	可采取支持 Support
表征知识 Representation knowledge	能否读懂符号与连线 Interpret symbols and edges	示范与有限词汇 Model conventions and limit vocabulary
领域基础 Domain knowledge	能否解释业务规则 Explain operating rules	案例与来源对照 Connect cases with sources
空间任务经验 Spatial experience	完成具体空间任务 Perform defined spatial tasks	逐步训练与反馈 Progressive practice and feedback
语言熟悉度 Language familiarity	术语与条件理解 Understand terms and conditions	双语词典与简明表达 Glossary and plain expression
感知与操作条件 Perception and interaction	能否看清并操作 See and operate the material	文本替代与键盘路径 Text alternatives and keyboard access
自信与判断 Confidence and judgment	确定程度是否匹配正确性 Compare certainty and accuracy	独立检查与反例 Independent checks and counterexamples

本报告的综合分析与设计建议 Report synthesis and design recommendations

4 学习效果的证据与边界
Evidence and boundary conditions for learning

4 1 研究支持什么程度的结论
The strength of conclusions supported by research

有关视觉学习的证据应按任务和比较条件阅读。Cromley 等人的元分析纳入 53 项研究、8111 名参与者，报告总体标准化效应量为 g 等于 0.69；事实、推论和迁移结果的估计分别为 0.85、0.44 和 0.22。[12]这些数值反映纳入比较的综合差异，不能解释成学习提高 69%，更不能当作 AI 代画图的预期收益。不同结果的差异提醒课程设计者，记住内容与把原理迁移到新情境，是不同难度的目标。

本报告据此提出一项评价要求：课程项目不能只问学生能否复述图中节点，还应给一个更换条件的新问题。学过酒店 RAG 流程后，可以把任务换成退款期限冲突，要求学生指出知识来源、日期条件和人工升级如何调整。能够照抄原图只能说明再现能力，能够说明修改理由才更接近迁移。图表作业因而需要原图、修改稿与解释记录，而不宜只收最终图片。

比较条件也决定研究结论的含义。如果画图组优于只读文字组，这并不能说明画图优于优秀的图文讲解、检索练习或有反馈的自我解释。时间投入不同还可能影响结果。选择教学策略时，需要同时看学习收益、用时与实施成本。某种活动即使有效，如果占去整节课却没有增进关键目标，也未必适合当前课程。这是把研究证据转化为教学判断时必须保留的权衡。

Evidence on visual learning should be read in relation to the task and comparison condition. Cromley and colleagues synthesized fifty-three studies involving 8,111 participants and reported an overall standardized effect of g = 0.69. Estimates for factual, inferential, and transfer outcomes were 0.85, 0.44, and 0.22 respectively.[12] These values summarize differences among the included comparisons. They do not mean a sixty-nine percent improvement, and they are not estimates of the benefit of having AI draw on a student's behalf. The differences among outcomes also show why remembering material and applying a principle in a new setting require separate assessment.

The report consequently recommends that course projects include a changed-condition problem. After studying a hotel retrieval workflow, students might encounter conflicting refund deadlines and have to revise the evidence source, date checks, and escalation route. Copying the original diagram demonstrates reproduction. Explaining why the changed conditions require a revision provides stronger evidence of transfer. Assessment should therefore collect an initial representation, a revision, and a reasoned explanation rather than only a final image.

Comparison conditions also constrain interpretation. An advantage over reading alone does not establish superiority over well-designed multimedia instruction, retrieval practice, or feedback-supported explanation. Unequal learning time may contribute to differences. Teaching decisions must consider learning, time, and implementation cost together. An activity can be beneficial yet still be unsuitable if it consumes an entire lesson without advancing its central objective. This trade-off should remain visible when research is translated into practice.

4 2 支架的价值在于帮助整合
Scaffolding for integration

Leutner 与 Biele 在 2025 年的元分析中综合了 14 项研究、16 个独立比较、1213 名参与者。总体绘图支持并没有自动改善结果，而能促进整合的支持与较好的理解相关，相关分组估计为 g 等于 0.35。[13]这一结果使“给模板就有效”的直觉需要修正。支架的价值不只在于减少画图劳动，也在于帮助学习者把图、文字和已有知识联系起来。该数值来自特定比较，不能扩展为任何 AI 支架的固定效果。

对 AI 导论课程而言，半成品图不应只是留出几个空框。更有意义的空缺，是要求学生判断关系：检索失败后为什么需要追问，生成答案之前为什么必须识别资料版本，人工升级是否应该保留原始证据。AI 可以按难度生成不同空缺，但教师要检查空缺是否对应概念目标。如果空缺只考软件按钮名称，活动主要训练的是记忆界面，而非理解系统。

反馈也应针对关系和推理。与其说“图还不够清晰”，不如指出“从客户问题到退款动作之间缺少身份核验和政策依据”。与其奖励节点数量，不如要求学生解释删去一个节点会带来什么后果。课程可以把标准示例安排在学生首次尝试之后，再让他们比较、修订和说明。这样的流程保留生成活动，同时避免让错误结构在没有反馈的情况下被反复巩固。

Leutner and Biele's 2025 meta-analysis included fourteen studies, sixteen independent comparisons, and 1,213 participants. Drawing support did not automatically improve outcomes overall. Support expected to facilitate integration was associated with better comprehension, with an estimated g of 0.35 in the relevant subgroup.[13] This challenges the assumption that providing any template is sufficient. A scaffold matters not only because it reduces drawing labor, but because it helps learners connect visual material, verbal material, and prior knowledge. The estimate does not specify a fixed benefit for AI scaffolding in general.

In an introductory AI course, a partially completed diagram should therefore contain meaningful relational gaps, rather than merely empty boxes. Students could decide why failed retrieval should lead to a clarifying question, why document versions need checking before generation, or whether escalation should preserve the original evidence. AI can generate gaps at different levels of difficulty, but the instructor must verify their connection to the learning objective. A gap that tests only the name of an interface button principally assesses interface recall.

Feedback should likewise address relationships and reasoning. Instead of saying that a diagram lacks clarity, identify the missing identity check and policy evidence between a customer request and a refund action. Instead of rewarding the number of nodes, ask what would happen if one node were removed. A model solution can appear after an initial attempt, followed by comparison, revision, and explanation. This retains generative activity while reducing the risk that unsupported errors are repeatedly reinforced.

4 3 多种表征和动画的条件
Conditions for multiple representations and animation

多种表征能够互补，也可能要求学习者完成额外转换。DeFT 框架提示设计者同时考虑表征设计、所承担的功能以及学习者需要完成的任务。[9]Rau 进一步讨论了学生尚未理解内容、同时又

不理解表征的困境。[10]因此，同一页并列公式、网络图、卡通人物和仪表盘，不一定比一个清楚的关系图更有帮助。关键是学生能否说出它们表达的是同一对象的哪些方面。

2024 年关于两种以上外部表征的综述纳入 46 项研究，元分析使用 132 个效应量；其绩效综合估计较小，报告为 g 等于 0.118，收益依赖适当支持。[14]这支持审慎增加表征，而不是追求种类数量。可以先用表格明确变量，再用图显示关系，最后让学生把图中的一个位置对应到表中的记录。表征之间的连接任务，比单纯同时展示更值得设计。

动画尤其容易被误用。Tversky 等人的分析强调动画的表达和理解条件，不能仅因内容在变化就认为动态展示更好。[16]在课程中，可以允许暂停、回看和逐步控制，并保留关键静态帧供比较。讲解智能体循环时，节点依次亮起可以突出执行顺序；若亮起速度超过学生理解速度，就应改成分步操作。动画的价值应通过理解与迁移检验，而不是靠热闹程度判断。

Multiple representations may complement one another, but they also require translation work. The DeFT framework asks designers to consider design, representational functions, and the tasks learners must perform.[9] Rau describes the difficulty of learning unfamiliar content through representations that are themselves unfamiliar.[10] Placing an equation, network, cartoon, and dashboard on one screen is consequently not necessarily better than a single clear relational diagram. The important question is whether students can explain which aspects of a shared object each form expresses.

A 2024 review of more than two external representations covered forty-six studies and used 132 effect sizes in its meta-analysis. The reported performance estimate was small, g = 0.118, and benefits depended on appropriate support.[14] This supports deliberate addition rather than maximizing the number of forms. A table might first define variables, a chart show their relationship, and a linking task require students to locate the corresponding record. Designing the connection between forms deserves more attention than merely displaying them together.

Animation also requires restraint. Tversky and colleagues' analysis emphasizes conditions of expression and comprehension, rather than assuming that dynamic content necessarily benefits from dynamic display.[16] Learners can be given pause, replay, and step controls, with key static frames retained for comparison. Sequential highlighting may clarify an agent loop, but if it moves faster than learners can understand, a controlled sequence is preferable. Understanding and transfer should determine the value of animation, rather than visual activity alone.

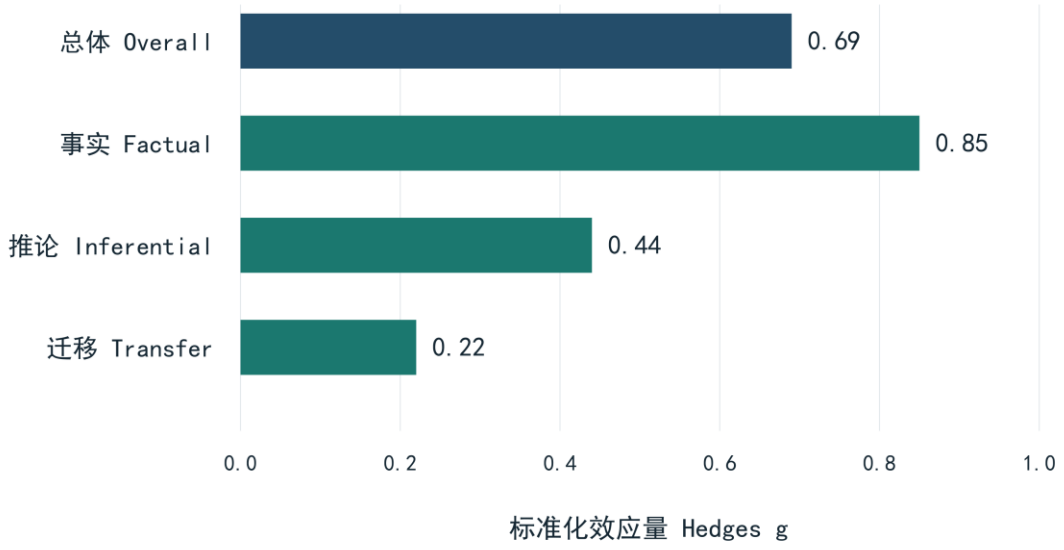


图 4 绘图学习元分析中的不同结果

Figure 4 Different outcomes in a drawing to learn meta analysis

数据来自[12] 不代表百分比 不同结果不是独立方法排名 未展示未取得的置信区间 Estimates are not percentage gains or independent method rankings

4 4 对 AI 辅助绘图的谨慎推断

Cautious inference about AI assisted drawing

传统绘图学习研究通常要求学习者亲自选择、组织和外化内容，而 AI 可以直接完成其中大量步骤。于是，“自己画图有效”不能直接推出“让 AI 画图同样有效”。这不是否定 AI，而是要求重新确定人参与的环节。如果学生只输入一句主题，拿到完成图并提交，主要的认知活动可能发生在系统中；如果学生比较两个候选、指出错误、重建关系并解释修改，AI 则可能降低制作门槛而保留思考任务。这一可能性需要专门实验。

近期绘图学习综述仍强调任务设计和边界条件的不一致性。[39]因此，本报告建议把“AI 生成图”“人修订 AI 图”“人先画再获 AI 反馈”“人使用标准图”作为不同条件分别研究。不能把所有含 AI 或含图片的教学干预归为一类。研究至少应报告先验知识、制作时间、图的准确性、反馈方式、延迟测验与新任务表现。否则，即使学生认为活动有趣，也难以判断它究竟改善了什么。

Traditional drawing-to-learn research usually requires the learner to select, organize, and externalize content. AI can perform much of that work directly. Evidence for learner-generated drawing therefore cannot simply be transferred to AI-generated drawing. The issue is to identify where human cognitive participation remains. A student who submits a completed graphic after entering a topic may have done little structural work. A student who compares candidates, identifies errors, rebuilds relationships, and explains revisions may retain substantial reasoning while reducing production effort. That possibility requires direct testing.

Recent review work continues to emphasize variation in drawing-task designs and boundary conditions.[39] The report consequently recommends treating AI-generated diagrams, human revision of AI diagrams, human-first drawing followed by AI feedback, and study of standard diagrams as distinct conditions. Interventions should report prior knowledge, production time, diagram accuracy, feedback, delayed testing, and transfer. Enjoyment alone cannot identify what has improved.

表 4 关键学习证据及解释边界
Table 4 Selected learning evidence and its limits

研究 Study	报告内容 Reported evidence	不能据此推出 Does not establish
Cromley 等 2020 [12]	53 项研究 8111 人 总体 g 0.69 53 studies 8111 participants overall g 0.69	不是提高 69% 或 AI 代画收益 Not a 69 percent gain or an AI drawing estimate
同一元分析 [12]	事实 0.85 推论 0.44 迁移 0.22 Factual 0.85 inferential 0.44 transfer 0.22	各结果不是独立随机对照排名 Outcomes are not an independent ranking
Leutner 与 Biele 2025 [13]	14 项研究 整合支持分组 g 0.35 14 studies integration-support subgroup g 0.35	不是任意模板都有效 Not support for every template
Rexigel 等 2024 [14]	46 项研究 绩效估计 g 0.118 46 studies performance estimate g 0.118	不是表征越多越好 Not an unlimited more-is-better rule
Fiorella 与 Zhang 2018 [11]	比较条件与指导影响解释 Comparators and guidance matter	不是画图总胜过优秀图文 Not universal superiority over good multimedia
Pashler 等 2008 [6]	学习风格匹配缺乏充分依据 Insufficient support for style matching	不否认差异或可访问需求 Does not deny differences or access needs

研究摘要与正式记录的选择性综合 Selected synthesis of abstracts and formal records

5 视觉表征的语法与设计
The grammar and design of visual representations

5 1 图形选择是一项建模决定
Representation choice as a modeling decision

图形的第一项决定不是配色，而是承认世界具有怎样的结构。树状图假定可以按层级展开，网络图允许多重连接，流程图突出顺序和条件，矩阵突出交叉比较，坐标图则要求可定义的变量。把一个存在反馈的酒店服务系统画成单向链条，会隐藏顾客重新提问与部门返工；把所有问题画成以 AI 为中心的辐射图，则可能掩盖业务目标与工具之间的区别。形式选择已经包含解释，不能交给审美偏好独立决定。

每次绘图之前可以完成一段表征说明：我要回答什么问题；图中对象是什么；连接代表什么；哪些信息不在图里；读者看完要做什么。比如“解释客诉处理的责任交接”比“画一张专业酒店流程图”更能约束输出。AI 若接收后一种要求，往往会补齐常见流程，让图看起来完整；前一种要求则促使系统突出交接记录、负责岗位和升级条件，便于人核查。

结构图还需要控制粒度。把“数据处理”画成一个节点，可能适合课程全景图，却不足以指导实操；把读取文件、清洗日期、删除重复记录全部展开，适合实现检查，却可能遮蔽业务主线。建议保留概览图与展开图，二者共享稳定节点编号。读者可以从全局进入细节，也能把细节放回全局。这样的层级比把所有信息挤进一张超大图更有利于持续修改。

The first decision in graphic design is a modeling decision, rather than a color decision. A tree assumes hierarchical organization. A network permits multiple connections. A process diagram emphasizes sequence and conditions. A matrix supports cross-comparison. A coordinate chart requires definable variables. Representing a service system with feedback as a one-way chain hides repeated questions and rework. Placing AI at the center of every radial map may obscure the difference between a business objective and a tool. The chosen form already contains an interpretation.

Before drawing, the author can write a short representation brief: what question is being answered, what the objects are, what a connection means, what is excluded, and what the audience should do next. Explaining responsibility handoffs in complaint handling constrains an output more effectively than requesting a professional-looking hotel diagram. The latter encourages AI to complete a familiar pattern. The former directs attention toward records, responsible roles, and escalation conditions that a human can inspect.

Granularity also needs control. A single data-processing node may suit a course overview but be inadequate for practical instruction. Expanding file loading, date cleaning, and duplicate removal may help implementation while obscuring the business story. An overview and a detailed view can share stable identifiers. Readers can then enter detail without losing its place in the whole. This supports revision more effectively than putting every detail into one enormous diagram.

5 2 节点和箭头必须有明确含义
Explicit meanings for nodes and arrows

节点可以代表岗位、工具、文件、动作、状态或概念，但不应在没有说明时混用。同样一条箭头可能表示时间顺序、数据传递、依赖、控制、因果或相似。如果图中“AI→收益提升”的箭头既没有数据也没有作用机制，它只是一个期待；如果“知识库→大模型”表示检索文本进入上下文，就应该注明传递的是资料而不是把整个数据库装进模型。准确命名连接，可以防止图形替作者作出未说明的承诺。

推荐为教学结构图采用有限词汇：读取、输出、检索、审核、触发、依赖、假设影响和人工确认。不同词汇可以通过线型或标签区分，但颜色不应成为唯一信息。关键连接需要给出来源或规则，例如“超过规定退款额度则升级人工”。当作者无法解释一条连线的意义时，应先删除或标为待确认，而不是用布局让它显得自然。

图中的边界同样重要。一个围框可以表示系统范围、组织职责或数据权限，含义不同。智能体内部的规划节点与外部支付系统之间，应画出权限和确认点。课程全景图中的“工具”与“能力”也应分开：能够点击绘图工具，不等于能够选择合理的视觉编码。学生需要学习阅读这些约定，并能够发现别人图中的语义冲突。

Nodes may represent roles, tools, files, actions, states, or concepts. These categories should not be mixed without explanation. Arrows may express sequence, data transfer, dependency, control, causation, or resemblance. An unsupported arrow from AI to increased revenue expresses an expectation. An arrow from a knowledge base to a language model should identify retrieved passages entering context, rather than imply that the entire database has become part of the model. Naming relationships prevents a diagram from making commitments that its author has not justified.

A limited instructional vocabulary is useful: reads, produces, retrieves, reviews, triggers, depends on, hypothetically influences, and requires human confirmation. Labels and line styles can distinguish these meanings, with color as a supplementary cue. Important edges should specify a rule or source, such as escalation when a refund exceeds an authorized limit. If an author cannot explain an edge, it should be removed or marked for confirmation rather than made plausible through layout.

Boundaries also carry meaning. A surrounding frame can denote system scope, organizational responsibility, or data permission. The relationship between an agent's planning component and an external payment service should make permission and confirmation visible. A course map should distinguish tools from capabilities: operating a charting application does not establish competence in choosing an encoding. Learners need to read these conventions and detect semantic conflicts in other people's diagrams.

5 3 数量图必须守住数据与尺度
Data definitions and quantitative scales

数量图的可信度来自数据定义与视觉编码的配合。Cleveland 与 McGill 关于图形知觉的研究提示，不同视觉编码对准确比较具有不同要求。[18]在实际报告中，若读者需要精确比较渠道收入，共用基线的条形图通常比大小不一的人物图标更容易检查；如果需要观察分布和异常，单一平均值又可能不够。选择图形时应首先考虑要完成的比较，而不是追求新奇。

每张数据图都应说明变量、单位、时间范围、样本或统计口径、缺失值处理和来源。百分比要说明分母，累计数要说明起点，双轴图要解释为何采用两种尺度。不得用模型生成的顺滑曲线填补没有观测的数据，再把它画成实际趋势。需要补值或预测时，应明确区分观测段、估计段和情境段，并让读者知道不确定性如何进入结果。

对教学演示数据，透明比仿真外观更重要。一个合成酒店例子可以采用简单需求函数，前提是标题和图注明确写出演示属性。简化模型的作用是让学生看见目标和假设如何影响判断，而不是假装来自真实酒店。学生应能重算至少几个数据点，改变一个参数，说明曲线为何变化。这比一张精致却无法复算的仪表盘更适合研究报告中的方法说明。

Trustworthy quantitative graphics connect data definitions with visual encodings. Cleveland and McGill's research on graphical perception shows why different encodings place different demands on accurate comparison.[18] If readers need to compare channel revenues, bars on a shared baseline are often easier to inspect than differently sized decorative icons. If they need distributions and outliers, a single mean may be inadequate. The intended comparison should guide the choice.

Each data chart should state its variables, units, time span, statistical definition, missing-data treatment, and source. Percentages need denominators, cumulative values need starting points, and dual axes need a reason for using separate scales. An AI-generated smooth curve must not fill unobserved periods and then appear as an actual trend. Observed, estimated, and scenario segments should be distinguishable, with uncertainty explained.

For teaching data, transparency matters more than realistic appearance. A synthetic hotel example can use a simple demand function if its title and note state that it is an illustration. Simplification allows learners to see how objectives and assumptions change a decision; it should not simulate the authority of actual hotel observations. Students should be able to recompute several points, change a parameter, and explain the resulting curve. This is more useful than a polished dashboard that cannot be reproduced.

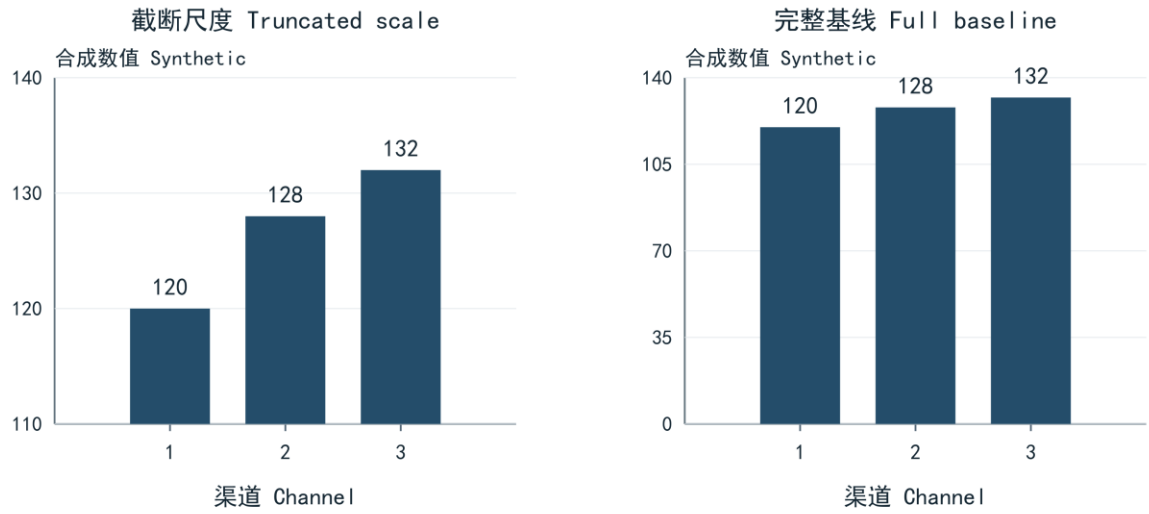


图 5 尺度改变视觉印象

Figure 5 Scale changes visual impression

合成演示数据 同一组数值 左图夸大视觉差异 截断轴并非在所有任务中错误 Synthetic demonstration with identical values

5 4 图文对照与可访问性

Bilingual alignment and accessibility

本报告采用同一小节内先中文、后英文的对照方式，而不是把两种语言塞进窄小并排栏。复杂图中的核心标签使用中英文短语，长解释放在图注和正文中；表格则对关键字段进行对照。这样既能保持对应关系，也能避免每个图框因为双语文字而失去空间。翻译的目标是保留逻辑、限制与证据强度，而不是逐字复制汉语语序。

W3C 关于复杂图像的指导强调提供能够表达关键信息的文本替代。[32]因此，图注不应只写“流程示意图”，还应概括主线、分支与例外。涉及颜色时应同时提供标签或线型，并检查必要的对比度。[33]对交互图，还要考虑键盘访问、默认状态、重置和静态替代。视觉思维的教育价值不应该以排除无法方便看图或操作图的人为代价。

This report places Chinese and corresponding English within the same subsection, instead of forcing long passages into narrow parallel columns. Core figure labels use short bilingual phrases, while longer explanations appear in captions and prose. Tables translate the fields needed for comparison. This preserves alignment without overcrowding the graphic. Translation aims to preserve logic, qualifications, and evidential strength, rather than reproduce Chinese word order.

W3C guidance for complex images emphasizes text alternatives that convey important information.[32] A caption should therefore summarize the main route, branches, and exceptions rather than merely identify a process illustration. Color should be supplemented by labels or line styles, and necessary contrast should be checked.[33] Interactive work also needs attention to keyboard access, initial states, reset controls, and

static alternatives. The educational value of visual thinking should remain accessible to people who cannot conveniently see or manipulate a graphic.

表 5 图表制作与可访问性核查
Table 5 Graphic and accessibility checks

项目 Item	验收问题 Acceptance question	修复方式 Repair
关系语法 Relation grammar	每条线表示什么 What does each edge mean	加标签或删除无依据连接 Label or remove unsupported edges
数量口径 Definitions	单位与分母是否可见 Are units and denominators visible	补字段词典与图注 Add dictionary and notes
布局层级 Hierarchy	主线是否容易找到 Can the main route be located	分层展开与稳定编号 Use layers and stable identifiers
颜色编码 Color encoding	不用颜色仍能理解吗 Is meaning available without color	增加线型与直接标签 Add line styles and direct labels
复杂图替代 Complex-image alternative	文本能否表达主要信息 Does text convey the important content	说明主线分支与限制 Describe main route branches and limits
交互访问 Interactive access	暂停重置键盘是否可用 Are pause reset and keyboard available	提供静态与替代输入 Provide static views and alternative input

本报告的综合分析与设计建议 Report synthesis and design recommendations

6 AI 时代的价值与能力重构
Value and capability development in the AI era

6 1 从信息生产转向问题表达
From producing information to framing questions

当 AI 可以大量生成材料时，稀缺能力逐步转向提出合适的问题、区分关键与次要信息，以及决定什么值得验证。视觉思维可以帮助人把一个模糊要求转成具有对象、范围和关系的问题。例如“帮助酒店提升经营”几乎不能指导研究；将其画成需求、价格、渠道、服务成本和客户体验之间的关系，就能进一步追问是短期现金流、贡献利润还是长期复购需要改善。问题被拆开以后，数据需求与价值冲突也随之出现。

这一变化并不降低语言的重要性。高质量视觉工作需要准确命名、明确边界和解释推断；语言又可借助图形检验是否自洽。AI 时代的优势不是用图取代文字，而是在人可理解的多种表征之间有目的地移动。能够把一段政策变成决策表，再把决策表变成流程，并指出各次转换丢失什么，是一种比“会让模型写报告”更稳定的能力。

在组织中，问题表达还影响协作。经营负责人、数据人员和一线员工可能使用同一词语却指向不同对象。一个共享流程图可以迫使大家说明“客户”是入住人、付款人还是会员，“订单”是预订记录还是已确认住宿。AI 可以帮助列出这些歧义，但只有相关人员共同确认，图才具备组织意义。视觉思维在这里承担的是建立共享问题空间，而非替团队作出所有判断。

When AI can produce abundant material, scarce capabilities increasingly concern asking suitable questions, distinguishing central from peripheral information, and deciding what deserves verification. Visual thinking can transform a vague request into a problem with objects, boundaries, and relationships. Helping a hotel improve performance is too broad to guide an investigation. Mapping demand, price, channels, service costs, and guest experience makes it possible to ask whether the objective concerns short-term cash flow, contribution profit, or repeat business. Data requirements and value conflicts become visible as the problem is decomposed.

Language remains essential. Effective visual work needs accurate names, explicit boundaries, and explanations of inference. Diagrams can in turn help test the consistency of language. The important capability is purposeful movement among representations. Translating a policy into a decision table, then into a process, while identifying what each transformation loses, is more durable than simply asking a model to write a report.

Problem representation also affects organizational collaboration. Managers, analysts, and frontline staff may use the same term for different things. A shared process can require them to clarify whether customer means guest, payer, or member, and whether order means a booking record or confirmed accommodation. AI can identify possible ambiguities, but the representation gains organizational meaning only through confirmation by the people concerned. Its role is to establish a shared problem space.

6 2 让 AI 的建议变成可质疑对象

Making AI recommendations open to challenge

文字回答的流畅性容易掩盖推断跳跃。把建议转成因果假设图、决策树或证据矩阵，可以让这些跳跃更容易定位。例如模型说“加强短视频推广能够提高入住率”，图中至少需要区分曝光、有效访问、预订转化、取消率与实际入住；还应考虑房价、季节和渠道佣金。任何一个连接缺少数据，都应标为假设，而不是被一根箭头偷偷升级为事实。

这种做法可以称为把回答外化为审查对象。人不必阅读模型声称的全部内部思考，而是要求公开可检查的依据、变量定义、计算步骤和方案差异。一个架构图不能证明模型内部真实采用了同样的过程；它能够展示的是系统规定的流程或实际记录的工具调用。报告严格区分解释性示意图、执行日志和内部机制推断，避免用可视化制造不应有的透明感。

有效的 AI 审查还需要反例。可以要求模型生成一个会使建议失败的条件，再把该条件放回图中检查。比如推广带来的客群对价格敏感，导致新增订单虽然增加，却挤占高贡献订单；或者客服自动回复节省时间，却把复杂投诉拖延到更晚。图使反例能够与原方案共享结构，团队就不必只在口头上讨论“可能存在风险”，而能指出风险发生的位置与可监测信号。

Fluent prose can conceal inferential gaps. Turning a recommendation into a causal-hypothesis diagram, decision tree, or evidence matrix makes those gaps easier to locate. If a model proposes short-video marketing to increase occupancy, the representation should distinguish exposure, qualified visits, bookings, cancellations, and completed stays. It should also consider price, seasonality, and channel commission. An unsupported connection remains a hypothesis; an arrow must not silently promote it into a fact.

This approach externalizes an answer as an object of review. It does not require access to all of a model's claimed internal thinking. It asks for inspectable evidence, variable definitions, calculations, and differences among alternatives. A workflow graphic does not prove that a model internally followed the same process. It can document a specified system procedure or recorded tool calls. Explanatory illustrations, execution records, and claims about internal mechanisms must remain distinct.

Review also needs counterexamples. A model can propose a condition under which its advice would fail, after which people place that condition into the representation. A promotion might attract price-sensitive customers who displace higher-contribution demand. Automated service might save initial response time while delaying resolution of complex complaints. Shared structure helps a team locate the failure mechanism and a monitoring signal, rather than discuss risk only in general terms.

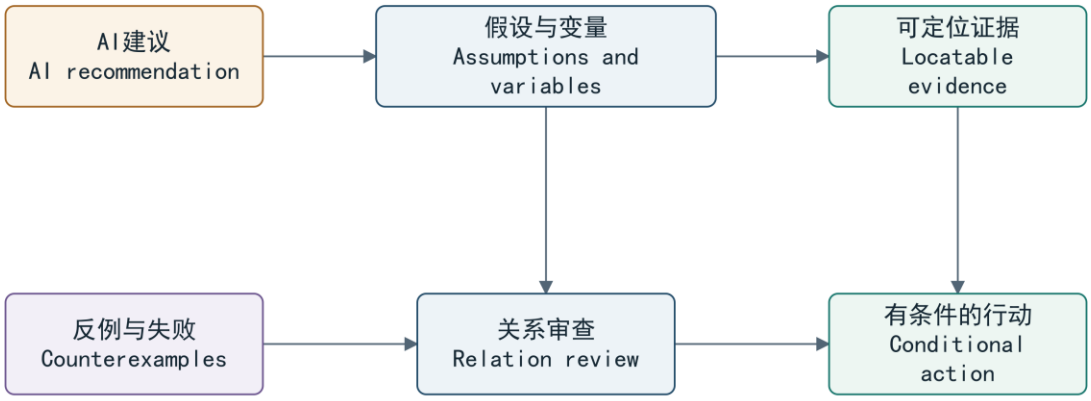


图 6 把流畅建议变成审查对象
Figure 6 Turning fluent advice into a reviewable object
本报告提出的概念与设计示意 Conceptual design proposed in this report

6 3 生产效率与思维能力不能混为一谈
Separating efficiency from thinking capability

一张图制作得更快，是生产效率的改进；一个学生学会判断关系，是能力的改进。两者可以同时发生，也可能相互背离。AI 自动生成课程地图，能节省教师整理素材的时间，却未必让学生建立地图中的关系。如果教师把节省时间用于互动、反馈和迁移任务，效率才可能转化为学习价值。因此，评价 AI 项目应同时记录产出速度、正确性和人的独立表现。

Lee 等人对 319 名知识工作者的调查收集了 936 个 AI 使用实例，发现对 AI 的信心与较少自报批判性思考相关。[28]这是调查关联，不能直接解释为 AI 导致永久能力衰退。对课程的合理启示，是设计可观察的核查活动：要求学生指出一处模型错误、解释一个改动，并在不使用 AI 时完成短任务。这样评价的是实际表现，而不是仅凭态度推断能力。

教师自身也需要保留独立判断。可以把备课过程中的时间分为找资料、设计结构、绘图排版、核对事实和试讲五部分。AI 可能显著减少前几项，却增加核查负担。完整成本需要把返工和错误修正算进去。对一个学期而言，可复用的清晰规范可能比单次惊艳生成更有价值；对学生而言，能够识别自己什么时候需要工具，也是一项重要的元认知能力。

Faster graphic production is an efficiency gain. A learner's improved ability to judge relationships is a capability gain. They may coincide or diverge. An automatically produced course map can save preparation time without ensuring that students understand its relationships. Efficiency becomes educational value when the released time supports interaction, feedback, and transfer. Evaluation should therefore record speed, correctness, and independent human performance.

Lee and colleagues surveyed 319 knowledge workers and collected 936 examples of AI use. Greater confidence in AI was associated with less self-reported critical thinking.[28] This is an association in survey data, not proof of permanent cognitive decline caused by AI. A reasonable instructional response is to require observable checks: identify a model error, justify a revision, and complete a short task without AI. Actual performance then provides evidence beyond attitudes.

Instructors also need independent judgment. Preparation time can be separated into finding material, designing structure, drawing and layout, checking facts, and rehearsal. AI may reduce some activities while increasing verification demands. Total cost should include correction and rework. Across a semester, reusable clear specifications may be more valuable than an impressive one-off generation. Students likewise need the metacognitive ability to recognize when a tool is useful and when independent work is necessary.

6 4 人机互补必须经过流程设计
Designing human AI complementarity

Vaccaro 等人的元分析纳入 106 项实验研究、370 个效应量，发现人机组合平均而言并没有超过其中表现更好的一方，且创作与决策任务的情况不同。[27]这提醒我们，增加一个“人工审核”节点或增加多个 AI 角色，并不能自动保证更好结果。互补需要明确谁处理什么、如何暴露错误，以及发生分歧时怎样决策。视觉工作流的价值在于把这些安排变成可检查的设计。

在教学报告系统中，可以让 AI 提出三种图形方案，让教师选择并说明理由；在经营决策中，可以让系统列出数据与情境，负责人确认目标后再进行分析。若人只是按下同意按钮，就很难构成有效互补。若 AI 一直重复人已经会做的工作，也可能增加沟通成本。合理分工需要通过真实任务中的错误类型、用时和最终判断来不断调整。

Vaccaro and colleagues' meta-analysis covered 106 experimental studies and 370 effect sizes. On average, human–AI combinations did not outperform the better individual contributor, and results differed between content creation and decision tasks.[27] Adding a human-review node or several AI roles is therefore not sufficient. Complementarity requires explicit responsibilities, visible errors, and a method for resolving disagreement. A visual workflow makes those arrangements inspectable.

In a teaching-material system, AI might propose three representation alternatives and an instructor choose one with a reason. In a business system, the software might present evidence and scenarios before a responsible person confirms the objective. Merely pressing an acceptance button offers limited complementarity. Repeating work that a human already performs easily may also add coordination cost. Division of labor should be adjusted using actual error patterns, time records, and decision quality.

表 6 人机分工与能力评价
Table 6 Division of work and capability assessment

任务 Task	AI 可辅助 AI contribution	人需要保留 Human responsibility
问题定义 Framing	整理歧义与候选问题 Organize ambiguities and candidates	目标价值与边界 Objectives values and scope

任务 Task	AI 可辅助 AI contribution	人需要保留 Human responsibility
资料研究 Research	检索线索与字段提取 Find leads and extract fields	核对来源与设计 Verify sources and designs
结构设计 Structuring	提出多个关系草稿 Propose relational drafts	选择并解释关系 Select and justify relations
绘制与转换 Rendering	代码布局与翻译初稿 Draft code layout and translation	检查语义与可读性 Check meaning and readability
数量推演 Computation	生成方案调用计算 Draft and invoke computation	确认假设与复算 Confirm assumptions and reproduce points
行动决策 Decision	比较情境与反例 Compare scenarios and counterexamples	承担权限与后果 Own authority and consequences

本报告的综合分析与设计建议 Report synthesis and design recommendations

7 AI 实现的技术路线
Technical routes for AI supported visual thinking

7 1 语言到结构再到图形
From language to structure to graphics

最适合课程与管理场景的起点，是把自然语言转换为明确结构，再由确定性工具渲染。输入可以是教材段落、业务规则、会议纪要或学生口述。模型先提取实体、关系、顺序和条件，输出中间表示；人检查语义后，系统再生成流程图、概念图或表格。这个分层避免把内容理解、布局选择和像素绘制混在一次不可检查的生成中。模型可以有创造性，但图的关键关系应当能被逐项检查。

中间表示可以非常简单：节点编号、节点类别、显示名称、来源、连接起点、终点、关系类型、条件和确认状态。比如一条连接写明从“检索政策”到“检查版本”，关系是执行顺序，来源是课程教师确认的操作规范。后来更换布局时，语义可以保持不变；后来发现政策版本判断应更早出现，也可以修改关系而不必重新手绘所有元素。稳定编号还能支持中文与英文标签保持一致。

Mermaid 等文本图形工具适合表达许多常见流程和结构，其官方文档列出相应语法与限制。[34]但一种工具不能自然覆盖所有需求。复杂的数量关系需要图表语法，精确空间需要坐标或建模工具，丰富的版式可能需要矢量编辑。报告建议先决定语义结构，再选择渲染器；不要因已经熟悉某种语法，就把所有问题改造成它擅长的形状。

A practical starting point for education and management is to transform natural language into an explicit structure, then render it with a deterministic tool. Inputs may include a textbook passage, operating rule, meeting record, or spoken explanation. The model extracts entities, relations, sequences, and conditions into an intermediate representation. A person checks its meaning before the system produces a diagram or table. This separates interpretation, layout, and pixel production instead of combining them in one opaque generation. Creative suggestions remain possible, while important relationships can be checked individually.

The intermediate representation can contain node identifiers, categories, labels, sources, edge endpoints, relation types, conditions, and confirmation states. A connection from policy retrieval to version checking might specify procedural sequence and cite an instructor-approved operating rule. Changing the layout need not change the meaning. Discovering that version checking belongs earlier can be addressed by editing relationships rather than redrawing everything. Stable identifiers also keep Chinese and English labels aligned.

Text-based diagram tools such as Mermaid support many common processes and structures, with their syntax and limitations described in official documentation.[34] They do not cover every need. Quantitative relations may require chart specifications, precise spatial work may require coordinates or modeling tools, and richer layouts may require vector editing. Semantics should determine the renderer, rather than familiarity with one renderer determining how every problem is represented.

7 2 数据到代码再到可复算图表
From data to code to reproducible charts

第二条路线从结构化数据出发。模型帮助解释字段、提出分析问题、生成清洗与绘图代码，人检查字段口径和变换步骤，再执行代码得到图。LIDA 展示了利用大模型组织自动可视化流程的研究方向。[21]Data Formulator 则通过高层概念与视觉通道的绑定，帮助用户表达意图，并展示变换后的数据以供检查。[22][23]这些系统提供了设计启发，不应被误解为所有自动图表已经具备可靠事实判断。

实际流程需要保留输入数据、转换脚本、输出数据和图表规范四种对象。只有图片而没有变换记录，读者无法判断重复订单是否被计了两次；只有代码而没有字段解释，也无法确认“收入”是否包含税费和佣金。一个可信的 AI 助手应能回答：“这根柱子用了哪些记录？采用什么过滤条件？缺失值怎么处理？为什么选择这个时间粒度？”这些问题比询问模型是否有信心更有用。

Vega-Lite 等声明式语法可以明确编码与转换，使数据、映射和呈现更容易分离。[35]Python 或其他分析语言则适合复杂计算。无论采用哪一种，都需要实际执行，而非把模型写出的代码当成已经得到的结果。执行后还应选取少量数据点手算核查，检查异常范围、排序与单位。对于高风险决策，还需要独立复核和真实业务验证。

A second route begins with structured data. AI helps interpret fields, propose analytical questions, and generate transformation and plotting code. People check definitions and operations before execution. LIDA illustrates research on organizing an automated visualization pipeline with language models.[21] Data Formulator helps users express intent through data concepts and visual-channel bindings, while exposing transformed data for inspection.[22][23] These systems provide useful design ideas; they do not establish that every automatically produced chart makes reliable factual judgments.

A practical workflow preserves four objects: input data, transformation code, transformed data, and the visualization specification. A picture alone cannot show whether duplicate bookings were counted twice. Code alone cannot establish whether revenue includes taxes or commissions. A trustworthy assistant should explain which records generated a bar, what filters were applied, how missing values were treated, and why a particular time interval was used. Such questions are more useful than asking whether the model feels confident.

Declarative languages such as Vega-Lite make encodings and transformations explicit.[35] General analytical languages can support more complex calculations. In either case, code must actually run; generated code is not an executed result. Selected points should then be recomputed independently, with checks on ranges, sorting, and units. Consequential decisions require additional review and validation against operating conditions.

7 3 草图与图像作为输入
Sketches and images as input

第三条路线允许人先画。教师可以拍下一张白板，学生可以提交粗糙流程，设计人员可以上传空间草图。多模态模型帮助转录标签、识别候选节点、整理层级并提出疑问。这里最大的价值

不是把笔迹变整齐，而是降低表达入口：人可以先用擅长的形式组织想法，再借 AI 转成可编辑结构。系统应保留原图与识别结果的对应，允许逐项纠正。

图像输入也会发生误读。箭头交叉不一定相连，近邻不一定表示从属，手写否定词一旦漏掉就可能反转含义。建议识别阶段输出两类内容：已经看清的观察与需要确认的解释。比如“看到 A 与 B 之间有一条线”是观察，“A 导致 B”是解释。让用户确认关系意义之后再清洁排版，可以减少把识别错误固化成漂亮成品的风险。

对学生作业，可以采用先解释再整图的顺序。模型先复述自己读到的结构，提出最多三项关键疑问，学生回答后再生成规范图。教师可以比较原图与整理图之间哪些概念发生变化，避免工具把不成熟但有价值的想法自动改成常见模板。保存这些差异，也能成为评价学生思考过程的材料。

A third route lets people begin with drawing. An instructor may photograph a whiteboard, a student submit a rough process, or a designer provide a spatial sketch. A multimodal model can transcribe labels, identify candidate nodes, organize hierarchy, and ask questions. The main value is lowering the entry barrier rather than merely improving handwriting. People can start with a comfortable form and convert it into editable structure. The original image should remain linked to the interpretation so that individual elements can be corrected.

Image input can also be misread. Crossing arrows do not necessarily connect. Proximity does not establish hierarchy. Missing a handwritten negative may reverse a rule. The recognition stage should distinguish observations from interpretations. Seeing a line between A and B is an observation; concluding that A causes B is an interpretation. Confirming the relationship before polishing the layout reduces the risk of turning a recognition error into a persuasive artifact.

For student work, the system can first explain what it believes the sketch means and ask a few critical questions. Only after the student answers should it produce a normalized diagram. Comparing the original and revised versions reveals conceptual changes. It also helps prevent an unusual but promising idea from being automatically replaced by a familiar template. Those differences can become evidence of thinking during assessment.

7 4 图像生成与视觉推理的区别
Image generation and visual reasoning

文生图能够帮助探索服务场景、空间气氛、产品外观和故事分镜，适合需要形象联想的任务。精确流程、统计图、公式与政策文字则需要其他控制机制。一幅图看似符合“酒店 AI 系统”的风格，并不能证明其中的按钮、数字与连接正确。建议把生成图中的艺术部分与文字数据层分离，关键文字用可编辑排版加入，关键关系由结构文件决定。

Visual Sketchpad 研究让多模态模型使用线条、框和标记等工具产生视觉中间材料，再基于这些材料继续解题。[20]它表明绘图动作可以进入 AI 推理流程，而不只是最后输出阶段。但这类基准结果仍针对特定模型和任务，不能证明 AI 拥有与人相同的心像体验，也不能保证它在真实课程或企业中具有同样收益。技术能力、认知解释与应用成效需要分别讨论。

Image generation can support exploration of service situations, atmosphere, product appearance, and storyboards. Precise procedures, statistical graphics, equations, and policy language need other control mechanisms. An image that resembles a hotel AI system does not establish that its controls, numbers, or connections are correct. Artistic imagery should be separated from editable text and data layers, with important relationships determined by explicit structural files.

Visual Sketchpad equips multimodal models with tools for producing lines, boxes, and marks, then uses those artifacts during further problem solving.[20] It shows that drawing actions can enter the reasoning workflow instead of appearing only at the end. Such results remain specific to models and benchmarks. They neither demonstrate human-like imagery experience nor guarantee equivalent gains in classrooms or businesses. Technical capability, cognitive explanation, and application effectiveness require separate evaluation.

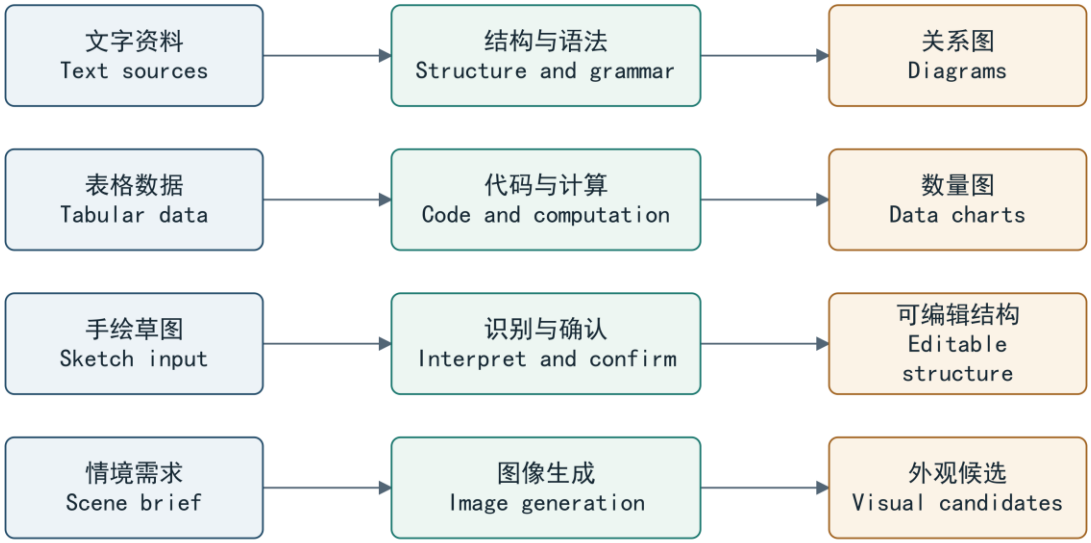


图 7 AI 支持视觉思维的主要路线
Figure 7 Main routes for AI supported visual thinking
本报告提出的概念与设计示意 Conceptual design proposed in this report

表 7 AI 实现路线及扩展能力比较
Table 7 AI implementation routes and extensions

路线 Route	主要产物 Output	适用条件与核查 Conditions and checks
语言到结构 Language to structure	节点关系与结构图 Nodes edges and diagrams	核对遗漏关系与来源 Check omissions relations and provenance
数据到代码 Data to code	变换数据与数量图 Transformed data and charts	实际执行并复算 Execute and independently recompute

路线 Route	主要产物 Output	适用条件与核查 Conditions and checks
草图到结构 Sketch to structure	转录与可编辑图 Transcription and editable structure	确认箭头与手写否定 Confirm edges and handwritten negation
情境到图像 Scenario to image	外观与场景候选 Appearance and scene candidates	精确文字另行排版 Set exact text separately
视觉中间材料 Visual intermediate artifacts	标记裁剪计算草图 Marks crops and computational sketches	检查中间材料和任务收益 Check artifacts and task benefit
交互计算 Interactive computation	参数控制与结果状态 Parameters and recorded states	区分模型探索与实证 Separate model exploration from validation

本报告的综合分析与设计建议 Report synthesis and design recommendations

8 可信的生成与验证架构
An architecture for trustworthy generation and validation

8 1 四层架构与稳定中间表示
Four layers and a stable intermediate representation

建议采用资料层、语义层、呈现层和验证层的四层架构。资料层保存来源、日期、访问范围与版本；语义层保存经过确认的概念、关系、规则和变量；呈现层负责布局、图形语法、语言与输出格式；验证层检查来源对应、结构约束、计算和可读性。四层可以由一个系统实现，也可以由多个工具配合；关键是逻辑上可以区分，出现错误时知道回到哪里修复。

例如，酒店政策变化是资料层更新，流程中增加人工确认是语义层修改，把中文标签翻成英文是呈现层修改，发现审批分支无法回到主流程是验证层问题。若这些修改都直接发生在最终图片上，后续版本很难保持一致。保留中间表示后，一次概念修正可以传播到 PPT、手册、作业和英文图中，同时留下修改记录。

中间表示还应包含不确定性。一个关系可以是来源明确、教师确认、模型推测或等待验证；一个数值可以是观察值、估计值或教学情境值。不同状态不是装饰标签，而会决定系统是否允许发布、是否需要图注和是否触发人工复核。这样，可信度管理进入生成流程，不再依赖最后一位审核者凭记忆补救。

The proposed architecture separates sources, semantics, presentation, and validation. The source layer stores origin, dates, access scope, and versions. The semantic layer contains confirmed concepts, relationships, rules, and variables. The presentation layer controls layout, notation, language, and export format. The validation layer checks source correspondence, structural constraints, calculations, and readability. One system or several tools may implement these layers. The essential requirement is logical separation so that an error can be repaired at the appropriate level.

A changed hotel policy is a source update. Adding human confirmation is a semantic revision. Translating labels is a presentation revision. Discovering that an authorization branch cannot rejoin the process is a validation finding. If all changes occur directly in a final image, consistency becomes difficult. With an intermediate representation, one conceptual correction can propagate into slides, manuals, assignments, and English figures while preserving its history.

Uncertainty should also be represented. A relationship may be source-supported, instructor-confirmed, model-proposed, or awaiting verification. A number may be observed, estimated, or synthetic. These states determine whether publication is allowed, whether a note is required, and whether human review is triggered. Trust management then becomes part of generation rather than a final reviewer having to remember everything.

8 2 四种验证各自解决不同问题

Four distinct forms of validation

第一种验证检查语法：代码是否能解析，节点编号是否重复，输出文件是否完整。第二种检查结构：是否存在悬空节点、无法到达的状态、缺失的分支条件或无限循环。第三种检查语义：一条关系是否符合原文，某个动作是否属于该岗位，一项术语是否在中英文中保持相同含义。第四种检查经验与后果：模型假设是否符合实际，方案是否在小规模试验中达到目标。通过前一种，并不意味着自动通过后一种。

图表也可以按这四种验证设计检查。代码运行成功只说明它被执行；图没有重叠只说明布局可读；数据映射正确才支持数值解释；真实业务试验才可能支持行动效果。AI 可以协助前三种检查，但自检容易受到同一错误解释影响。关键事实、计算和权限应尽可能采用独立规则、原始资料或人员复核。让同一模型重复说“检查通过”不构成独立证据。

本报告建议每张关键图附一个简短验证记录：来源已核对，关系已确认，计算已执行，限制已说明。记录还应允许“未检验”状态。例如合成收益曲线可以通过计算核验，但经验有效性应标为未验证。把未验证明确写出来，通常比给整张图一个笼统的绿色通过图标更诚实，也更能指导下一步工作。

Four kinds of validation address different questions. Syntax checks whether code parses, identifiers are unique, and output files are complete. Structure checks for disconnected nodes, unreachable states, missing branch conditions, and uncontrolled loops. Semantics checks whether a relationship matches its source, an action belongs to a role, and terminology remains aligned across languages. Empirical and consequence-oriented validation asks whether assumptions fit operating conditions and whether a small trial achieves its objective. Passing one level does not automatically pass the next.

The same distinction applies to charts. Successful execution establishes that code ran. A readable layout establishes that elements can be seen. Correct data mapping supports numerical interpretation. A real operating trial is needed to assess action effects. AI can assist several checks, but self-review may repeat the original misconception. Important facts, calculations, and permissions should use independent rules, primary sources, or human review where possible. Repeated approval by the same model is not independent evidence.

A short record can accompany each important graphic: sources checked, relationships confirmed, calculations executed, and limitations stated. The record should permit an untested state. A synthetic revenue curve may pass computational validation while its empirical applicability remains untested. Stating this explicitly is more informative than assigning the entire figure a single green approval symbol.

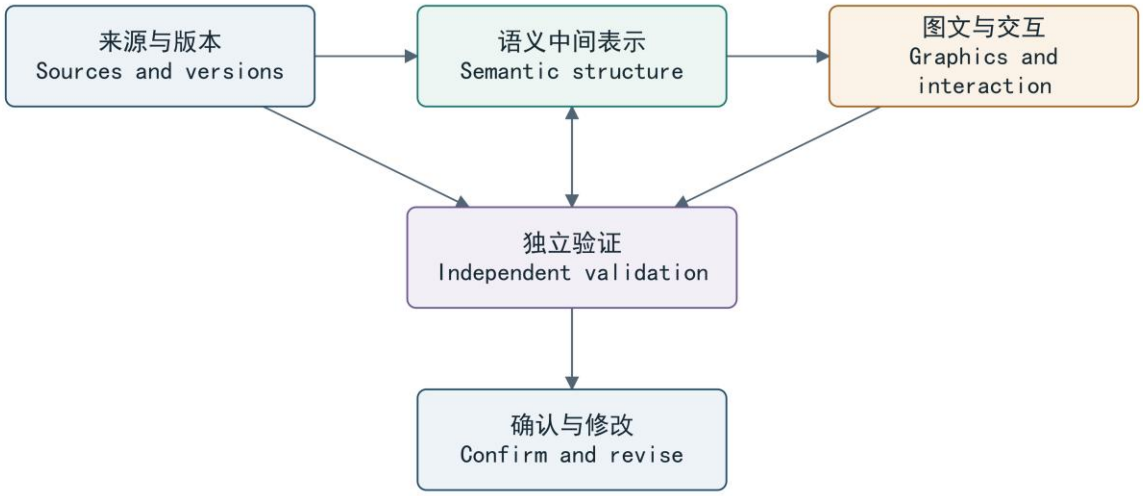


图 8 来源语义呈现与验证分层
Figure 8 Sources semantics presentation and validation
本报告提出的概念与设计示意 Conceptual design proposed in this report

8 3 证据如何连接到图中的一个元素

Linking evidence to graphical elements

可追溯性应细到重要节点和关键连线，而不只是文末列出一串参考文献。一个政策决策节点可以连接到原始条款，数据柱可以连接到过滤后的记录，研究证据图可以连接到原文中的效应量说明。为了降低维护负担，不要求每一个装饰元素都有来源，但任何影响判断的事实、数字和规范都应可定位。AI 生成的补充解释应与来源陈述分别记录，避免把合理推测伪装成引用。

具体操作可采用来源编号加元素编号。例如源文件 S03 中的第三条政策支持节点 N12；分析脚本 T02 产生图表 F06 中的序列 B。中英文标签共享 N12，而不是各自维护两个事实副本。引用定位不必暴露敏感数据，读者可以看到受控说明与访问条件。对学生课程，教师可以使用去标识化资料或模拟数据，让可追溯练习不依赖真实顾客信息。

证据还需要时间维度。一次回答引用正确的旧政策，仍可能在当前场景出错。资料层应存生效时间与版本，而不只是上传日期。对于搜索资料，记录检索日期与访问位置；对于计算，记录参数与运行版本；对于人工判断，记录确认人和理由。这样的记录能够解释为什么上周正确的图在本周需要更新。

Traceability should reach important nodes and edges rather than stop at a bibliography. A policy decision can link to its clause, a bar to the filtered records that produced it, and an evidence graphic to the reported estimate. Decorative elements do not all need citations, but facts, numbers, and rules that affect judgment should be locatable. AI-generated explanation should be recorded separately from source statements so that plausible inference does not become a disguised quotation.

A simple system combines source and element identifiers. A clause in source S03 supports node N12; transformation T02 produces series B in figure F06. Chinese and English labels share N12 rather than maintain separate factual copies. Traceability need not expose sensitive data. A reader can receive a controlled description and access conditions. In education, de-identified or simulated records can support the exercise without requiring real guest information.

Time is part of evidence. An answer may accurately quote an old policy and still be wrong for the current situation. Effective dates and versions matter alongside upload dates. Search material needs retrieval information, computations need parameter and execution versions, and human judgments need a responsible person and reason. These records explain why a diagram that was correct last week may require revision now.

8 4 权限与失败路径进入设计
Permissions and failure paths

课程和企业系统都应画出失败情况。资料缺失、工具报错、数据越权、模型不确定和用户改变目标，不是罕见例外，而是正常运行条件。一个工作流至少应说明停止、追问、重试和转人工的条件。重试应有次数或时间上限，否则自动化可能在没有新信息的情况下持续消耗资源。对外发送、退款、价格修改和数据删除等动作，应具有与后果相称的授权控制。

NIST 的生成式 AI 风险框架为识别、测量和管理风险提供通用参照。[30]在本报告的课程设计中，这被转化为具体图形要求：权限边界可见，关键动作可回溯，异常状态有出口，责任人可识别。它们是教学与工程建议，而不是宣称满足某一图形格式就已经满足全部法规或组织政策。真实部署仍应使用所在机构的要求进行核对。

Failure paths belong in both classroom and business systems. Missing evidence, tool errors, permission conflicts, uncertainty, and changed objectives are ordinary operating conditions. A workflow should specify when to stop, ask, retry, or escalate. Retries need limits so that automation does not continue consuming resources without new information. External messages, refunds, price changes, and deletion need authorization proportionate to their consequences.

NIST's generative-AI risk framework provides a general reference for identifying, measuring, and managing risks.[30] Here it informs practical design requirements: visible permission boundaries, traceable consequential actions, exits from exceptional states, and identifiable responsibility. These are educational and engineering recommendations. A particular diagram format does not itself establish regulatory or organizational compliance; deployment needs local requirements to be checked.

表 8 分层架构的错误定位
Table 8 Locating errors by architecture layer

层次 Layer	典型错误 Typical error	修复与责任 Repair and ownership
资料层 Sources	政策过期或范围不明 Expired or ambiguous policy	资料负责人核对有效期 Source owner checks validity

层次 Layer	典型错误 Typical error	修复与责任 Repair and ownership
语义层 Semantics	顺序误作因果 Sequence presented as causation	领域人员确认关系 Domain reviewer confirms relation
数据变换 Transformation	重复订单被多次计数 Duplicate bookings counted	分析人员检查规则 Analyst checks operations
呈现层 Presentation	标签拥挤或单位缺失 Crowded labels or missing units	设计者调整编码布局 Designer revises encoding and layout
执行层 Execution	越权或无重试上限 Unauthorized action or unlimited retry	系统负责人限制权限 Owner constrains authority
经验验证 Empirical validation	改善机制未被支持 Unconfirmed improvement mechanism	小范围测试并修订 Pilot and revise

本报告的综合分析与设计建议 Report synthesis and design recommendations

9 人与 AI 共同思考的工作方法
A working method for thinking with AI

9 1 以问题与判断标准开始
Start with the question and criteria

人与 AI 协作的第一步，是把任务写成能够判断完成与否的问题。不要从“帮我做一张思维导图”开始，而要从“我要判断酒店客诉为何在夜间积压，需要区分客流、人员与升级规则”开始。前者指定了产品，后者指定了认知任务。只有认知任务清楚，系统才知道应索取哪些资料、保留哪些例外以及采用什么表征。图的形式可以在比较候选后决定。

任务说明至少包含对象、受众、行动目的、可用资料和限制。对课程来说，还应加入学生基础与需要证明的能力。对行业项目，还应加入时间范围、业务指标、数据权限和哪些动作必须人工确认。AI 可以把缺失项整理成问题清单，但不应为追求完整而擅自填入事实。无法获得的数据可以标为缺口，无法确认的关系可以标为假设，任务仍然可以在这些边界内前进。

判断标准需要提前明确。比如一张客诉流程图的合格标准是：正常路径可执行、责任明确、证据来源可查、三种异常有出口、学生能解释升级理由。美观只占其中一部分。若直到生成后才决定标准，人很容易被成品外观吸引，从而降低对内容的要求。先定标准，也使不同方案之间的比较更公平。

The first step is to express the task as a question whose completion can be judged. Instead of beginning with a request for a mind map, specify that the investigation concerns why complaints accumulate at night and must distinguish demand, staffing, and escalation rules. The former specifies a product; the latter specifies a cognitive task. A clear cognitive task tells the system what evidence to request, which exceptions to preserve, and what representations may be useful. The form can be chosen after alternatives are compared.

A task brief includes the object, audience, action purpose, available evidence, and constraints. Educational work also needs prior knowledge and the capability learners must demonstrate. Industry work needs a time span, business measures, permissions, and actions requiring confirmation. AI can organize missing information into questions, but it should not invent facts for completeness. Missing data and unconfirmed relationships can remain visible while work proceeds within those limits.

Success criteria should be defined before generation. A complaint workflow might require an executable normal route, clear ownership, traceable evidence, exits for three exceptions, and an explanation of escalation. Appearance is only one criterion. If standards are chosen after seeing an attractive artifact, people may unconsciously relax content requirements. Prior criteria support fairer comparison among candidates.

9 2 先形成多个低成本候选
Generate inexpensive alternatives

视觉思维的一项优势，是能够在方案还不昂贵时比较不同结构。AI 可以根据同一问题生成一张责任泳道图、一张瓶颈时间图和一张原因假设图。人分别说明它们能回答什么、不能回答什么，

再决定下一步。这比过早把一种形式精修到完成状态更适合探索。候选不必都生成完整图片，简短结构草稿就足够判断方向。

比较时应记录放弃理由。例如原因图缺少数据，暂时用于提出调查问题；责任泳道图已有规范依据，适合安排跨部门讨论；时间图需要收集时间戳，暂时不能完成。这样的记录让项目逐步积累知识，而不是只留下最终成品。模型的价值在于扩大候选空间，人的价值在于根据场景与证据缩小空间。两者相互支持，才能避免生成数量替代判断质量。

对学生，可以把候选比较设计成短讨论。每组领取两种图，不立即选择“最好”的，而是解释各自适合哪个问题。随后加入新的条件，例如夜班人手不能增加，要求重新选择。学生会看到形式与目的之间的关系，也能学习承认一个图的局限。AI 在此承担对话素材提供者，而非权威答案提供者。

Visual work makes alternative structures inexpensive to compare early. For one problem, AI can propose a responsibility swimlane, a bottleneck timeline, and a causal-hypothesis map. People explain what each can and cannot answer before choosing the next step. Simple structural drafts are sufficient; every candidate need not become a polished image. Early comparison is often more useful than refining one format before its suitability is established.

Reasons for rejection should be preserved. The causal map may lack data and remain useful only for planning an investigation. The swimlane may have documentary support and be ready for a cross-department meeting. The timeline may require timestamps that have not been collected. Such records accumulate knowledge instead of leaving only a final artifact. AI broadens the candidate space, while people narrow it using evidence and context. More generation then supports better judgment rather than substituting for it.

Students can compare two representations without immediately naming a winner. Each group explains which question each form serves. A new constraint, such as a fixed night-shift workforce, then requires reconsideration. This develops understanding of the relationship between form and purpose and the ability to acknowledge limitations. AI supplies material for discussion rather than an authoritative answer.

9 3 用解释 修改与反例完成迭代

Iterate through explanation revision and counterexamples

一次有效迭代至少包含三种活动：解释当前结构，修改一个重要关系，检验一个反例。解释防止只看表面；修改检验是否理解依赖；反例检验适用边界。以 RAG 流程为例，学生先说明检索资料怎样进入回答，再把“资料冲突”加入流程，最后处理“查到资料但资料已过期”的情形。完成这三步，比单纯把原图换色更能显示理解。

AI 反馈可以采用问题而不是直接替换。例如“这一动作是否需要客户身份？”“两个政策日期不同，图中在哪里判断？”“删除审核节点后，什么错误会被放行？”系统不应在每次回答后都自动给出完整正确图，否则学生容易只做接受操作。教师可把反馈分为提示、局部示例和完整解释三个层级，按学生需要逐步开放。

修改还应区分概念变化与版式变化。节点位置调整可能提高可读性，却不一定改变推理；新增授权条件则改变流程含义。学生提交时应分别写出“我改了什么”和“为什么改”。这种记录既有助于评价，也能让 AI 在后续任务中使用准确的偏好与规则，而不是根据外观推测学生的思考。成熟的协作保留有理由的改变，而不是无限生成。

A meaningful iteration includes explanation, consequential revision, and a counterexample. Explanation guards against surface reading. Revision tests dependency understanding. A counterexample tests scope. In a retrieval workflow, a student can first explain how evidence enters a response, then add conflicting sources, and finally handle a retrieved but expired policy. These activities reveal more than changing the diagram's colors.

AI feedback can take the form of questions rather than immediate replacement: does this action require identity verification, where are conflicting dates compared, and what error could pass if review were removed? Automatically displaying a complete correct diagram after every response can reduce the learner's role to acceptance. Hints, partial examples, and full explanations can instead be released progressively.

Conceptual changes should be distinguished from layout changes. Moving a node may improve readability without changing reasoning. Adding an authorization condition changes the process itself. Students should state what changed and why. These records help assessment and allow later AI assistance to use confirmed rules rather than infer thinking from appearance. Mature collaboration preserves reasoned changes instead of producing endless alternatives.

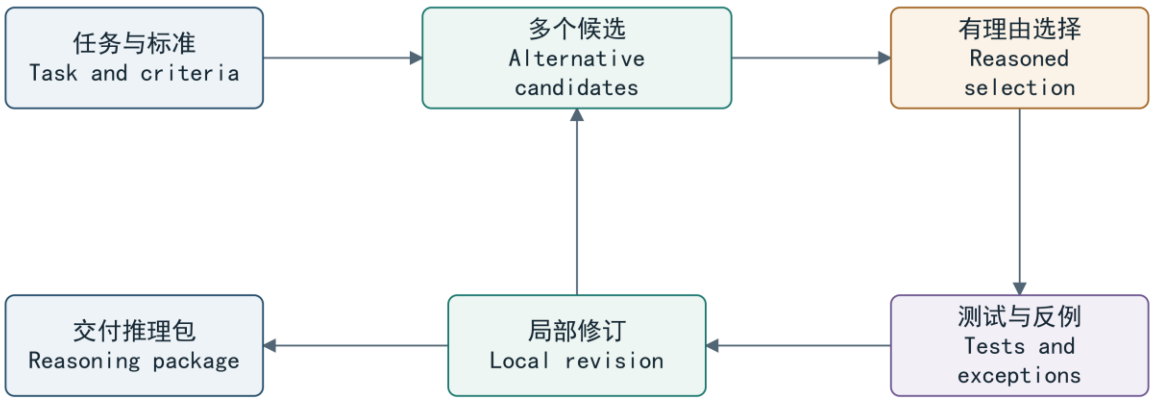


图 9 人机协作中的解释修订与反例
Figure 9 Explanation revision and counterexamples
本报告提出的概念与设计示意 Conceptual design proposed in this report

9 4 形成可交付的推理包
Deliver a reasoning package

建议最终交付一个相互对应的推理包：问题说明、主要图、数据或来源表、关键假设、替代方案、验证记录和下一步行动。图不能独自承担全部责任，文字也不必重复图中所有节点。两者分工应当清楚：图显示结构，文字解释意义，表保留可比较记录，附录支持复算。这样的组合适合课堂答辩，也适合企业中的跨岗位讨论。

交付还应设定阅读顺序。受众先看到问题和主要结论，再查看关系与证据，需要时进入细节。对不同受众可以保留相同事实结构而改变展开程度，避免为领导、教师与学生各写一份事实不一致的材料。中英文版本也遵循相同原则：共享编号与数据，分别优化表达。视觉思维的成果最终应让别人能够继续思考，而不是只能欣赏作者已经完成的画面。

A final reasoning package can include a problem statement, principal graphic, evidence or data table, assumptions, alternatives, validation record, and next action. A diagram cannot carry every responsibility by itself, and prose need not repeat every node. Graphics show structure, prose interprets meaning, tables preserve comparable records, and appendices support reproduction. The combination suits both classroom defense and cross-functional business discussion.

Reading order should also be designed. An audience first encounters the question and conclusion, then relationships and evidence, with detail available when needed. Different audiences can receive different levels of expansion while retaining one factual structure. Chinese and English versions should likewise share identifiers and data while using natural expressions in each language. The purpose is to enable others to continue the reasoning.

表 9 一次视觉推理工作会的产物
Table 9 Artifacts from a visual reasoning session

阶段 Stage	需要留下 Record	检查重点 Check
澄清 Clarify	问题与成功标准 Question and success criteria	是否可以判断完成 Can completion be judged
候选 Propose	两三种不同结构 Two or three different structures	差异是否涉及概念 Are differences conceptual
选择 Select	采用与放弃理由 Selection and rejection reasons	是否依据任务与证据 Do reasons follow evidence
检验 Test	反例异常与复算点 Exceptions and reproducible points	是否挑战主要假设 Do tests challenge assumptions
修订 Revise	概念与版式变化 Semantic and layout changes	是否同步相关产物 Are related outputs updated
交付 Deliver	推理包与责任人 Reasoning package and owner	别人能否继续使用 Can another person continue

本报告的综合分析与设计建议 Report synthesis and design recommendations

10 酒店运营与服务设计案例
Hotel operations and service design cases

10 1 用服务蓝图看见前后台联系
Connecting frontstage and backstage through blueprints

酒店服务适合说明视觉思维，因为顾客看到的体验与后台实际工作并不重合。一次入住涉及预订、确认、到店、登记、入住、服务请求、结账和后续评价；后台还包括房态更新、清洁排班、库存、支付和渠道同步。服务蓝图可以把顾客动作、可见接触、后台动作和支持系统分层排列，使团队看到交接与等待。AI 可先根据材料提出候选，但实际岗位与规则需由酒店人员核实。

以提前入住请求为例，顾客看到的是一句“能否现在入住”，背后需要判断房间是否清洁、原住客是否退房、是否存在会员权益、是否收费以及谁能批准。若仅画一个“AI 客服回复”节点，就掩盖了实际依赖。合适的图应标出资料读取与行动权限：AI 可以解释规则、查询已授权状态、提出候选，但不应把没有确认的房间标为可入住。服务蓝图使能力边界具体到动作。

课堂任务可提供一份简化资料，让学生找出三个潜在等待点，并解释需要什么数据验证。图中的“等待清洁”只是候选瓶颈，不能凭图认定其造成满意度下降。学生还应考虑到店时间分布、房型差异和人员安排。AI 可以帮助提出调查字段，但真正的改善需要基于记录与试验。这样，图同时连接体验理解与运营证据。

Hotels illustrate visual thinking because visible guest experience and backstage work do not coincide. A stay involves booking, confirmation, arrival, check-in, room use, service requests, settlement, and feedback. Supporting work includes room-status updates, cleaning schedules, inventory, payment, and channel synchronization. A service blueprint separates guest actions, visible contact, backstage activity, and support systems so that handoffs and waiting become visible. AI can propose a draft, while hotel staff confirm actual roles and rules.

An early-arrival request may sound simple but depend on cleaning status, the previous guest's departure, membership entitlements, charges, and authorization. A single AI-customer-service node hides these dependencies. A useful representation distinguishes evidence access from action permission. AI may explain rules, query authorized status, and suggest options, while an unconfirmed room must not become available merely because the model proposes it. The blueprint makes capability boundaries specific to actions.

Students can inspect simplified records, identify three possible waiting points, and specify the evidence needed to test them. Waiting for cleaning remains a candidate bottleneck, not a proven cause of dissatisfaction. Arrival distributions, room types, and staffing may also matter. AI can suggest fields to collect, while improvement requires records and trials. The representation connects guest experience with operating evidence.

10 2 酒店知识问答中的证据流程

Evidence flows in hotel knowledge questions

酒店 RAG 教学可以使用早餐时间、退订政策、宠物规则、设施开放和特殊需求等去标识化资料。系统先理解问题与时间条件，再检索候选资料、检查适用范围、生成带依据的答复，并在资料冲突或权限不足时转人工。这个图必须保留“不知道”路径。若教学只展示所有步骤顺利完成，学生容易把检索增强误解为事实保证。

案例可以安排两份日期不同的取消政策。学生首先画出最简单流程，再处理三个测试：用户未提供预订渠道；资料分别适用于团队与散客；当前政策与已签合同冲突。每次测试都需要修改条件或升级路径。AI 可生成回复草稿和测试问题，学生则必须说明哪一条资料适用，哪些信息仍需确认。图的准确性由测试情境决定，而非节点名称是否专业。

评价指标可包括证据引用准确性、适用范围判断、必要追问、错误承诺与升级完整性。不能只统计回答速度和“看起来有帮助”。例如自动承诺免费取消可能让用户满意，却不符合资料或权限；无限转人工虽然安全，却失去辅助价值。课程可让学生比较这两种极端，寻找可解释的分工。真实业务还需将模型输出置于酒店既有服务流程和授权体系之中。

A hotel retrieval exercise can use de-identified material about breakfast, cancellation, pets, facilities, and special requests. The process interprets the question and its temporal conditions, retrieves candidates, checks scope, generates a supported response, and escalates conflicts or permission gaps. It needs an explicit unknown route. An entirely successful demonstration can otherwise make retrieval augmentation appear to guarantee factual accuracy.

Two cancellation policies with different dates support useful tests. Students first build a simple process, then handle an unspecified booking channel, policies for groups versus individual guests, and a conflict between a current rule and a signed agreement. Each test requires a condition or escalation revision. AI may draft responses and test questions, while students justify applicability and identify missing information. Scenario performance, rather than professional-sounding component names, determines accuracy.

Assessment can examine citation accuracy, scope decisions, necessary clarification, unauthorized promises, and escalation completeness. Speed and apparent helpfulness are insufficient. Automatically promising free cancellation may please a guest while violating evidence or authority. Escalating every question avoids some errors but offers little assistance. Comparing these extremes helps students design a reasoned division of work. Real use must fit the hotel's existing service and authorization system.

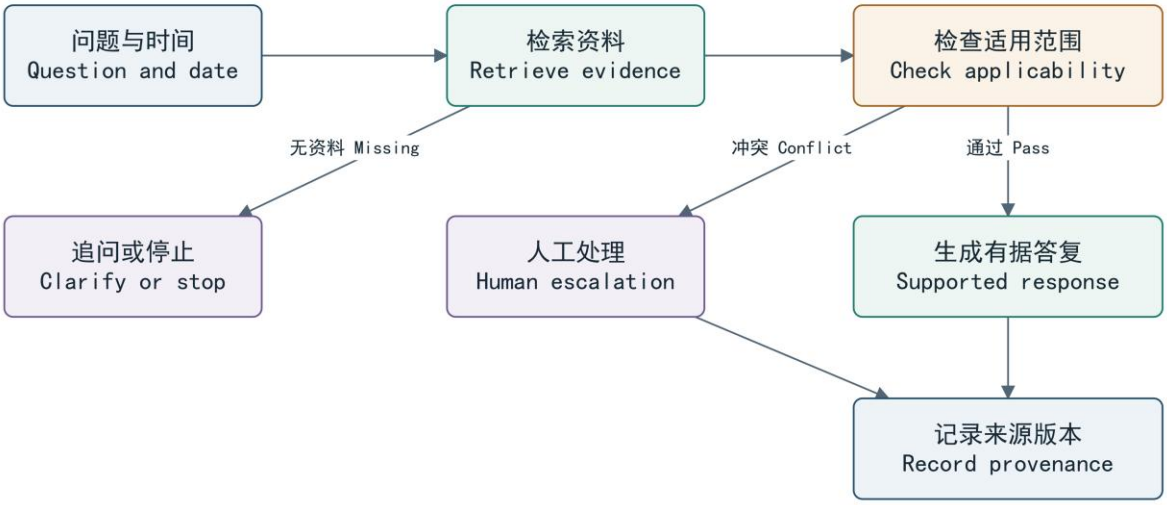


图 10 酒店 RAG 的证据与升级路径
Figure 10 Evidence and escalation in hotel retrieval
本报告提出的概念与设计示意 Conceptual design proposed in this report

10 3 收益与利润目标的计算示例
A computed revenue and profit scenario

以下为合成教学情境，不是任何真实酒店的需求估计。设酒店有 100 间可售客房，房价为 p，预期售出间夜为 95 减去 p 的十分之一；在本例讨论的价格区间内，需求不超过容量。每间已售客房变动成本为 80，每日固定成本为 8000。营业收入等于房价乘以售出间夜，情境利润等于房价减去 80 后乘以售出间夜，再减去 8000。这里允许小数间夜，因为它表示期望值而非某一天的实际房间计数。

Wolfram 符号与数值核算得到：收入最大时房价为 475，预期售出 47.5 间夜，收入 22562.5；利润最大时房价为 515，预期售出 43.5 间夜，利润 10922.5。若房价为 300，预期售出 65 间夜，收入 19500，利润 6300。由此可见，在这一特定假设内，提高入住率、提高收入和提高利润并不是同一个目标。模型没有计入房型差异、渠道佣金、服务容量、客户体验和长期需求，因此不能用来给真实酒店直接定价。

图表教学可让学生先预测两条曲线的峰值是否相同，再查看计算结果，最后改变变动成本或需求斜率。学生需要解释峰值为何移动，而不是只报出最佳价格。若需求函数是错的，精确求出的最佳值也可能没有经营意义。这个案例的价值是展示目标、假设、公式与图如何互相制约，并训练学生区分计算正确和经验正确。

The following is a synthetic teaching scenario, not an estimate of demand for an actual hotel. There are one hundred available rooms. At price p, expected room nights sold equal ninety-five minus one tenth of p. Demand remains below capacity over the illustrated price interval. Variable cost is eighty per occupied room and daily fixed cost is eight thousand. Revenue equals price multiplied by expected room nights. Scenario

profit equals price minus eighty, multiplied by expected room nights, minus eight thousand. Fractional room nights represent expectations rather than a literal daily room count.

Symbolic and numerical checks with Wolfram give a revenue-maximizing price of 475, expected sales of 47.5 room nights, and revenue of 22,562.5. The profit-maximizing price is 515, with 43.5 expected room nights and profit of 10,922.5. At price 300, expected sales are sixty-five, revenue is 19,500, and profit is 6,300. Within these assumptions, maximizing occupancy, revenue, and profit are different objectives. The model excludes room-type differences, commissions, service capacity, guest experience, and long-run demand. It does not supply an actual pricing recommendation.

Students can predict whether the peaks coincide, inspect the computation, and then change variable cost or the demand slope. They should explain why an optimum moves rather than merely report a number. An exact optimum under an incorrect demand function may have little operating value. The exercise demonstrates mutual constraints among objectives, assumptions, formulas, and graphics, while distinguishing computational correctness from empirical adequacy.

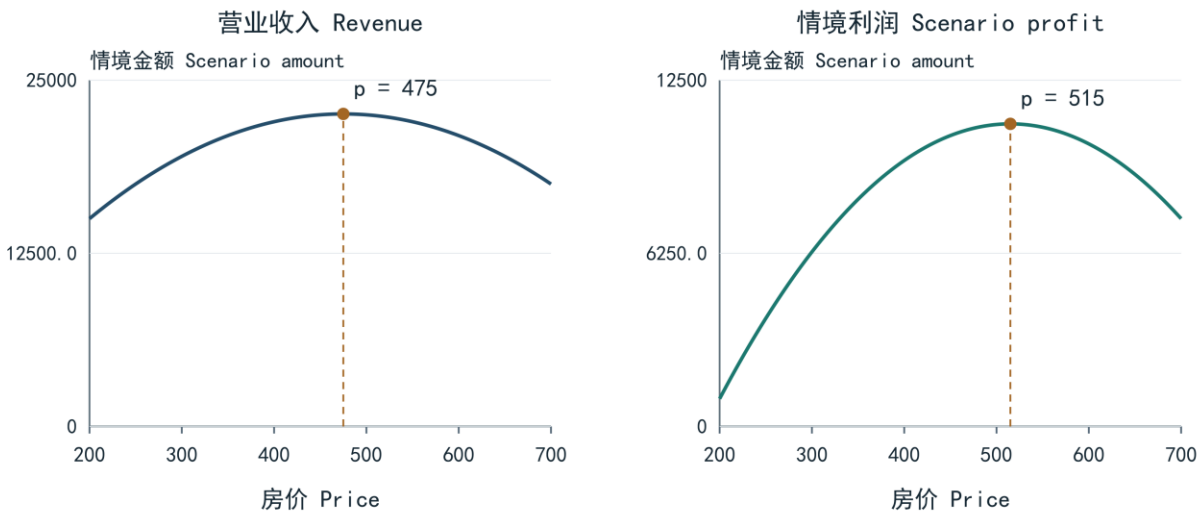


图 11 收入与利润最优点可以不同
Figure 11 Revenue and profit can have different optima
合成教学情境 Wolfram 核算 q 等于 95 减 p 的十分之一 变动成本 80 固定成本 8000 不是实际酒店定价模型 Synthetic scenario not an empirical pricing model

表 10 合成酒店情境的计算结果
Table 10 Computed results for a synthetic hotel scenario

房价 Price	期望间夜 Expected nights	收入 Revenue	情境利润 Scenario profit
300	65	19500	6300
400	55	22000	9600

房价 Price	期望间夜 Expected nights	收入 Revenue	情境利润 Scenario profit
475	47.5	22562.5	10762.5
500	45	22500	10900
515	43.5	22402.5	10922.5
600	35	21000	10200

合成教学数据 Wolfram 核算 不是实际定价建议 Synthetic teaching data checked with Wolfram not an actual pricing recommendation

10 4 从单点图表到可执行试验

From a chart to an executable pilot

若把上述课堂方法用于酒店改进项目，第一步应选一个可控且后果有限的环节，例如内部知识查询或服务交接提示。先建立现状图，记录基线表现；再用 AI 提出改进图，明确预期机制；最后在小范围比较结果。指标可以同时包括完成时间、错误、顾客体验和员工负担，避免只优化其中一项而把成本转移给其他岗位。

实施记录应保存被否定的假设。例如团队最初以为资料搜索慢是主要问题，后来发现真正延迟来自审批职责不清。图的更新记录就能展示学习发生在哪里。对于没有改善的试验，也应分析是模型判断错误、资料不足、流程未执行，还是指标不合适。视觉思维的价值不在于为所有项目制造成功故事，而在于把成功或失败转化为可解释、可复用的经验。

An operational pilot should begin with a controllable, limited-consequence activity such as internal knowledge lookup or handoff prompts. Establish a current-state representation and baseline, use AI to propose a revised process with an explicit mechanism, then compare outcomes on a small scale. Measures can include completion time, errors, guest experience, and staff burden so that one improvement does not merely transfer costs to another role.

Rejected hypotheses deserve preservation. A team may initially blame slow search and later discover unclear approval responsibility. Revision history then shows where learning occurred. An unsuccessful pilot should distinguish poor model judgment, inadequate evidence, incomplete execution, and unsuitable measures. Visual thinking helps turn both success and failure into interpretable experience.

11 市场营销与文旅金融案例
Marketing and tourism finance cases

11 1 把营销漏斗改成可核查链条
Making a marketing funnel auditable

营销材料常用漏斗呈现曝光、访问、咨询、预订与实际消费，但漏斗本身不会解释损失发生的原因。一个严谨的营销图需要定义每一层的事件、时间窗口和分母，区分人数、次数与订单。例如同一用户多次访问不能自动视为多个潜在顾客，预订也不能自动等于完成入住。AI 可以帮助梳理字段，分析人员则必须确认数据采集方式是否支持这些定义。

课堂可以使用一个合成漏斗：10 万次曝光带来 5000 次有效访问，产生 200 笔预订，最后完成 150 笔入住。按这些口径，访问到预订转化率为 4%，预订到入住完成率为 75%。这些比例不能直接解释广告的因果效果，因为没有对照，也没有排除原本就会预订的人。图注应明确它只是观察或演示链条。学生需要提出额外证据，才能回答“广告是否真正带来新增住宿”。

视觉思维在这里有两个用途。首先，把定义不一致暴露出来，防止不同部门各用一套转化率。其次，将解释与观察分开：数据图显示在哪一层下降，假设图提出可能机制，实验设计再判断某个干预是否有效。三个步骤可以相互连接，却不能压缩为一张带箭头的“增长闭环图”。AI 越擅长生成营销叙事，人越需要检查这些层次。

Marketing funnels commonly show exposure, visits, inquiries, bookings, and consumption. A funnel does not explain why losses occur. A defensible representation defines each event, time window, and denominator, distinguishing people, events, and orders. Repeated visits by one person do not automatically represent several potential guests, and a booking is not a completed stay. AI can help organize fields, while analysts confirm whether collection practices support the definitions.

A synthetic classroom funnel might contain one hundred thousand exposures, five thousand qualified visits, two hundred bookings, and one hundred fifty completed stays. Under those definitions, visit-to-booking conversion is four percent and booking completion is seventy-five percent. These ratios do not establish the campaign's causal effect. There is no control condition and no exclusion of people who would have booked anyway. Students must identify additional evidence needed to estimate incremental stays.

Visual thinking first exposes inconsistent definitions and then separates observation from explanation. A data chart shows where a decline occurs, a hypothesis diagram proposes mechanisms, and an experiment investigates an intervention. These steps can connect, but they cannot be compressed into a persuasive growth-loop graphic that appears to prove everything. Stronger AI-generated marketing narratives increase the importance of checking these distinctions.

11 2 客群画像与用户旅程的边界
Boundaries of personas and journeys

AI 可以根据访谈、评价和服务记录提出客群主题，再整理为用户旅程或需求地图。但用户画像容易把推测画成真人，赋予没有证据的年龄、收入、动机和情绪。一个看似真实的头像，可能使团队忘记画像只是工作假设。建议把画像分为观察事实、合理推测和待验证问题，使用来源片段支持主要需求，并明确样本覆盖范围。

例如，评论中多次出现“到店后等待”，可以作为服务体验线索；进一步推断该群体都偏好某种高价套餐，就需要额外证据。旅程图可以标明观察到的接触点、记录的情绪表达与团队假设，避免把所有格子填满。AI 能够聚类与摘要，但小样本中的沉默需求、少数群体经验和服务失败的严重程度，仍可能被频次导向的总结压低。

教学可以要求学生为一张 AI 画像做证据审计。每个属性必须能连接到资料，或被清楚标为假设；删去没有证据的细节后，重新判断图是否仍能支持设计。这个练习通常能够揭示“细节多”与“信息可靠”之间的差异。更成熟的画像不一定更像一个完整的人，却应该更准确地表达我们知道什么以及还不知道什么。

AI can identify themes in interviews, reviews, and service records and organize them into journeys or need maps. Personas can nevertheless turn speculation into an apparently real person by adding unsupported age, income, motives, and emotions. A realistic portrait may conceal the fact that a persona is a working hypothesis. Observations, inferences, and unanswered questions should remain separate, with source excerpts supporting important needs and sample coverage made explicit.

Repeated comments about waiting after arrival provide a service clue. Inferring that the same group prefers an expensive package requires additional evidence. A journey can distinguish observed touchpoints, recorded emotional expressions, and team hypotheses instead of filling every cell. AI clustering and summaries may also underrepresent silent needs, minority experiences, or infrequent but severe failures.

Students can audit an AI persona by linking every attribute to evidence or labeling it as a hypothesis. After removing unsupported detail, they reassess what decisions remain justified. The exercise reveals the difference between detail and reliability. A better persona may look less like a complete individual while more accurately showing what is known and unknown.

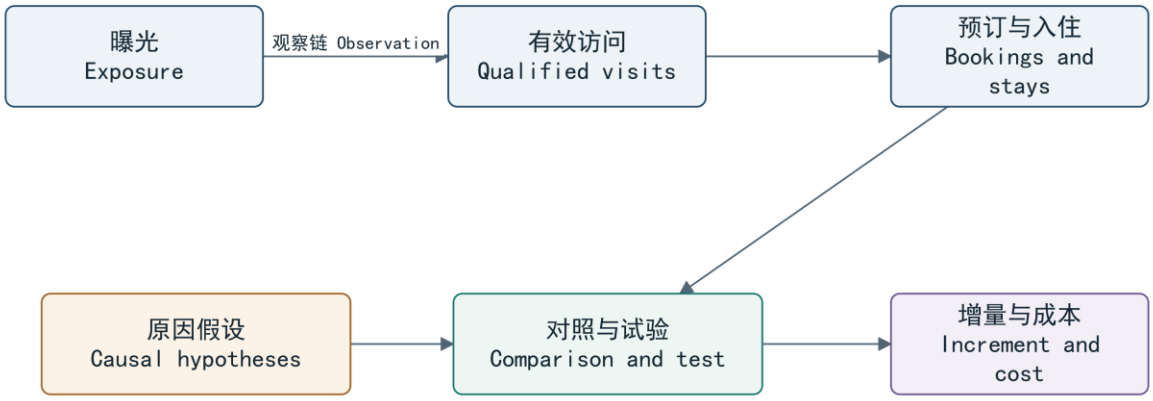


图 12 营销观察假设与因果检验
Figure 12 Marketing observation hypotheses and causal tests
本报告提出的概念与设计示意 Conceptual design proposed in this report

11 3 文旅金融中的现金流与情境图
Cash flow and scenarios in tourism finance

文旅与金融结合的课堂案例，可以围绕酒店改造、季节经营与项目现金流展开。视觉思维能够把初始投入、营运收入、固定支出、维护支出和融资安排放在同一时间线上。AI 可以帮助提取假设、生成情境表与计算草稿，计算工具负责执行。报告将此视为教育性建模方法，不对任何实际投资项目给出判断。

重要的是同时呈现收益与脆弱点。一个项目即使在平均情境下盈利，也可能因淡季现金不足而出现困难；一个短期收入很高的方案，也可能依赖不可持续折扣。情境图应明确基本、乐观和压力假设，避免只展示一条上升曲线。学生需要说明哪些变量来自资料，哪些是自己设定，以及变量之间是否相互独立。若把相关变量当成彼此独立，漂亮的情境结果仍可能失真。

对净现值、回收期等概念，先用现金流时间线帮助理解，再引入相应计算与限制。折现率、税费、融资结构和残值等假设会改变结果；不同指标回答的问题也不同。AI 应给出可检查的参数表和公式，而不应以“综合分析认为可行”替代计算。学生最终提交的应是有假设、有敏感性、有反例的模型说明，而不是投资口号。

A tourism-finance exercise can examine hotel refurbishment, seasonal operations, and project cash flows. Visual thinking places initial expenditure, operating receipts, fixed costs, maintenance, and financing on one timeline. AI can extract assumptions and draft scenario tables, while computational tools execute the model. This is an educational modeling approach rather than an assessment of any actual investment.

Returns and vulnerabilities should appear together. A project profitable under average conditions may still experience a seasonal cash shortfall. A high-revenue option may rely on unsustainable discounts. Base,

optimistic, and stress scenarios should make assumptions visible instead of presenting one rising line. Students identify which variables come from evidence, which are chosen, and whether dependencies have been represented. Treating correlated variables as independent can distort an otherwise attractive scenario display.

A cash-flow timeline can precede calculations of present value or payback. Discount rates, taxes, financing, and residual values affect results, and different measures answer different questions. AI should provide inspectable parameters and formulas rather than a general declaration of feasibility. The assessed artifact is a model explanation containing assumptions, sensitivity, and counterexamples.

11 4 跨领域任务需要共同词典
A shared dictionary across domains

酒店、营销和金融分别使用入住率、转化率、现金流、收入和利润等词语，跨领域团队若不统一定义，很容易把不同指标直接相连。建议在项目开始建立变量词典，记录名称、单位、时间粒度、统计口径和负责人。关系图中的每一个数量节点都引用词典编号，这样更新定义时能够检查所有相关图表，而不会只修正文中的某一处。

跨领域项目还需要分清描述、预测、决策与执行。描述显示过去，预测表达条件下的未来估计，决策比较目标与约束，执行改变现实状态。AI 可以参与每一步，但权限和验证要求不同。把四者画成不同阶段，能使学生理解“系统给出了预测”与“机构可以采取行动”之间仍有一段需要人负责的过程。

最终答辩可以围绕三问展开：你们的图改变了哪个判断；这个判断依据什么；如果关键假设不成立，方案怎样调整。这样，行业案例不仅训练工具使用，也训练跨领域沟通和责任意识。视觉思维提供共同结构，专业知识决定结构是否站得住，AI 则帮助降低表达与比较方案的成本。三者必须同时存在，才可能形成可靠应用。

Hotel operations, marketing, and finance use measures such as occupancy, conversion, cash flow, revenue, and profit. Without shared definitions, teams may connect incompatible quantities. A project dictionary should record names, units, temporal granularity, definitions, and ownership. Quantitative nodes refer to dictionary identifiers so that a changed definition can be checked across all related graphics.

Cross-domain work also needs to distinguish description, prediction, decision, and execution. Description concerns past observations. Prediction estimates future conditions. Decision compares objectives and constraints. Execution changes the world. AI can participate in each, but permissions and validation differ. Showing separate stages helps students understand the responsibilities between a forecast and authorized action.

A final defense can ask what judgment the representation changed, what evidence supports that judgment, and how the plan changes if an assumption fails. Industry cases then develop communication and responsibility alongside tool use. Visual thinking supplies shared structure, professional knowledge determines whether it is defensible, and AI lowers the cost of expression and comparison.

表 11 营销与文旅金融的判断边界
Table 11 Interpretation boundaries in marketing and tourism finance

观察或模型 Observation or model	可以回答 Supports	仍需补充 Needs more evidence
营销漏斗 Marketing funnel	各层定义下的转换 Conversions under definitions	新增效果与成本 Incremental effect and cost
客群主题 Customer themes	资料中反复出现的需求 Recurring documented needs	代表性与沉默需求 Representativeness and unspoken needs
用户旅程 User journey	接触点与假设痛点 Touchpoints and hypothesized friction	实际行为与严重程度 Behavior and severity
现金流情境 Cash-flow scenario	假设内的时间分布 Timing within assumptions	参数证据与外部变化 Parameter evidence and external change
敏感性图 Sensitivity chart	输入变化怎样传递 How input change propagates	结构错误与相关性 Structural error and dependence
行动建议 Action proposal	候选与验证路径 Candidates and testing route	授权资源与真实试验 Authority resources and field trial

本报告的综合分析与设计建议 Report synthesis and design recommendations

12 AI 导论课程的整体重构
Redesigning an introductory AI course

12 1 以可证明的能力组织课程
Organizing a course around demonstrable capabilities

面向酒店管理、市场营销和文旅金融学生的 AI 导论，可以以“用 AI 解决有边界的真实问题”为总体目标。学生需要形成四项相互联系的能力：把问题转成可分析结构，选择适当 AI 与计算工具，验证输出，以及把结果转为可解释行动。视觉思维贯穿其中，却不取代基础概念与必要的数量知识。课程应让学生能够讲清原理的最低要求，同时避免把所有时间用于技术细节。

课程地图可以分成问题层、资料层、工具层、流程层和评价层。问题层放业务目标，资料层放文本、表格与图像，工具层放语言模型、检索、计算与绘图，流程层显示协作与权限，评价层说明正确性、效果与限制。这样的地图比“工具依次升级”的同心圆更能反映真实工作，因为工具之间未必是高低关系，简单表格有时比复杂智能体更合适。

每个模块都应产出一种可以检查的能力证据。学习提示词后，学生能解释约束如何改变结果；学习数据分析后，能复算一个图表；学习 RAG 后，能判断资料适用性；学习智能体后，能指出工具失败与授权边界；学习视觉思维后，能在同一问题上比较不同表征。最终项目则把这些证据连接成完整链条，而不是突然要求学生做一个规模过大的产品。

For students in hospitality, marketing, and tourism finance, an introductory AI course can aim at solving bounded real problems with AI. Four connected capabilities are central: representing a problem, selecting appropriate AI and computational tools, validating outputs, and translating results into explainable action. Visual thinking runs through the course without replacing foundational concepts or necessary quantitative knowledge. Students need a minimum explanatory understanding while avoiding a course dominated by implementation detail.

A course map can contain problem, evidence, tool, workflow, and evaluation layers. Business objectives occupy the first, text and data the second, models and computational tools the third, collaboration and permission the fourth, and correctness and limitations the fifth. This is more faithful to work than a hierarchy in which tools simply become more advanced. A simple table may be more suitable than a complex agent.

Every module should produce inspectable evidence of capability. After prompting, learners explain how constraints alter outputs. After analysis, they reproduce a chart value. After retrieval, they judge source applicability. After agents, they identify tool failures and permission boundaries. After visual thinking, they compare representations of one problem. The final project connects these achievements rather than abruptly requiring an oversized product.

12 2 一个可调整的十六周框架

An adaptable sixteen week framework

十六周方案可以按四个阶段安排。第一阶段用前四周建立问题表达、AI 边界、资料与视觉语法；第二阶段用第五至第八周训练数据分析、检索与证据；第三阶段用第九至第十二周学习工作流、智能体、行业服务与情境推演；第四阶段用最后四周完成项目、评价、修订和展示。这是建议框架，学校课时、学生基础与学期安排应决定实际节奏。

每周保持一个核心问题、一项主要表征和一个可验证成果。例如数据周的核心问题是“这张图怎样由记录产生”，主要表征是字段表加图，成果是可复算说明；智能体周的问题是“什么时候应该停止或转人工”，主要表征是带条件流程，成果是异常测试。这样，学生不会只记住软件名称，而能把每周学习放进同一解决问题路径。

课程地图应动态更新，但不能每周无限增长。可以保留一张稳定总图，再用专题图展开。每次更新要求学生指出新增连接与原有知识之间的关系，并说明一个仍不理解的问题。教师由此判断班级是否真正建立结构。AI 可以为学生提供不同复杂度的地图，但评价仍围绕共同目标，而不是谁的地图节点更多。

A flexible sixteen-week design can use four phases. Weeks one to four establish problem representation, AI boundaries, evidence, and visual conventions. Weeks five to eight develop data analysis, retrieval, and verification. Weeks nine to twelve address workflows, agents, industry services, and scenarios. The final four weeks support project work, evaluation, revision, and presentation. This is a proposal to adapt to local schedules and prior knowledge.

Each week retains one central question, one principal representation, and one verifiable outcome. A data week asks how records produce a chart and yields a reproducible explanation. An agent week asks when a system should stop or escalate and yields exception tests. Students can then locate weekly learning within a consistent problem-solving process instead of remembering only application names.

The map should evolve without expanding indefinitely. A stable overview can link to topic-specific detail. Students explain new connections and identify one unresolved question at each revision. This helps an instructor assess structural understanding. AI may offer maps at different levels of complexity, while assessment remains tied to shared objectives rather than node counts.

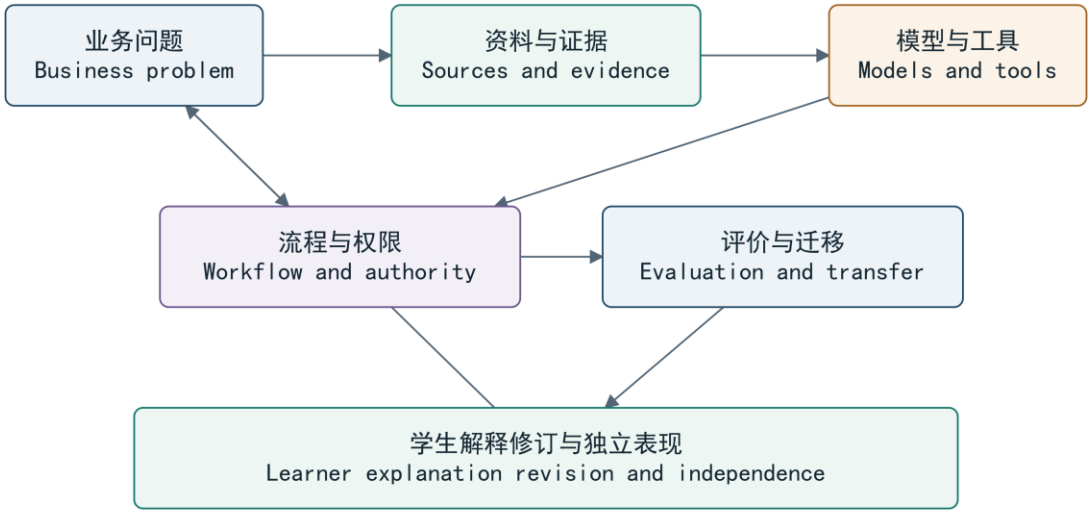


图 13 AI 导论的五层能力地图
Figure 13 Five layers of an introductory AI course
本报告提出的概念与设计示意 Conceptual design proposed in this report

表 12 十六周课程建议框架
Table 12 Proposed sixteen-week course

周次 Weeks	问题与内容 Question and content	能力证据 Evidence of capability
1 至 2 1 to 2	AI 能做什么与问题边界 AI capabilities and problem scope	问题图与能力边界说明 Problem map and capability limits
3 至 4 3 to 4	提示资料与表征语法 Prompts sources and visual grammar	图文互译与错图诊断 Translation and error diagnosis
5 至 6 5 to 6	数据定义与数量图 Data definitions and charts	可复算图表与字段词典 Reproducible chart and dictionary
7 至 8 7 to 8	检索与证据约束 Retrieval and evidence	带来源回答与冲突测试 Sourced answer and conflict test
9 至 10 9 to 10	工作流与智能体 Workflows and agents	权限失败与人工出口 Permission failure and escalation
11 至 12 11 to 12	行业蓝图与情境 Industry blueprints and scenarios	服务流程与假设比较 Service process and assumptions

周次 Weeks	问题与内容 Question and content	能力证据 Evidence of capability
13 至 14 13 to 14	项目试用与修订 Pilot and revision	测试记录与变化解释 Tests and justified changes
15 至 16 15 to 16	答辩与独立迁移 Defense and transfer	个人解释与复现材料 Individual explanation and reproduction

本报告的综合分析与设计建议 Report synthesis and design recommendations

12 3 把教师示范转成学生建构

From instructor demonstration to learner construction

视觉课程容易出现一种表面改革：教师制作更多图，学生仍只观看。要改变这一点，需要安排逐步移交。早期教师示范如何从原文提取节点，随后学生补全关系，再独立选择表征，最后对别人的图提出有依据的修订。每一步都应有短解释，防止活动变成描摹。AI 可以协助反馈，但不应让所有学生只接收同样成品。

课堂可以使用“预测再展示”“补图并解释”“错图诊断”“换条件重画”和“图文互译”五类活动。预测促使学生先形成模型；补图聚焦关系；错图诊断训练审查；换条件重画训练迁移；图文互译检查表征之间的对应。活动选择应取决于教学目标，不需要每节课全部使用。教师记录哪一种活动在哪个知识点帮助最大，逐步形成自己的证据。

学习支架需要随表现变化，而不是只按周数撤去。若学生仍不能区分数据与控制流，就保留箭头词汇表；若已经熟练，就要求自行确定类型。对工具操作困难的学生，可以提供结构化模板或纸笔入口，但仍要求解释相同关系。这样既降低不必要门槛，也保留核心学习要求。视觉思维成为可以发展的能力，而不是一种被预设的天赋。

A visually rich course can still leave students passive. Responsibility needs to transfer gradually. An instructor first models extraction from a source, students then complete relationships, later choose representations independently, and finally justify revisions to others' work. Short explanations accompany each stage so that activity does not become tracing. AI feedback can help without giving everyone an identical completed artifact.

Five useful activity types are prediction before reveal, completion with explanation, error diagnosis, redrawing under changed conditions, and translation between prose and graphics. They respectively support initial modeling, relational focus, scrutiny, transfer, and correspondence. A lesson need not include all five. Instructors can record which activities help with particular concepts and build local evidence.

Scaffolds should change with demonstrated performance. A learner who confuses data and control flow may retain an arrow vocabulary, while a proficient learner defines the types independently. Templates or paper entry can reduce interface barriers while preserving the same explanatory requirements. Visual competence becomes something to develop rather than a presumed talent.

12 4 评价构成与学习档案
Assessment and a learning portfolio

建议评价同时覆盖个人与小组、过程与成品、使用 AI 与不使用 AI 的表现。一个可调整方案是概念与短测 20%，视觉推理档案 25%，行业项目 35%，个人迁移答辩 20%。这些比例属于课程设计建议，并非研究证明的最佳比例。学校的考核制度和课程目标应决定最终分配。关键在于不能让小组精美成品掩盖个人理解不足。

学习档案可以保存四类证据：首次问题图、AI 反馈记录、修订说明和独立迁移任务。教师无需检查全部对话，只要求与关键概念变化有关的片段。学生还应声明使用了什么工具，以及哪些部分由自己确认。这样的透明记录既有助于诚信，也能训练职业协作中对来源和责任的说明。评价因此成为课程能力链的一部分，而不是最后单独加上的评分表。

Assessment should cover individual and group work, process and product, and performance with and without AI. One adjustable allocation is twenty percent concepts and short tests, twenty-five percent a visual-reasoning portfolio, thirty-five percent an industry project, and twenty percent an individual transfer defense. These are design suggestions rather than empirically optimal percentages. Local requirements and learning objectives determine the final allocation. The essential point is that a polished group artifact should not conceal weak individual understanding.

A portfolio can preserve an initial problem map, relevant AI feedback, a revision explanation, and an independent transfer task. The instructor need not inspect every conversation, only evidence related to important conceptual changes. Students identify their tools and what they personally verified. This supports integrity and professional communication about sources and responsibility. Assessment becomes part of the capability sequence.

13 课堂实施与完整课时方案
Classroom implementation and a complete lesson design

13 1 九十分钟课时的目标与准备
Objectives and preparation for ninety minutes

以“酒店知识问答中的 RAG 与人工升级”为例，九十分钟课时可设三个结果目标：学生能够用图解释资料检索与生成的区别，能够为资料冲突设计升级路径，能够用测试情境检查方案。课程不以完成某个商业平台的全部配置为目标，因为界面步骤容易占去概念学习时间。工具搭建可以作为延伸任务，课堂首先保证流程与证据关系得到理解。

教师提前准备一份短政策集、一张不完整流程图、三个测试情境、一份字段词典和一个可用的演示界面。政策集使用模拟或公开资料，不包含真实客户信息；其中故意保留两处需要判断的日期或适用范围差异。不完整图应缺少有教学意义的判断节点，而不是随机删字。演示系统需有静态截图与手工替代方案，网络或平台故障时仍能完成主要学习任务。

学生先做一个不超过三分钟的进入测验：用自己的话解释“模型回答流畅”和“回答有依据”的区别，并指出一份政策中的时间条件。结果帮助教师判断是否需要额外解释。测验只诊断基础，不据此给学生贴类型标签。对软件经验不足的学生，可以提供纸笔版；所有学生仍需完成相同的概念解释与测试判断。

A ninety-minute lesson on retrieval and escalation in hotel knowledge questions can target three outcomes. Learners explain the distinction between retrieval and generation, design an escalation route for conflicting evidence, and test their proposal with scenarios. Completing every configuration step on a commercial platform is not the primary objective. Interface work can consume conceptual learning time, so implementation may become an extension after evidence relationships are understood.

Preparation includes a short policy set, an incomplete process, three scenarios, a field dictionary, and an available demonstration interface. Policies use simulated or public information without real guest data. Two deliberate differences in dates or applicability create decisions worth discussing. The incomplete process removes meaningful judgments rather than random words. Screenshots and a manual alternative preserve the central activity if a platform or network fails.

A short entry assessment asks students to explain the difference between a fluent answer and a supported answer and identify a temporal condition in a policy. This diagnoses background knowledge without assigning cognitive types. Paper access can support learners with limited software experience while preserving the same conceptual and testing requirements.

13 2 从导入到首次建构
Introduction and initial construction

前十分钟展示两个同样流畅却依据不同的回答，让学生判断哪一个值得采用，并说明还缺什么信息。教师此时不立即公布标准图，而是把学生提出的核查需求写在白板上。接下来十五分钟，

用一个具体问题逐步演示问题解释、资料检索、适用范围检查和回答生成。每新增一个节点，都同时说明输入、输出和可能失败的情形，使图不是事后包装。

随后十五分钟，小组补全半成品图。每组必须补入一个判断节点和一个异常出口，并为两条连接命名。AI 可作为提问助手，提示遗漏，但暂不直接生成完整图。教师巡视时重点问“这个箭头传什么”和“这里依据什么决定”，而不是关注配色。小组完成后用一分钟解释最重要的修改；其他同学只能提出针对关系或证据的问题。

这一阶段应允许学生犯可修复的错误。例如把所有资料冲突都交给模型“综合”，正好可以引出适用范围和权限。教师不应急于把每组改成同一版图，而应保留两三种有代表性的结构，作为下一阶段测试材料。课堂由此形成比较对象，学生也能看到错误并非丢脸，而是检验模型的机会。

The first ten minutes present two equally fluent responses supported by different evidence. Students judge which can be used and identify missing information. Their checks appear on a whiteboard before a standard diagram is revealed. During the next fifteen minutes, an instructor builds a concrete process through question interpretation, retrieval, scope checking, and response generation. Each node is introduced with its input, output, and possible failure, making the diagram part of explanation rather than an after-the-fact decoration.

Groups then spend fifteen minutes completing a partial process. Each adds one decision, one exceptional exit, and labels for two relationships. AI may ask questions about omissions but does not immediately provide a complete solution. The instructor asks what an arrow carries and what evidence supports a decision. Groups explain their most consequential revision, while peer questions focus on relationships and evidence.

Repairable errors are useful. A proposal to let the model combine all conflicting policies can prompt discussion of scope and authority. The instructor can preserve several representative structures rather than normalize every group immediately. These become test objects in the next phase, showing that errors are opportunities for examination.

13 3 测试 修订与独立迁移
Testing revision and independent transfer

第四十至第六十分钟，各组交换图，并使用相同测试情境检查：资料缺失、资料日期冲突以及要求超出授权。测试者必须指出问题停在哪个节点，以及需要怎样改变流程。不能只说“模型会处理”。被测试组根据反馈修订，并记录一条被否定的假设。若系统可用，可运行少量演示；若不能运行，按流程逐步手工扮演也能检查许多结构问题。

第六十至第七十五分钟，教师展示一个参考方案，比较它与学生方案的差异。重点不是宣布唯一正确布局，而是明确哪些规则不可省略，哪些属于可选择设计。随后每个学生独立完成一个新情境：把酒店政策换成市场活动优惠规则，说明原流程可保留什么、必须更改什么。这个短任务让教师看到个人迁移，而不会被小组成员代劳所掩盖。

最后十五分钟用于复盘和退出测验。学生提交一句核心解释、一处修改及其依据、一项未解决问题，并用自己的话指出 AI 在此次活动中帮助了什么。教师根据答案安排下次课的支架。课后

延伸可以要求实现一个最小系统，但验收应包括异常测试和证据记录。技术完成与概念理解在成绩中分别记录，便于发现学生究竟卡在哪里。

From minute forty to sixty, groups exchange diagrams and apply shared tests involving missing information, conflicting dates, and requests beyond authority. Testers identify the node where the problem occurs and the required revision. Saying that the model will handle it is insufficient. The receiving group revises its process and records one rejected assumption. A working interface permits a few runs, while manual role-play can still expose many structural problems.

From minute sixty to seventy-five, the instructor compares a reference with student proposals. The discussion distinguishes indispensable rules from optional design choices rather than declaring one layout correct. Each learner then independently transfers the process to a promotional-policy scenario, explaining what can remain and what must change. This reveals individual understanding beyond group performance.

The final fifteen minutes support reflection and an exit assessment. Students submit a central explanation, a justified revision, and an unresolved question, along with a description of AI's contribution. The next lesson's scaffolds follow those responses. An extension may implement a minimal system, but acceptance includes exceptions and evidence records. Technical completion and conceptual understanding should remain separately visible in assessment.

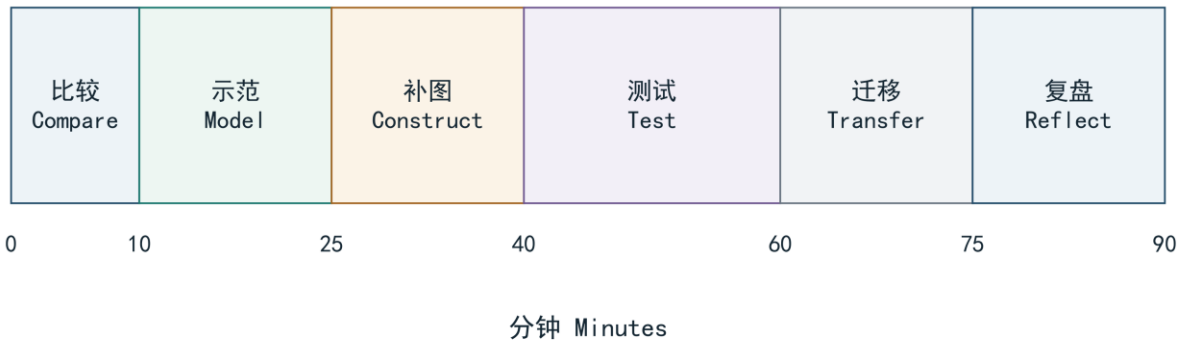


图 14 九十分钟课时的活动节奏
Figure 14 Activity sequence for a ninety minute lesson
建议教学时间 可按学生表现调整 Proposed timing adjustable to learner performance

表 13 九十分钟课时安排
Table 13 Ninety-minute lesson sequence

时间 Minutes	活动 Activity	可观察结果 Observable outcome
0 至 10 0 to 10	比较流畅但依据不同的答案 Compare fluent answers with different evidence	提出核查问题 Identify checks
10 至 25 10 to 25	教师逐步建构流程 Instructor builds the process	解释输入输出与失败 Explain inputs outputs and failure
25 至 40 25 to 40	小组补图并命名关系 Complete and label relations	一项判断与异常出口 A decision and exceptional exit
40 至 60 40 to 60	交换图并使用统一测试 Exchange and apply shared tests	定位问题与修订 Locate issues and revise
60 至 75 60 to 75	参考比较与个人迁移 Reference comparison and transfer	新条件下的合理变化 Justified change under new conditions
75 至 90 75 to 90	复盘与退出测验 Reflection and exit assessment	解释修改与未知 Explain revisions and uncertainty

13 4 四十五分钟与不同基础的适配
Adapting to forty five minutes and prior knowledge

压缩到四十五分钟时，应保留问题、建构、测试和迁移的完整链条，而不是删去验证只留下演示。可以减少政策条数，把小组建构改为全班共同补图，保留一个异常情境和一个个人解释。界面配置移到课前或课后。这样缩短的是材料量与操作量，而不是核心思考环节。教师可根据学生基础决定是否预先提供变量词典或来源标记。

对有编程基础的班级，可以增加结构验证或自动测试；对商科入门班级，则重点训练变量口径、证据判断和流程边界。二者不应被理解为高低层次，而是不同课程目标。若学生已经能熟练画图，教师应提高解释和迁移要求；若学生画图困难，则先提供有限节点和关系选项。教学调整依据实际表现，而不是一套固定的“视觉型学生教材”。

课堂之后还应检查材料是否产生了意外误解。学生是否把 RAG 当成数据库本身，是否认为智能体具有默认行动权限，是否认为带引用的回答一定正确。教师可以把这些常见误解转成下一次错图诊断题。AI 可辅助生成题目，但教师必须确认错误位置与概念目标相符。持续收集这种小证据，比单次展示一套完美教案更能推动课程改进。

A forty-five-minute adaptation should retain the sequence of question, construction, testing, and transfer. Reduce the number of policies, replace some group construction with shared completion, and preserve one exception and one individual explanation. Interface configuration moves outside the lesson. Material and operation volume decrease while the essential reasoning remains. Dictionaries or source markers can be supplied according to prior knowledge.

A programming-oriented class may add structural validation and automated tests. A business introduction may emphasize definitions, applicability, and process boundaries. These are different objectives rather than a hierarchy of worth. Proficient diagram makers need more demanding explanations and transfer tasks; learners struggling with representation can begin with limited nodes and relation options. Adaptation follows performance rather than fixed learner categories.

After teaching, check for unintended misconceptions: equating retrieval augmentation with the database, assuming an agent has default authority, or believing that citations guarantee correctness. These can become the next lesson's error-diagnosis items. AI can draft them, but the instructor confirms the conceptual target. Accumulating such evidence supports continuing course improvement.

14 Skill 与专家协作体系
Skills and expert role collaboration

14 1 把附件中的生产思路变成可检查规范
Turning attachment proposals into specifications

附件提出 Skill、专家与专家团的递进结构，适合用来组织课程生产职责。但这些名称本身不保证能力或质量。一个 Skill 可以是可复用指导、脚本、模板或工具组合；一个专家角色可以是承担特定责任的配置；专家团则是一种协作工作流。报告使用这些词描述架构，不据此宣称某个平台当前具备全部接口，也不认为角色名称意味着真实专业资格。

要把设想落地，应为每项能力写输入、输出、约束、依赖和验收标准。课程结构能力输入目标与资料，输出知识关系和活动安排；绘图能力输入已确认结构，输出图形与源文件；验证能力输入图与证据，输出具体问题和修改建议。若只有“你是一位优秀专家”这样的角色描述，系统仍可能在关键细节上任意发挥。职责应由可检查的产物定义。

附件中的固定规则可以改为默认值加例外条件。例如单卡文字限制可帮助控制密度，但概念复杂时应允许展开；课件页数可用于初步规划，却不能替代课时节奏；作业要求交图可以训练表征，但仍应允许必要的文字、计算和访问替代。规则的价值是稳定质量，而不是让所有课时具有相同外观。教师应记录何时偏离默认值及其理由。

The attachments propose a progression from skills to expert roles and expert teams. This can organize production responsibilities, but the names do not guarantee competence or quality. A skill may contain reusable guidance, scripts, templates, or tools. An expert role is a configuration with a defined responsibility. A team is a collaboration workflow. These terms describe architecture here; they do not establish current support by a particular platform or confer professional qualifications.

Implementation requires inputs, outputs, constraints, dependencies, and acceptance criteria. A course-structure capability receives objectives and sources and produces relationships and activities. A drawing capability receives confirmed semantics and produces graphics and editable sources. A validation capability receives artifacts and evidence and returns specific findings. Merely describing a role as excellent leaves important behavior underspecified. Inspectable products should define responsibility.

Fixed attachment rules can become defaults with exceptions. A card-length limit can manage density while allowing expansion for difficult concepts. Slide counts can support planning without replacing lesson timing. Requiring a diagram can develop representation while preserving necessary prose, calculations, and accessible alternatives. Rules should stabilize quality rather than impose identical appearance on every lesson. Deviations need reasons.

14 2 单一工作流与多角色分工的选择
Choosing a single workflow or divided roles

不是每次备课都需要多个智能体。若资料较少、目标清楚，一个模型配合确定性绘图与检查步骤，可能更易管理。若任务包含学术核查、行业案例、视觉设计与教学评价，分角色可以使责任更清楚。但角色越多，传递信息、统一术语和解决冲突的成本也越高。应比较完整工作流的质量、时间和返工，而不是把角色数量当作先进程度。

可采用四类职责：课程架构负责目标与概念，行业审查负责场景真实性，表征设计负责图形与可读性，质量审查负责来源、计算和评价一致性。职责可以由人或 AI 承担不同部分。关键审查应尽可能使用独立材料和规则，避免所有角色共享同一个错误假设后互相赞同。课程负责人保留最终解释责任，并能追溯每项重要修改的来源。

协作流程也不应只有线性传递。行业审查可能要求回到问题定义，质量审查可能发现需要更换图形，学生测试可能推翻活动安排。建议把返工路径画出来，明确谁可以提出修改、谁确认影响、哪些产物需要同步更新。这样专家协作成为可迭代工程，而不是一次从输入到成品的理想化流水线。

Not every lesson needs several agents. With limited sources and clear objectives, one model plus deterministic rendering and checks may be easier to manage. Tasks combining scholarly verification, industry cases, visual design, and assessment may benefit from differentiated responsibilities. More roles also increase communication, terminology, and conflict-resolution costs. Overall quality, time, and rework should determine the choice rather than the number of roles.

Four responsibilities are useful: course architecture for objectives and concepts, industry review for scenario validity, representation design for visual form, and quality review for sources, calculations, and assessment alignment. People and AI can share these responsibilities. Important review should use independent evidence and rules where possible so that shared assumptions do not simply receive repeated agreement. A course owner retains responsibility for interpretation and can trace important revisions.

The process should include returns. Industry review may change the problem definition, quality review may require a different representation, and student testing may overturn an activity plan. Rework routes specify who proposes a change, who confirms its effects, and which artifacts update together. Collaboration becomes an iterative process rather than an idealized one-way production line.

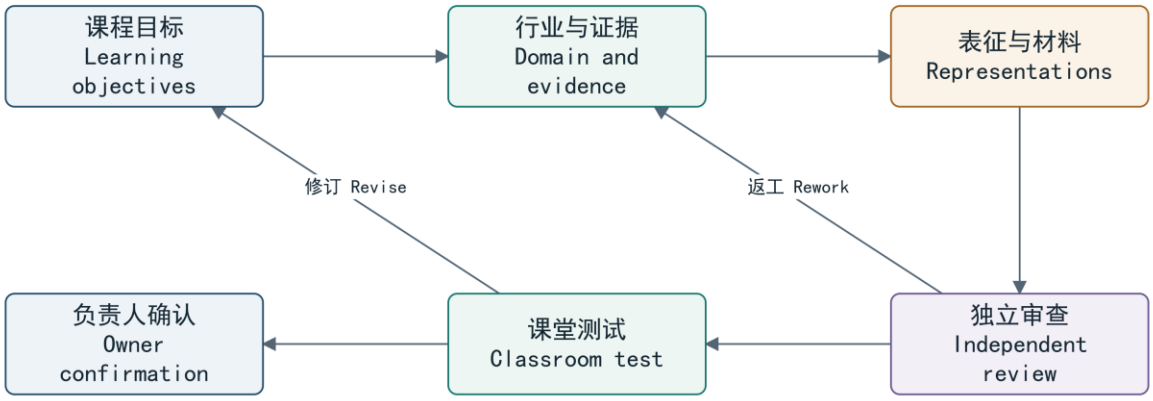


图 15 课程生产职责与返工路径
Figure 15 Production responsibilities and rework
本报告提出的概念与设计示意 Conceptual design proposed in this report

14 3 一份生产合同应包含什么

Contents of a production contract

课程生产合同可以从五组字段构成。目标字段说明受众、先验知识、学习结果与课时；资料字段说明来源、有效期、许可与缺口；结构字段说明概念、关系与图形选择理由；交付字段说明 PPT、手册、互动和作业之间的对应；质量字段说明事实、可读性、异常与评价检查。合同不必很长，但要足以让不同步骤共享同一任务定义。

每项输出还应携带状态。草稿允许未确认内容存在，审核稿要求关键关系已有依据，课堂版要求演示与替代方案可用，归档版则保留源文件和修改记录。不要只用“最终版”覆盖所有中间状态。状态可以帮助系统决定哪些动作可以自动执行、哪些需要课程负责人判断。一次新的政策或工具界面变化，也应触发相关输出退回待复核状态。

对于中英文材料，生产合同应要求术语表、共享编号与数值一致性。中文“收益”在不同情境可能对应 revenue、return 或 yield，不能由模型随意选择。图题、表头和正文应使用一致译法；如果一个词具有多个含义，应按变量定义分别处理。双语质量检查不能只看语法流畅，还要检查逻辑条件、否定和不确定性是否被保留。

A production contract can contain five field groups. Objectives specify audience, prior knowledge, outcomes, and time. Sources specify provenance, validity, permission, and gaps. Structure specifies concepts, relationships, and reasons for representation choices. Deliverables connect slides, manuals, activities, and assignments. Quality fields define factual, readability, exception, and assessment checks. The contract can be concise while giving all stages one task definition.

Outputs also need states. A draft permits unresolved content. A reviewed version requires support for important relations. A classroom version requires a usable demonstration and fallback. An archive preserves sources and revisions. One final label should not erase these distinctions. States help determine what can proceed automatically and what needs judgment. A changed policy or interface can return related outputs to review.

Bilingual work needs terminology, shared identifiers, and numerical consistency. The Chinese term for returns or revenue can require different English words in different contexts. Labels, headings, and prose should follow variable definitions. Review must examine conditions, negation, and uncertainty as well as fluent grammar.

14 4 如何检验是否真的提高效率
Testing whether efficiency actually improves

效率试验应记录完整成本，而不是只计模型生成时间。至少分别记录准备资料、提出目标、生成候选、人工修改、事实核查、视觉检查、课堂试用和返工。比较时使用相近难度的课时与相同质量标准，保留未使用 AI 的基线，避免把首次建立模板的时间与长期复用的时间混为一谈。若系统减轻了排版却增加了核查，结果应如实报告。

质量也应有独立指标，例如关键事实错误、错误连线、学生误解、可编辑性和不同交付物之间的不一致。时间下降但严重错误上升，不能简单宣布成功。一个更有用的结果可能是：常规课时准备更快，复杂概念仍需教师大量参与；或者图表一致性提高，但行业案例需要额外核实。这样的条件性结论比“全自动十分钟完成”更能指导持续使用。

长期改进可建立问题库，把每次发现的错误转成可复用检查规则。例如“过期政策未标日期”“把因果与顺序混用”“图表单位丢失”。规则应经过实际错误验证，避免无限扩张成难以执行的清单。课程生产系统由此积累的是经检验的工作知识，而不仅是更多提示词。人仍然参与改进，但参与位置更清楚、价值更集中。

An efficiency study should record the entire cost, not only model-generation time. Separate source preparation, objective specification, candidate generation, human revision, fact-checking, visual review, classroom testing, and rework. Compare lessons of similar difficulty under common quality criteria and preserve a non-AI baseline. Initial template setup should be distinguished from repeated use. If layout time falls while verification rises, both belong in the result.

Quality needs independent measures such as important factual errors, incorrect edges, learner misconceptions, editability, and inconsistency across deliverables. Faster production with more serious errors is not a simple success. A conditional finding may be more useful: routine lessons become faster while difficult concepts still require substantial instructor involvement, or visual consistency improves while cases need additional checks.

A maintained error library can turn recurring failures into reusable checks: missing effective dates, causal arrows confused with sequence, or absent units. Checks should respond to observed errors rather than expand indefinitely. The system then accumulates tested working knowledge alongside prompts, with human participation concentrated where it adds value.

表 14 课程生产职责与验收
Table 14 Production responsibilities and acceptance

职责 Responsibility	主要产物 Artifact	验收 Acceptance
课程架构 Architecture	目标概念与活动对应 Objectives concepts and activities	活动能证明目标 Activities demonstrate objectives
行业审查 Domain review	真实规则与场景边界 Rules and scenario boundaries	事实与授权明确 Facts and authority clear
表征设计 Representation	图形规范与源结构 Graphics and source structure	关系可读且可维护 Readable maintainable relations
资料核对 Source review	来源日期与适用范围 Origins dates and scope	重要判断可追溯 Important claims traceable
质量检查 Quality	错误记录与最小修改 Findings and minimal repairs	数量双语与异常一致 Numbers languages and exceptions aligned
课程负责人 Course owner	确认与发布版本 Confirmed release	承担解释与更新责任 Own interpretation and updates

本报告的综合分析与设计建议 Report synthesis and design recommendations

15 评价实验与研究设计
Evaluation experiments and research design

15 1 区分制作质量 理解与迁移
Separating product quality understanding and transfer

评价视觉思维至少包含三个层面。制作质量关注图是否准确、清晰、可访问与可维护；理解关注学习者是否能解释图中关系、识别条件并发现错误；迁移关注能否在新场景中选择或修改表征。三者有联系，却不能互相代替。一张由 AI 生成的高质量图，可能出自理解有限的学生；一个尚不熟练排版的学生，也可能提出准确而有洞察的结构。评分应把这些维度拆开。

可以使用分析性量规，每个维度给出可观察描述。语义准确看关键节点与连接是否符合资料；推理完整看条件、例外与反例是否处理；证据可追溯看主要判断能否定位；表达可读看编码、层级和文本是否清楚；迁移看新条件下修改是否合理。美观可以作为表达的一部分，但不应压过证据与推理。评分例子应同时包含漂亮错图和朴素正确图，帮助学生理解标准。

评价还应检查解释能否脱离成品成立。教师可以随机选一条连接，要求学生说明其来源、作用和删除后的后果；也可以提供一个新的资料片段，要求现场修改。这样的短答辩不需要很长，却能揭示图是否成为学生自己的理解。对大班课程，可以采用小样本抽查与全班独立小测结合，避免评价成本失控。

Evaluation has at least three levels. Production quality concerns accuracy, clarity, accessibility, and maintainability. Understanding concerns explanation of relationships, conditions, and errors. Transfer concerns selecting or revising a representation in a new situation. The levels relate but cannot substitute for one another. An excellent AI-generated graphic may be submitted by someone with limited understanding, while a visually rough sketch may encode accurate and insightful structure. Assessment should separate them.

An analytic rubric can describe observable performance. Semantic accuracy checks important nodes and edges against evidence. Reasoning completeness checks conditions, exceptions, and counterexamples. Traceability checks whether important claims can be located. Readability checks encoding, hierarchy, and text. Transfer checks revisions under changed conditions. Appearance belongs within communication, but should not outweigh evidence and reasoning. Examples should include attractive errors and plain correct work.

Understanding should also survive beyond the finished artifact. An instructor can select an edge and ask for its source, role, and consequences if removed, or provide a new source passage for immediate revision. A short defense reveals ownership of the reasoning. Large classes can combine sampled defenses with independent short assessments to control cost.

15 2 一个可检验 AI 代画与 AI 反馈的实验

An experiment on AI generation and feedback

若要研究 AI 怎样支持视觉思维，可以比较四个条件：学生阅读教师提供的图；学生独立画图；学生获得 AI 成图后解释修订；学生先画再获得 AI 问题式反馈。所有条件使用相同学习内容、相近总时间和共同测验。研究假设可以预先写明：AI 成图可能提高制作速度，问题式反馈可能更有利于关系解释，但这些只是待检验假设，不应在设计阶段宣布结果。

主要结果可以设为新情境中的关系推断或迁移表现，次要结果包括知识保持、用时、认知负荷、图的准确性和自信校准。立即测验与一周后的延迟测验分别反映不同问题。没有 AI 的短迁移任务尤其重要，因为最终目标是人的能力。若只评价使用 AI 时的成品，就无法区分能力提高与工具代劳。研究还应记录学生实际做了哪些修改。

随机分配条件可以降低部分混杂，但仍需考虑班级、教师、先验知识与合作关系。若只能按班级实施，应报告班级层面的限制与分析方法。样本量不应凭习惯指定，可依据有意义的最小差异、测量可靠性与聚类结构进行功效分析。本报告不提供一个脱离这些条件的通用样本数，也不把探索性课堂活动包装成已完成的随机试验。

A study of AI-supported visual thinking can compare four conditions: studying an instructor-provided diagram, drawing independently, explaining and revising an AI-generated diagram, and drawing first followed by question-based AI feedback. Content, approximate total time, and assessment remain comparable. Preregistered hypotheses might predict faster production from direct generation and stronger relational explanation from feedback. These remain hypotheses rather than results announced in advance.

A primary outcome could be relational inference or transfer in a new setting. Secondary outcomes include retention, time, cognitive load, graphic accuracy, and confidence calibration. Immediate and delayed testing address different questions. An unaided transfer task is particularly important because the intended outcome concerns human capability. Product quality during AI use alone cannot distinguish learning from tool substitution. Actual revisions should also be recorded.

Random allocation addresses some confounding, but classes, instructors, prior knowledge, and collaboration still matter. Class-level implementation needs corresponding limitations and analysis. Sample size should follow the smallest meaningful difference, measurement reliability, and clustering through a power analysis. There is no context-free universal sample size, and an exploratory classroom activity should not be presented as a completed randomized trial.

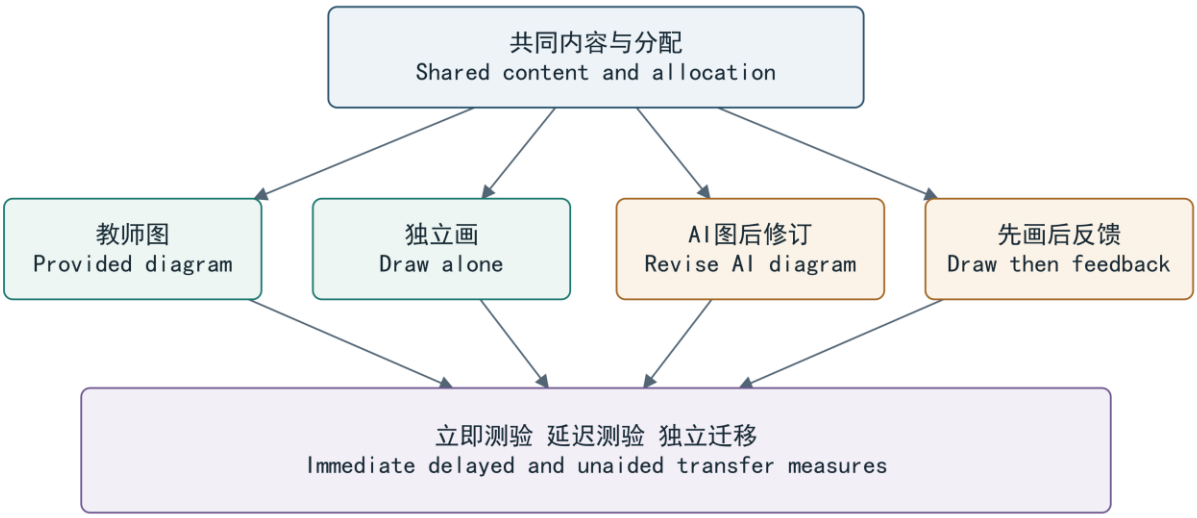


图 16 AI 辅助绘图的四条件研究设计
Figure 16 Four conditions for studying AI assisted diagrams
本报告提出的概念与设计示意 Conceptual design proposed in this report

15 3 测量过程与识别机制
Measuring processes and mechanisms

过程数据可以帮助解释为什么某种条件有效或无效。记录学生阅读来源、修改节点、请求提示和回看反馈的顺序，可能显示他们是否进行整合。记录不应无限收集个人行为，而应围绕明确研究问题，经过适当告知和权限安排。对普通教学改进，保留少量关键版本与解释通常已很有价值，未必需要复杂眼动或全程监控。

机制解释应谨慎。例如修订次数与成绩相关，可能因为能力强的学生更愿意修订，也可能因为修订帮助学习，不能仅凭相关决定方向。若要检验反馈是否通过提高关系准确性影响迁移，需要适当设计与分析，并承认中介推断的条件。AI 日志的丰富性不等于因果证据的充分性。报告建议先建立可靠测量，再追求复杂模型。

自信校准是值得加入的指标。学生在每次判断后估计自己有多确定，再与正确性比较。AI 可能使成品看起来更完整，从而提高信心，却未同步提高准确性。比较这一差距，比笼统询问学生是否喜欢 AI 更能发现问题。教师还可以观察学生是否能说出“我不知道哪条关系成立”，把适当的不确定性视为专业判断的一部分。

Process evidence can explain success or failure. The order of source reading, node revision, hint requests, and feedback review may show whether integration occurred. Collection should serve a clear question with appropriate information and permissions rather than gather behavior indefinitely. For routine course improvement, a few important versions and explanations may suffice without elaborate monitoring.

Mechanistic inference requires caution. A correlation between revisions and achievement may mean that stronger learners revise more, that revision supports learning, or both. Testing whether feedback affects

transfer through relational accuracy requires suitable design and assumptions. Rich logs do not automatically provide causal evidence. Reliable measurement should precede complicated explanatory models.

Confidence calibration is also useful. Learners estimate certainty after a judgment, then compare it with accuracy. AI may make work look complete and raise confidence without a matching accuracy gain. The gap can reveal more than a general preference survey. The ability to identify an uncertain relationship should count as part of professional judgment.

15 4 分析 报告与复现
Analysis reporting and reproduction

结果报告应同时给出估计、区间与任务描述，不只列显著与否。若某种方法对低基础学生帮助较大，应说明这一分析是否预先计划，避免从许多分组中挑出偶然结果。图表应展示合适的分布或个体变化，不把平均值当作所有学生的经历。没有显著差异也不等于完全等效，需要结合测量精度和研究设计解释。

研究包应保存匿名化任务、教学材料、评分规则、分析代码、图表数据和版本说明。不能公开的数据应说明限制与可替代材料。模型名称、使用日期、提示模板和关键参数可能影响结果，也应记录；但无需公开学生无关私人对话。双语研究材料应检查任务难度是否因翻译变化，必要时采用语言熟悉度作为解释条件。

最后，课堂证据应进入下一轮改进。若 AI 反馈提高准确性却耗时过长，可以缩短反馈并保留关键问题；若学生成图更快但迁移下降，应增加独立建构；若对部分任务无明显价值，就允许不用 AI。课程研究的目标是找到条件与机制，而不是证明所有新工具都有效。视觉思维提供的可见过程，恰好有助于形成这种可修订的教学知识。

Reports should present estimates, intervals, and task descriptions rather than significance labels alone. A larger effect for learners with less prior knowledge should be identified as planned or exploratory. Graphics should show suitable distributions or individual change instead of treating the mean as everyone's experience. A nonsignificant result does not establish equivalence without considering precision and design.

A reproducibility package can contain anonymized tasks, instructional materials, rubrics, analytical code, chart data, and versions. Restricted data need a stated limitation and possible substitutes. Model names, dates, prompt templates, and consequential parameters may affect outcomes and should be recorded, while irrelevant private conversations need not be disclosed. Bilingual tasks require checks that translation has not changed difficulty.

Findings should guide the next iteration. Accurate but slow feedback may need shortening. Faster production with weaker transfer may require more independent construction. Some tasks may not benefit from AI. The aim is to identify conditions and mechanisms rather than establish universal value for new tools. Visible processes support this revisable educational knowledge.

表 15 视觉推理分析性量规
Table 15 Analytic rubric for visual reasoning

维度 Dimension	基础表现 Basic	较强表现 Stronger
语义准确 Semantic accuracy	主要节点可识别 Main nodes identifiable	关系与资料一致 Relations match evidence
推理完整 Reasoning	正常路径大致连贯 Normal route coherent	例外条件反例得到处理 Exceptions and counterexamples handled
证据 Evidence	给出总体来源 General sources supplied	关键元素可定位依据 Important elements traceable
可读性 Readability	文字与结构可辨 Text and structure legible	编码服务任务且可访问 Task-serving accessible encoding
修订 Revision	能够修改外观 Can change appearance	解释概念变化与影响 Explains conceptual effects
迁移 Transfer	复现已学案例 Reproduces learned case	新条件下合理调整 Adapts to new conditions

本报告的综合分析与设计建议 Report synthesis and design recommendations

16 风险治理与实施路线
Risk management and implementation

16 1 视觉幻觉与错误的权威感
Visual hallucination and apparent authority

AI 视觉错误不只包括物体画错，还包括数量失真、连接缺失、标签反转、比例误导和来源混淆。图形的整洁与专业感可能使这些错误更难被质疑。ChartQA 和 MathVista 等基准把图表理解与视觉数学问题分成可测试任务，为能力评估提供了具体对象。[24][26]但任何一个基准都只覆盖部分任务，不能作为真实图表全部可靠的保证。

针对细致空间信息的研究还报告了视觉语言模型在简单几何关系上的困难。[25]课程可以据此安排“低级但关键”的检查，例如数清节点、确认相交与相连、检查箭头方向和比对图中数值。这些工作看似不如让模型写总结高级，却直接关系到图的正确性。尤其在复杂关系网中，一个误连可能改变整个解释，而模型自信的描述无法弥补它。

治理应从错误后果出发。教学示意图可以允许草稿试错，但进入公开课程材料前需核查；内部探索图可以带假设使用，但用于经营决策前需更强证据；涉及自动执行的图还需要权限和回退。不同风险不宜共用同一个审核门槛。报告建议把错误严重度、可发现性和可逆性作为判断依据，确定哪里需要人、规则或独立计算。

AI visual errors include distorted quantities, missing connections, reversed labels, misleading proportions, and confused sources, as well as incorrectly depicted objects. Professional appearance may make errors harder to challenge. ChartQA and MathVista define testable chart-understanding and visual-mathematical tasks.[24][26] Their scope remains partial, so benchmark performance cannot guarantee reliability across real graphics.

Research on fine spatial information also reports difficulties for vision-language models on simple geometric relationships.[25] Instruction should include basic but consequential checks: counting nodes, distinguishing intersection from connection, confirming arrow direction, and matching numbers. These checks may appear less sophisticated than generated summaries, yet a single incorrect edge can change an explanation. Confident prose cannot repair it.

Review should follow consequences. Draft teaching graphics can support experimentation but need checks before distribution. Internal exploratory diagrams can retain hypotheses but need stronger evidence before decisions. Executable workflows additionally need permissions and recovery. Severity, detectability, and reversibility help determine where human review, rules, or independent calculation are appropriate.

16 2 数据 版权与来源责任
Data copyright and provenance responsibility

图文材料可能包含顾客记录、内部政策、未公开研究、品牌资产和第三方图片。AI 使用前应明确资料是否允许输入所选服务，输出是否允许传播，以及是否需要去标识化。课程可以优先使

用公开或合成资料，避免为了演示效果上传不必要的个人信息。具体许可与法律问题应按机构和适用规则处理，本报告不把一般建议当成个案法律结论。

来源责任也不能被“AI 生成”四个字替代。研究图中使用的数值仍需引用，改绘别人的结构仍需说明依据，合成图片仍不应冒充现场记录。C2PA 规范提供内容来源与完整性相关机制，但有效凭证并不证明内容所表达的事实为真。[38]来源可信、文件未改动和解释正确，是不同问题，课程应让学生能够分别说明。

对学生项目，建议建立简单来源表，记录素材、用途、来源链接、处理方式和可分享范围。报告正文只放与理解有关的来源，完整记录放在附件或课程档案。教师不需要把所有作品变成法律审查项目，但应让专业习惯在小任务中形成。知道何时需要询问资料拥有者，比事后笼统声明“仅供学习”更有实际作用。

Visual materials may contain guest records, internal policies, unpublished research, brand assets, or third-party images. Before AI use, determine whether the selected service may receive the material, whether outputs may be shared, and whether de-identification is needed. Public or synthetic classroom data can avoid unnecessary personal information. Specific permissions and legal questions require applicable institutional rules; general guidance here is not a case-specific legal conclusion.

An AI-generated label does not remove source responsibility. Research numbers still need citations, adapted structures need attribution, and synthetic images must not impersonate field records. C2PA provides mechanisms concerning provenance and integrity, but a valid credential does not establish that depicted claims are true.[38] Trustworthy origin, unchanged files, and correct interpretation remain distinct.

Student projects can maintain a concise source register with purpose, link, processing, and sharing scope. Important sources appear in the report and full records remain in the portfolio. Small tasks can develop professional habits without making every assignment a legal audit. Knowing when to consult a material owner is more useful than relying on a general educational-use disclaimer afterward.

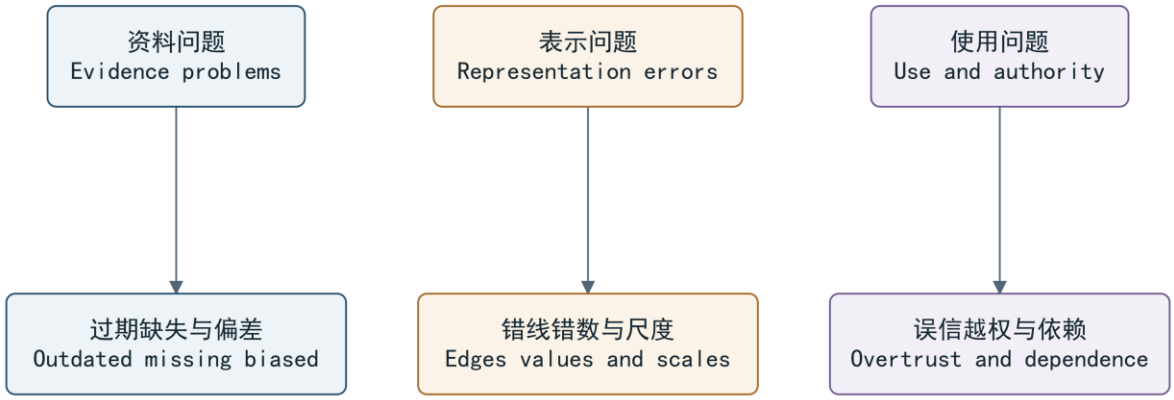


图 17 视觉 AI 风险的三个来源
Figure 17 Three sources of risk in visual AI
本报告提出的概念与设计示意 Conceptual design proposed in this report

16 3 依赖 可访问性与组织负担

Dependence accessibility and organizational burden

AI 辅助视觉思维可能降低绘图门槛，也可能形成新的依赖：学生不会在没有模型时画出基本关系，教师无法维护自动生成的复杂文件，机构被锁在某个平台格式中。治理不只是防止错误，还要保留自主性。建议核心结构采用可导出的文本、表格或开放图形格式，重要教学活动保留纸笔与静态替代，关键知识通过独立任务检验。

UNESCO 的生成式 AI 教育指导强调人的能动性、包容性与教育目标。[31]在本报告中，这转化为三个实施判断：工具应服务学习目标，师生应知道哪些决定由系统提出，无法使用某种界面的人仍应有可行路径。工具成本、网络、账号和设备也属于课程条件；不能把无法登录误判为不努力或能力不足。视觉教学需要同时审视内容与获得内容的条件。

组织负担还包括维护、培训与支持。一个系统可能在演示时表现出色，却需要每周修复模板、更新界面截图和处理文件兼容。实施计划应给这些活动分配负责人和时间。若没有维护资源，选择更简单、可手工修订的方案可能更合理。可持续性通常来自合适复杂度，而不是最大自动化程度。

AI may lower drawing barriers while creating new dependencies: learners cannot reconstruct basic relationships without it, instructors cannot maintain complex generated files, or institutions become tied to one format. Governance should preserve autonomy as well as prevent errors. Core structures need exportable text, tables, or open graphics, important activities need static or paper alternatives, and independent tasks should check central knowledge.

UNESCO guidance emphasizes human agency, inclusion, and educational purposes.[31] In practice, tools should serve objectives, users should know which decisions the system proposed, and people unable to use an interface should retain a viable route. Cost, connectivity, accounts, and devices are instructional conditions. Failure to log in should not be interpreted as low effort or ability.

Maintenance, training, and support are additional burdens. A convincing demonstration may still require weekly template repairs, screenshot updates, and compatibility work. Responsibility and time should be assigned. Without maintenance resources, a simpler, manually revisable system may be preferable. Sustainability depends on appropriate complexity.

16 4 十二周实施与退出条件
A twelve week pilot and exit conditions

一个十二周试点可以分为三阶段。前四周确定目标、选择低风险任务、整理资料和基线；中间四周生成最小材料、进行课堂或内部试用、记录错误；最后四周比较质量与时间、修订规范、决定继续、调整或退出。试点不应以“必须证明有效”为前提，而应允许发现某些任务不值得自动化。每阶段都应有具体产物，而不是只报告平台已经开通。

建议设定明确退出条件：关键错误持续无法降低，维护成本超过可接受范围，学生独立表现下降，资料权限无法满足，或使用者无法理解系统输出。退出并不意味着视觉思维没有价值，可能只意味着所选实现不合适。保留结构源文件与教学经验，就能在更简单工具中继续使用。这是避免技术选择绑架教育目标的重要安排。

扩展时先复制经过验证的任务模式，再增加领域和复杂度。一个酒店政策案例稳定后，可以转向活动优惠，再考虑更复杂服务；不要从一个成功演示直接推断所有课程可全自动生产。每次扩展都重新检查证据、权限与评价。这样形成的系统增长较慢，却更容易被教师理解、维护与信任。

A twelve-week pilot can have three phases. The first four weeks define objectives, select a limited-consequence task, organize sources, and establish a baseline. The next four create minimal materials, test them, and record errors. The final four compare quality and time, revise specifications, and decide whether to continue, change, or stop. The pilot should permit findings that some tasks do not justify automation.

Exit conditions may include persistent important errors, excessive maintenance, weaker independent student performance, unresolved data permissions, or users' inability to understand outputs. Stopping one implementation does not invalidate visual thinking. Preserved structures and instructional experience can move to simpler tools, keeping educational aims independent of a platform.

Expansion should copy validated task patterns before adding domains and complexity. A stable policy case may extend to promotional rules and then more complex services. One successful demonstration does not establish that every course can be automated. Each extension rechecks evidence, authority, and assessment, supporting a system instructors can understand and maintain.

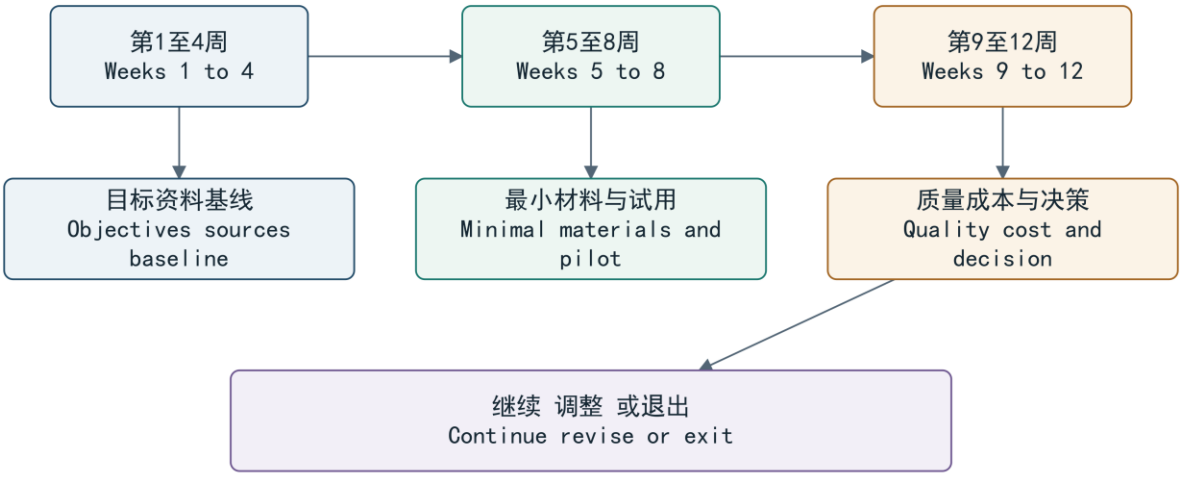


图 18 十二周试点与继续或退出

Figure 18 Twelve week pilot with continuation or exit

本报告提出的概念与设计示意 Conceptual design proposed in this report

表 16 试点风险与退出条件

Table 16 Pilot risks and exit conditions

风险 Risk	监测信号 Signal	响应 Response
事实错误 Factual error	关键资料持续错用 Repeated misuse of sources	暂停相关输出并修复 Suspend and repair affected output
学习替代 Learning substitution	成品好而独立表现下降 Good products weaker independence	增加先尝试与迁移 Restore attempts and transfer
维护过重 Maintenance burden	返工抵消节省时间 Rework exceeds saved time	缩小范围或更换方案 Reduce scope or replace
权限不清 Unclear permission	资料或行动无授权 Unauthorized data or action	限制能力与确认责任 Constrain and confirm ownership
可访问障碍 Access barriers	部分学生无法参与 Some learners cannot participate	提供替代路径与支持 Provide alternative access
不可复现 Irreproducibility	无法重建关键结果 Important results cannot be rebuilt	保留数据结构与版本 Preserve data semantics and versions

本报告的综合分析与设计建议 Report synthesis and design recommendations

17 可复用任务与提示模板
Reusable tasks and prompting templates

17 1 问题澄清与表征选择模板
Clarification and representation selection template

下面的模板是本报告提出的工作建议，使用时需要替换情境并检查结果。它不要求模型披露内部思维，而要求提供可检查的任务定义、依据和候选结构。模板一适合开始研究或备课：“我需要解决的问题是某一明确问题，受众是某类人，最终要支持某项判断。已知资料包括若干来源，限制包括时间、权限与学生基础。请先区分已经给出的事实、尚未确认的假设与缺失信息，再提出三种适合的视觉表征，分别说明能回答的问题、不能回答的问题、需要的数据和选择理由。资料中没有的信息请标为未知。”

教师可以追加：“本节课学生需要证明的能力是关系解释与迁移，请为每种候选表征提出一项学生操作，而不是只给观看建议。不要因为主题是 AI 就把 AI 放在所有图的中心。若一张表比复杂图更合适，请明确说明。”这段约束的作用，是把图形选择与认知目标相连，防止系统按常见视觉套路直接生成成品。

使用后应检查三点。候选是否真正不同，还是同一内容换了外观；每种形式是否对应明确问题；缺失资料是否被偷偷补齐。若模型把所有候选都说成适合，要求它指出各自失败条件。若模型仍无法区分形式，应先缩小任务。提示词不能补救一个没有边界的问题，好的澄清往往比更长指令更重要。

The following templates are proposed working aids. Contexts must be replaced and outputs checked. They request inspectable definitions and evidence rather than hidden internal reasoning. Template one supports research or lesson preparation: “The problem is a specified question, the audience is a specified group, and the result should support a specified judgment. Available evidence consists of the listed sources. Constraints concern time, permission, and prior knowledge. First separate supplied facts, unconfirmed assumptions, and missing information. Then propose three suitable representations. For each, explain the question it answers, the question it cannot answer, required data, and the reason for choosing it. Mark absent information as unknown.”

An instructor can add: “The learning objective is relational explanation and transfer. Propose one learner operation for each representation rather than only a viewing activity. Do not place AI at the center merely because the topic concerns AI. State when a table is more appropriate than a complex graphic.” This connects form to cognitive purpose.

Check whether alternatives genuinely differ, whether each answers a clear question, and whether missing evidence has been invented. Ask for failure conditions if every option is presented as universally suitable. If distinctions remain unclear, narrow the task. A longer prompt cannot repair an unbounded problem.

17 2 来源到结构的转换模板

Source to structure template

模板二用于把资料变成可检查结构：“请仅依据所给资料提取概念、动作、状态、角色和条件。为每个元素分配稳定编号，记录原文支持位置。连接必须标明关系类型，可选顺序、数据传递、依赖、分类、假设影响和授权。不要把相邻段落自动视为因果。输出元素表、关系表、三项关键歧义和建议图形。未经确认的关系标为待核查，不要通过常识补齐。”

对双语输出，再追加：“中文和英文共享元素编号。保留数字、单位、否定、适用范围与不确定性。遇到中文术语可能对应多个英文词时，先说明情境中的含义，再确定译法。不要生成两套独立事实结构。”这可以降低双语材料各自演化的风险，但仍需人工检查，尤其是政策条件、例外和期限。

对学生使用，可把输出限制为半成品：“先提供关键节点，不补全部连接，向我提出三个关系判断问题；在我回答前不展示完整结构。”这样保留了建构机会。对教师使用，可以要求完整候选并附来源，但关键关系仍由教师确认。相同模板根据使用者目标改变开放程度，避免把便利与学习混为一谈。

Template two converts sources into inspectable structure: “Using only the supplied material, extract concepts, actions, states, roles, and conditions. Give each a stable identifier and record its supporting location. Every edge needs a type: sequence, data transfer, dependency, classification, hypothetical influence, or authorization. Do not infer causation from adjacent paragraphs. Return element and relationship tables, three important ambiguities, and a suggested representation. Mark unconfirmed relations for checking rather than completing them from general knowledge.”

For bilingual work, add: “Chinese and English share identifiers. Preserve numbers, units, negation, scope, and uncertainty. When a Chinese term has several possible English equivalents, identify its contextual meaning before choosing. Do not create two independent factual structures.” Human review remains necessary, particularly for conditions, exceptions, and deadlines.

A student version can request a partial structure: “Provide key nodes without all connections and ask three relational questions. Do not reveal the complete structure before I answer.” An instructor may instead request a complete sourced candidate. The degree of completion follows the user's objective, keeping convenience distinct from learning.

17 3 数据图与计算核验模板

Chart generation and computational checking template

模板三用于数据分析：“请先检查字段名称、单位、时间粒度、缺失值与重复记录，列出不能从数据判断的内容。根据问题提出两种图形，说明编码和统计口径。生成可执行的变换与绘图方案，保留原始数据和变换后数据。执行后提供三个可独立复算的点，并说明是否存在预测、补值或合成情境。若不能执行，请明确标为未执行，不要把代码描述成结果。”

可以追加：“百分比必须说明分母，金额必须说明是否含税费与佣金，累计值必须说明起点。不要自动把缺失月份补成平滑趋势。图的主要结论与可能反例分别写出。若选择双轴、截断轴或

非线性尺度，请解释原因，并给出更容易比较的替代方案。”这些要求把图表可信度落到具体行为，比“请严谨”更容易验收。

对合成课堂数据，再加：“所有标题与图注标明教学演示，不能使用真实酒店或机构名称暗示实测。列出模型假设与有效范围，并设计一个改变参数的任务。”此模板适合收益、营销漏斗和现金流案例。使用时仍要检查模型是否擅自增加变量，或把没有数据的情境解释成预测。清楚的图注是必要条件，却不能替代计算与语义核查。

Template three supports analysis: “Inspect field names, units, time granularity, missing values, and duplicates. List what the data cannot establish. Propose two chart forms for the question and explain their encodings and definitions. Generate an executable transformation and visualization plan, preserving original and transformed data. After execution, provide three independently reproducible points and identify prediction, imputation, or synthetic scenarios. If execution is unavailable, label the code unexecuted rather than present it as a result.”

Additional constraints can require denominators for percentages, tax and commission definitions for money, and starting points for cumulative measures. Ask the system not to smooth missing months, to distinguish conclusions from counterexamples, and to justify dual, truncated, or nonlinear axes with a more readily comparable alternative. These are observable requirements rather than a general request for rigor.

For synthetic lessons, require demonstration labels, explicit assumptions, valid ranges, and a parameter-change task. Avoid real institution names that imply field observations. Check for invented variables and scenario outputs mislabeled as forecasts. Clear captions are necessary but do not replace computational and semantic review.

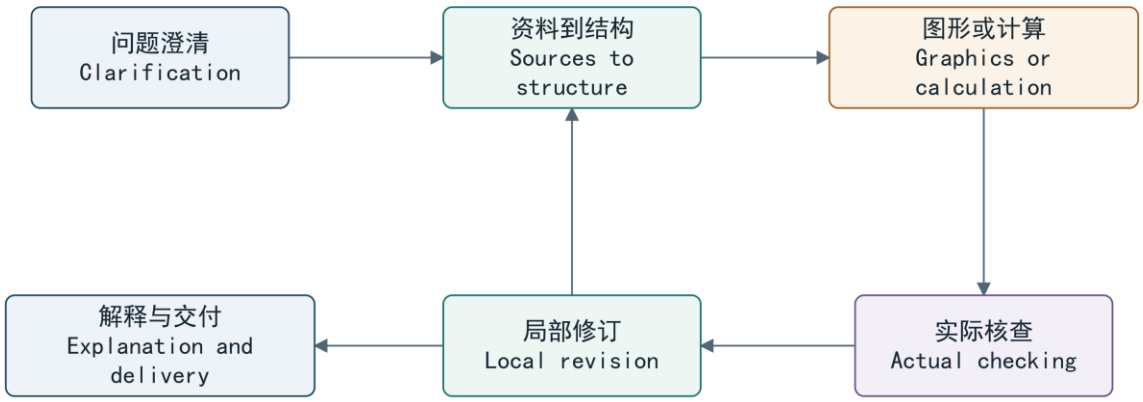


图 19 提示模板如何进入可检查工作流
Figure 19 Prompts within an inspectable workflow
本报告提出的概念与设计示意 Conceptual design proposed in this report

17 4 评审与教学反馈模板

Review and instructional feedback template

模板四用于审查：“请检查这份图文方案，不要直接重写全部内容。按事实、关系、数量、来源、可读性、权限和学习目标分别列出具体问题。每个问题注明位置、依据、可能后果和最小修改。区分确定错误、需要确认与可选优化。不能核查的项目请标为未验证。优先处理会改变判断的错误，而不是只优化风格。”

用于学生反馈时，可改为：“先指出一项结构正确的地方，再提出不超过三个需要我解释的问题。不要立即给出答案。我的解释后，再根据原资料指出一项最重要的修改，并要求我说明修改理由。”这里的正向反馈应基于可见表现，避免空泛鼓励。教师应检查模型是否给出事实正确、难度适当的问题，防止反馈把学生引向错误方向。

最后可以加入交付核查：“请比较中文与英文、图与表、正文与附录的编号和数字。列出冲突，不能确定哪版正确时不要自行选择。检查所有情境值是否标明演示属性，所有研究数值是否有来源，所有行动是否有负责人与权限说明。”这些模板的共同原则是把 AI 输出变成可检查产物，人的确认集中于重要关系与后果。它们需要随真实错误迭代，不能视为永久有效的咒语。

Template four supports review: “Inspect this visual proposal without rewriting everything. Identify specific factual, relational, quantitative, sourcing, readability, permission, and learning-alignment issues. Give the location, basis, consequence, and smallest repair. Separate confirmed errors, questions requiring confirmation, and optional improvements. Mark uncheckable items as unverified. Prioritize errors that change judgment rather than style alone.”

A learning version can first recognize one demonstrably correct structure, then ask up to three explanatory questions without revealing answers. After the learner responds, it identifies one important revision using the sources and requests a reason. Positive feedback should refer to visible performance. Instructors still check accuracy and difficulty.

A final consistency request can compare identifiers and numbers across languages, figures, tables, prose, and appendices. It should report conflicts instead of silently choosing a version and check demonstration labels, research sources, responsibility, and permissions. These templates make outputs inspectable and concentrate human confirmation on consequential relationships. They need revision through use rather than being treated as permanent formulas.

表 17 提示模板的输入与验收
Table 17 Prompt inputs and acceptance checks

模板 Template	必需输入 Input	输出验收 Acceptance
问题澄清 Clarification	对象受众判断与限制 Object audience judgment constraints	事实假设未知分开 Facts assumptions unknowns separate

模板 Template	必需输入 Input	输出验收 Acceptance
表征选择 Form selection	任务动词与数据条件 Task and available evidence	候选有适用与失败条件 Candidates have scope and failures
资料转换 Source conversion	可定位资料与术语 Locatable sources and terms	编号关系来源可查 Identifiers relations traceable
数量绘图 Quantitative chart	字段口径单位与目标 Fields units definitions objective	实际执行且可复算 Executed and reproducible
学生反馈 Learner feedback	首次尝试与学习目标 Attempt and learning objective	问题促使解释而非代答 Questions elicit explanation
一致性检查 Consistency	各语言与交付版本 Languages and deliverable versions	冲突定位并保留未知 Conflicts located unknowns retained

本报告的综合分析与设计建议 Report synthesis and design recommendations

18 术语与前沿研究议程
Terminology and a research agenda

18 1 从感知到生成再到行动
From perception to generation and action

近年来的多模态研究逐步把视觉从输入扩展为推理过程的一部分。较早系统常把图片转成语言描述，再以文字进行后续推断；工具增强方法则可以在图上画框、裁剪、标记或计算；新的统一模型探索文本与视觉生成交替进行。不同路线在信息保留、可解释性、成本和控制方面存在不同取舍，不能只用一个“多模态”标签概括。

2025 年的 Latent Sketchpad 预印本探索在模型中生成视觉潜在表示，并通过解码形成可解释草图。[42]2026 年的一项世界模型研究提出，在适合视觉建模的任务上交替视觉与语言生成可能有益，但在其他任务上未显示明确优势。[41]这些工作提示表征与任务匹配仍然重要。它们属于特定技术研究，不能直接证明人类教育效果，也不应从标题中的“类人”推断相同认知机制。

对应用者而言，判断重点可以分成三问：系统是否真正使用了新的视觉信息；视觉中间材料是否保留重要约束；加入视觉步骤是否在准确性、成本或人的理解上带来收益。若只是在文字回答旁边附上一张图，它可能是展示，而不是视觉推理。若图被用于下一步操作，也仍要检验图本身是否正确。技术前沿没有取消基本验证要求，反而使验证对象更复杂。

Multimodal research increasingly treats vision as part of a process rather than only an input. Some systems convert images into language before continuing. Tool-augmented methods can mark, crop, calculate, or draw. Unified approaches explore interleaved verbal and visual generation. These routes make different trade-offs in information preservation, interpretability, cost, and control; one multimodal label is insufficient.

The 2025 Latent Sketchpad preprint explores generated visual latent representations that can be decoded into sketches.[42] A 2026 world-model study reports benefits of interleaved visual and verbal generation for tasks suited to visual modeling, without a clear advantage elsewhere.[41] These findings reinforce task matching. They concern specific technical settings, not educational effectiveness or proof of identical human and machine cognition.

Application review can ask whether the system uses additional visual information, whether intermediate artifacts preserve constraints, and whether the visual step improves accuracy, cost, or human understanding. A picture attached to prose may be presentation rather than reasoning. A picture used in a subsequent operation still requires checking. New methods increase the complexity of validation rather than remove it.

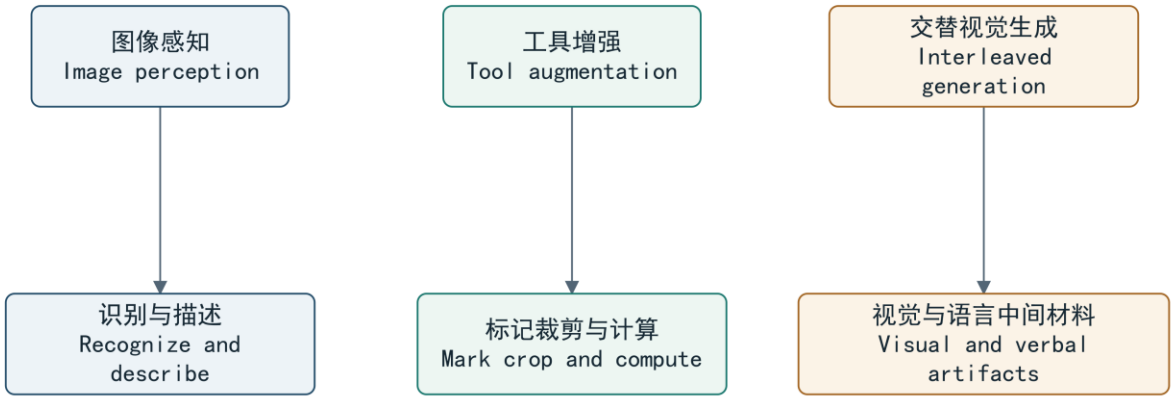


图 20 视觉 AI 研究路线的层次
Figure 20 Levels of visual AI research approaches
根据[20][25][41][42]绘制 路线比较不是能力排名 Approach comparison not a capability ranking

18 2 需要统一的几个术语

Terms that need consistent use

视觉思维指使用视觉与空间表征进行认知活动；视觉素养指阅读、评价与制作视觉信息的能力；视觉表征指承担信息功能的图形、空间或符号形式；可视化指把数据或结构映射到视觉形式的过程；图解推理指借助图中关系完成推断；可视分析指把分析与交互视觉结合的活动。教学中可以用简短语言解释，但应避免把这些词交替当作同义词。

对象想象与空间想象描述不同维度，不能分别等同于“会画画”与“会逻辑”。概念图强调概念之间有含义的关系，思维导图通常以层级或辐射组织联想，流程图强调顺序与条件，因果图强调假设的影响结构。知识图谱还涉及类型化数据与可查询关系。一个带箭头的图可以同时具备部分特点，但作者必须说明连接语义。

AI 相关术语也应保持精确。模型、工具、工作流与智能体不是同一层次；检索不等于验证，生成不等于执行，执行成功不等于结果正确。提示模板可以指导行为，却不必然提供确定性能力。专家角色是职责安排，不自动等于专业可靠性。把这些概念放在一个层次清楚的术语表中，有助于学生在不同平台之间迁移，而不依赖某个产品的命名。

Visual thinking concerns cognitive activity using visual and spatial representations. Visual literacy concerns reading, evaluating, and making visual information. A representation is a graphical, spatial, or symbolic form with an informational role. Visualization maps data or structure into visual form. Diagrammatic reasoning uses graphical relationships for inference. Visual analytics combines analysis with interactive visual methods. Accessible language is useful, but the terms should not become interchangeable.

Object and spatial imagery are dimensions, not simple equivalents of artistic and logical talent. Concept maps emphasize meaningful conceptual relations. Mind maps commonly organize associations hierarchically or radially. Processes emphasize sequence and conditions. Causal diagrams express hypothesized influence. Knowledge graphs additionally concern typed, queryable data. An author should state what an edge means even when forms overlap.

Models, tools, workflows, and agents also occupy different levels. Retrieval is not verification, generation is not execution, and successful execution is not correctness. A prompt can guide behavior without creating deterministic competence. Expert roles allocate duties without guaranteeing professional reliability. A clear glossary supports transfer across platforms.

表 18 核心术语中英文对照
Table 18 Core bilingual terminology

中文 Chinese	英文 English	使用边界 Boundary
视觉思维	Visual thinking	认知实践 不等于美化 Cognitive practice rather than decoration
视觉素养	Visual literacy	读图评价与制作 Interpret evaluate and create
对象想象	Object imagery	对象细节维度 非固定类型 Object-detail dimension not a fixed type
空间想象	Spatial imagery	关系转换 不自动等于因果 Spatial relations not automatically causation
图解推理	Diagrammatic reasoning	基于图中关系完成推断 Inference through graphical relations
多重外部表征	Multiple external representations	多种形式需要对应与整合 Forms require correspondence and integration
可视分析	Visual analytics	交互视觉与分析结合 Interactive visual analysis
检索增强生成	Retrieval augmented generation	检索约束生成 仍需核查 Retrieved context still needs verification
认识性行动	Epistemic action	操作帮助发现与理解 Action supports discovery and understanding
可追溯性	Traceability	判断可定位到资料与变换 Claims link to sources and transformations

本报告的综合分析与设计建议 Report synthesis and design recommendations

18 3 五个值得优先研究的问题
Five research priorities

第一，AI 降低制作劳动之后，人是否仍进行足够的选择、组织与整合。第二，视觉反馈比文字反馈在哪些任务与基础条件下更有效。第三，学生能否把在 AI 辅助中形成的结构带到独立任务。第四，如何让图中的不确定性被正确理解，而不被误读为系统不可靠或可以忽略。第五，双语图文能否帮助跨语言理解，以及何时增加额外负担。这些问题应通过具体任务研究，而不是凭一般态度回答。

每个问题都需要区分短期与长期。第一次使用时的新鲜感、界面熟悉度和教师支持可能影响结果；持续使用后，工具习惯与能力发展又可能改变。建议研究包含多次任务与延迟测量，关注个体变化和任务差异。对无法进行大规模实验的教师，可以采用连续的小规模改进研究，保留相同测量并坦诚说明限制。

还应研究失败而不只是成功示范。哪些图让人过度自信，哪些中间表示难以维护，哪些任务不值得加入 AI，哪些群体因界面或语言受到排除。失败案例如果结构、来源与修改过程清楚，同样具有知识价值。研究领域需要的不只是更强生成能力，也包括更成熟的比较、测量与报告方法。

Five priorities deserve study: whether reduced production effort preserves selection and integration; when visual feedback outperforms verbal feedback; whether AI-supported structure transfers to independent work; how uncertainty can be interpreted appropriately; and when bilingual graphics support or burden understanding. These questions require defined tasks rather than general attitude measures.

Short- and long-term effects should be separated. Novelty, interface familiarity, and instructor support affect initial use, while habits and skills may change later. Repeated tasks and delayed measures can examine these developments. Instructors without large experimental resources can conduct smaller continuing improvements with stable measures and explicit limitations.

Failures also deserve study: graphics that inflate confidence, intermediate representations that resist maintenance, tasks that do not justify AI, and interfaces that exclude users. Well-documented failures contribute knowledge. Better generation needs to be accompanied by better comparison, measurement, and reporting.

18 4 对理论与实践的综合判断
An integrated judgment on theory and practice

视觉思维在 AI 时代的作用，不是因为机器让文字失去价值，而是因为信息生产越来越容易，结构判断与证据判断更加需要训练。人可以借图看见联系，也可能被图的完整感误导；AI 可以扩展表征，也可能复制错误。因此，最值得建设的是一个有来源、有操作、有反馈和有独立检验的认知环境。环境中的每一种表征都承担任务，每一种自动化都说明责任。

对本报告所关注的 AI 导论课程，建议把课程地图、行业问题、工具使用和学习评价放在同一设计中。教师用图解释，学生用图建构，AI 帮助转换与比较，计算工具核验数量关系，真实任务检验迁移。这是一个可以逐步实施、检验和修订的方向，而不是一次采购平台或安装 Skill 就能完成的改革。其价值最终取决于学生是否更会理解问题、解释依据并负责任地行动。

Visual thinking matters because easier information production increases the need for structural and evidential judgment. Graphics can reveal relationships or create a misleading sense of completeness. AI can expand representations or reproduce errors. The worthwhile objective is an environment with sources, manipulation, feedback, and independent testing, where each representation serves a task and each automated action has a responsibility.

For an introductory AI course, course maps, industry problems, tool use, and assessment should be designed together. Instructors explain, learners construct, AI assists transformation and comparison, computational tools verify quantities, and new tasks test transfer. This is a direction for progressive implementation and revision. Its value depends on learners becoming better at understanding problems, explaining evidence, and acting responsibly.

19 科研阅读与知识发现应用
Applications in research reading and knowledge discovery

19 1 从文献清单到问题结构
From a reading list to a question structure

科研中的视觉思维可以从文献整理开始，但一张作者关键词网络不能替代对研究问题的理解。研究者首先需要区分理论主张、测量方式、实验条件、结果变量与推断限制。AI 可以协助从英文摘要中提取这些字段，形成证据矩阵；研究者再核对重要原文，决定哪些研究可以比较。图的作用是帮助看见差异与缺口，而不是把所有文献连接成看似统一的结论。

例如，研究绘图学习时，阅读文字后自己画图、观看教师画图、使用计算机拖拽图形和让 AI 直接生成图，是不同干预；知识记忆、理解与迁移也是不同结果。若在图中只用“视觉学习有效”一个节点概括，就会抹去重要差异。更好的证据地图可以把干预、比较条件、支持方式和测量分别编码，允许读者查看哪些连接有多项研究，哪些只是理论推测。

AI 在这个过程中应输出出处与不确定状态。摘要没有说明随机分配，就标为未说明；没有延迟测验，就不要自动补一个长期效果。研究者可以让系统列出最影响结论的缺失信息，再决定是否寻找全文。这样能把有限阅读时间集中在关键问题上。自动摘要的便利应当服务精读，而不能成为跳过关键证据的理由。

Visual thinking can organize research reading, but an author or keyword network does not replace understanding a question. Researchers need to distinguish theoretical claims, measures, interventions, outcomes, and inferential limits. AI can extract fields from English abstracts into an evidence matrix. Researchers then check important originals and decide which studies are comparable. The map reveals differences and gaps rather than forcing all papers into one conclusion.

Drawing after reading, observing an instructor draw, manipulating computer shapes, and receiving an AI-generated diagram are different interventions. Recall, comprehension, and transfer are different outcomes. A single visual-learning-is-effective node erases those distinctions. A useful evidence map encodes intervention, comparator, support, and measurement separately, showing which relationships have repeated evidence and which remain theoretical.

AI extraction should preserve provenance and uncertainty. If random assignment is not described, it remains unspecified. If no delayed test appears, a long-term effect must not be invented. A system can identify missing information most likely to change interpretation, helping researchers prioritize full-text reading. Summary convenience should support close reading rather than replace it.

19 2 证据地图怎样避免制造共识
Avoiding manufactured consensus in evidence maps

证据地图应同时容纳支持、无明确差异、混合结果和边界条件。不同颜色可以辅助识别，但需要文字标签与定义。连接粗细若表示研究数量，应说明数量不等于质量；节点大小若表示样本

量，应避免把大样本相关研究自动视为强因果证据。一个图形编码承担一个清楚意义，比同时用大小、颜色、亮度和位置暗示“重要性”更容易审查。

绘图研究中的一个有用例子，是 2025 年关于绘图与检索练习结合的研究记录：闭卷后开卷的安排未显示总体优势，同时涉及检索收益与心理模型质量的权衡。[40]这个结果提示，听起来合理的两种有效策略组合，不一定自动产生更好效果。研究者可以在证据地图中把“促进提取”和“影响模型质量”分别列出，形成下一步研究问题，而不是把不显著结果简单归为失败。

对 AI 辅助视觉研究，也应保留反例和成本。某种方法可能提高准确率却增加时间，可能改善新手成品却降低独立解释，可能在空间题有效却在文字分类中没有优势。证据地图的价值正是帮助读者看到这种条件性。若它最终只剩一张不断向“AI 有效”汇聚的箭头图，就需要重新检查是否有选择性呈现。

Evidence maps should accommodate support, unclear differences, mixed findings, and boundaries. Color needs labels and definitions. Edge width representing study count must not imply quality. Node size representing sample size must not automatically imply causal strength. One explicit meaning per encoding is easier to review than several visual properties jointly suggesting importance.

A 2025 study record on combining drawing with retrieval practice reports no overall superiority of a closed-then-open-book arrangement and discusses a trade-off involving retrieval and mental-model quality.[40] Combining two plausible strategies does not necessarily improve learning. An evidence map can separate the proposed mechanisms and generate further questions instead of treating an unclear advantage as a simple failure.

AI-assisted visual research should similarly retain counterexamples and costs. Accuracy may improve while time increases; novice products may improve without independent explanation; spatial tasks may benefit while classification does not. Making conditionality visible is the map's purpose. A diagram in which every path converges on a general AI-success claim deserves scrutiny for selective presentation.

19 3 用概念模型连接理论与测量
Connecting theory and measurement

研究方案图可以区分干预、过程、结果和背景条件。例如 AI 问题式反馈是干预，关系修订与自我解释是可能过程，迁移表现是结果，先验知识与语言熟悉度是背景条件。箭头应明确标为假设，不能因为图画得完整就当作机制已经证明。每个概念还需连接可观察指标，否则模型无法指导数据收集。

AI 能够帮助寻找替代解释。若 AI 组成绩更高，可能与更多学习时间、更清楚提示或新鲜感有关，而不只是视觉反馈。把这些因素画到方案图中，有助于设计对照。研究者也可以让模型提出最强反对意见，再检查实验能否排除。这里 AI 扮演的是产生候选的助手，设计是否成立仍由方法与证据决定。

在报告结果时，概念图与统计图应互相对应。理论中声称检验的关系，结果部分应能找到相应分析；未直接测量的机制，应保持推测状态。一个常见问题是讨论部分又加入许多漂亮机制，

却没有数据支持。保留从研究问题到变量、分析和结论的可视索引，可以帮助作者控制这种扩张，也帮助审稿者定位论证。

A study model can separate intervention, process, outcome, and background. Question-based AI feedback is an intervention; relational revision and self-explanation are possible processes; transfer is an outcome; prior knowledge and language familiarity are contextual conditions. Arrows remain hypotheses. Each concept also needs an observable measure if the model is to guide collection.

AI can propose alternatives. Better performance may reflect more time, clearer prompts, or novelty rather than visual feedback alone. Adding these possibilities helps design comparisons. A model can also generate a strong objection for researchers to test against the design. It supplies candidates; methods and evidence determine adequacy.

Conceptual and statistical displays should correspond. A relationship claimed as tested needs an identifiable analysis, while unmeasured mechanisms remain speculative. A visual index connecting questions, variables, analyses, and conclusions helps prevent discussion sections from expanding beyond the evidence and helps reviewers locate the argument.

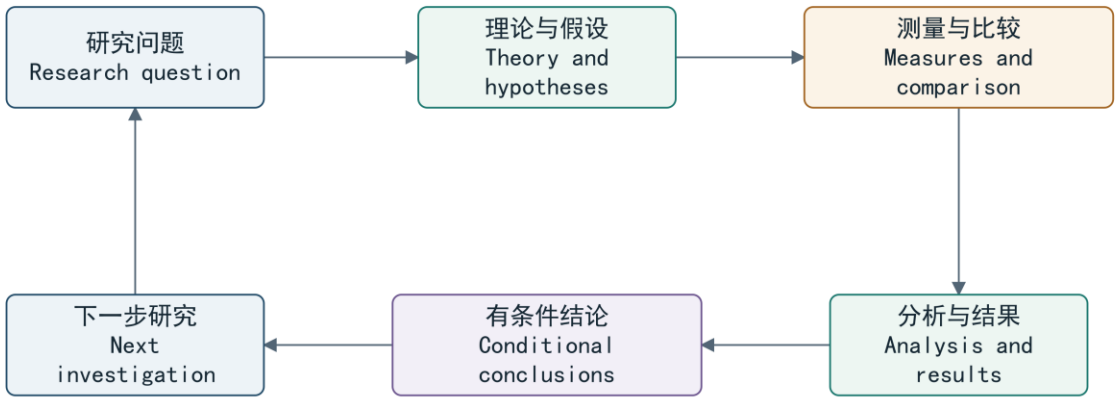


图 21 科研中的问题证据与结论索引
Figure 21 Linking research questions evidence and conclusions
本报告提出的概念与设计示意 Conceptual design proposed in this report

19 4 英文资料研究中的人机分工
Division of work in English language research

英文资料检索可分为发现、筛选、核对、综合和写作五步。AI 与学术工具擅长扩大线索、整理摘要和发现术语变体，人需要检查来源身份、研究设计、语境和结论强度。检索时应组合概念词、任务词和方法词，例如 visual thinking、diagrammatic reasoning、drawing to learn、meta-analysis 与 human AI collaboration，避免只用一个大众术语漏掉相关研究传统。

引用前应核对标题、作者、年份与正式发表状态。在线年份与期刊卷期年份可能不同，预印本可能更新，检索工具也可能提供不一致元数据。本报告对绘图元分析采用 2020 年正式卷期，并说明 2019 年首次在线，正是为了避免把两个日期当成两篇研究。类似核对应成为学生学术素养的一部分，而不只是教师替他们完成的后台工作。

综合时应使用自己的问题结构，而不是把每篇摘要顺序翻译。可以围绕“什么任务有效、为何有效、何时无效、如何实现、如何验证”组织内容，并把原始证据与自己的建议分别标明。双语写作则要求两种语言共享相同结论强度。中文写“可能支持”，英文不能变成“proves”。这种细节直接影响研究报告的可信度。

English-language research can be organized into discovery, screening, verification, synthesis, and writing. AI and scholarly tools broaden leads, summarize abstracts, and identify terminology. People check source identity, design, context, and strength. Searches should combine conceptual, task, and methodological terms rather than rely on one popular phrase, which may miss related traditions.

Before citation, verify title, authors, year, and publication status. Online and issue dates may differ, preprints change, and discovery services can disagree. This report uses the formal 2020 issue year for the drawing meta-analysis while noting first online publication in 2019, avoiding duplicate treatment. Such checking belongs in student scholarly competence.

Synthesis should follow a question structure rather than translate abstracts sequentially. Organize around tasks, mechanisms, boundaries, implementation, and validation, separating evidence from recommendations. Bilingual conclusions need equal strength: a Chinese statement of possibility must not become an English claim of proof. This detail directly affects credibility.

表 19 文献提取与证据审查字段
Table 19 Literature extraction and review fields

字段 Field	需要记录 Record	常见错误 Error
发表身份 Publication status	年份版本与正式状态 Year version and status	把更新版算作新研究 Counting versions as new studies
研究设计 Design	实验调查综述或原型 Experiment survey review prototype	技术演示等同学习证据 Demonstration treated as learning evidence
干预 Intervention	谁做何种视觉操作 Who performs which visual operation	AI 代画与人画混同 Conflating AI and learner generation
比较 Comparator	时间提示与材料条件 Time guidance and materials	忽略弱对照 Ignoring weak comparisons

字段 Field	需要记录 Record	常见错误 Error
结果 Outcome	记忆理解迁移与用时 Retention comprehension transfer time	用成品质量替代能力 Product quality replaces capability
限制 Limits	样本任务与可见资料范围 Sample task and accessible evidence	摘要没有仍擅自补齐 Inventing absent details

本报告的综合分析与设计建议 Report synthesis and design recommendations

20 设计实践与组织协作应用
Design practice and organizational collaboration

20 1 从草图到可讨论原型
From sketches to discussable prototypes

在产品、空间和服务设计中，视觉思维让尚未确定的想法成为可以共同讨论的对象。草图的粗糙有时是一种优势，因为它提示方案仍可改变。AI 可以迅速生成不同布局、界面和场景，使团队比较更多方向；但过早生成逼真图像，也可能让参与者误以为成本、结构与可行性已经解决。原型应明确表达完成程度，避免外观精致掩盖未确定部分。

Sutherland 的 Sketchpad 早已展示图形交互与约束在计算机辅助设计中的潜力。[19]今天的 AI 增加自然语言解释、候选生成和跨形式转换，却没有替代约束本身。一个酒店大堂效果图仍需检查通行、消防、工作动线与设施尺寸；一个漂亮服务界面仍需检查数据和权限。AI 视觉生成应进入有约束的设计过程，而不应仅作为最终渲染。

课堂可以让学生先画低精度服务界面，标出用户需要做的判断，再让 AI 产生两个外观方案。比较时先检查任务是否顺利完成，再讨论视觉风格。若新方案删掉了重要说明，即使更美观也应修订。学生由此理解，设计质量来自目的、约束与体验的共同满足，而不只是图像吸引力。

In product, spatial, and service design, sketches make uncertain ideas available for discussion. Roughness can be useful because it signals that change remains possible. AI quickly generates layouts, interfaces, and situations, broadening comparison. Premature realism can nevertheless imply that cost, structure, and feasibility are settled. A prototype should communicate its level of completion.

Sutherland's Sketchpad demonstrated the potential of graphical interaction and constraints in computer-aided work decades ago.[19] Contemporary AI adds language interaction and candidate generation without replacing constraints. A lobby rendering still needs checks on movement, dimensions, and operating requirements. A service interface still needs data and permission checks. Generation belongs inside a constrained process.

Students can sketch a low-fidelity interface and identify the decisions users must make before asking AI for two visual alternatives. They compare task completion before style. An attractive version that removes an important explanation needs revision. Design quality then concerns purpose, constraints, and experience together.

20 2 共享图形如何帮助跨岗位协作
Shared graphics across organizational roles

组织中的图可以成为不同专业之间的共同参照。业务人员解释目标，数据人员解释字段，技术人员解释实现，管理人员解释责任。共同参照不意味着所有人理解完全相同，而是让分歧有可以定位的位置。若有人说“这个流程不对”，主持者可以请他指出节点、连接或规则，并提出修改依据。讨论因此从抽象意见转向可处理差异。

AI 会议助手可以把讨论整理为候选图与决策记录，但必须区分建议、已同意事项和未解决分歧。不能把多数人反复提到的内容自动标为共识，也不能把会议中沉默理解为同意。图中的确认状态应来自明确记录。对于重要改变，应保留提出者、负责岗位和后续核实动作，避免图变成没有人真正承担责任的漂亮总结。

团队还可以使用分层视图。管理层看目标与风险，实施人员看接口与数据，一线人员看动作和例外；三者共享关键编号，允许向下展开。这样既避免每张图承载所有细节，也避免不同岗位各维护一套互相矛盾的解释。AI 适合帮助生成不同视图，但共享语义结构与变更控制仍是组织工作。

Shared graphics provide a reference across professional roles. Business staff explain objectives, analysts explain fields, developers explain implementation, and managers explain responsibility. Shared reference does not require identical interpretation; it makes disagreements locatable. Someone who objects to a process can identify a node, edge, or rule and provide a basis for change. Discussion becomes a tractable difference rather than a general opinion.

An AI meeting assistant may draft diagrams and decision records, but it must distinguish proposals, agreements, and unresolved disagreements. Repetition does not establish consensus, and silence does not establish approval. Confirmation states require explicit records. Important changes need ownership and follow-up so that the diagram does not become an attractive summary without responsibility.

Layered views support different needs. Managers inspect objectives and risks, implementers inspect interfaces and data, and frontline staff inspect actions and exceptions. Shared identifiers connect these levels. AI can produce views while common semantics and change control remain organizational responsibilities.

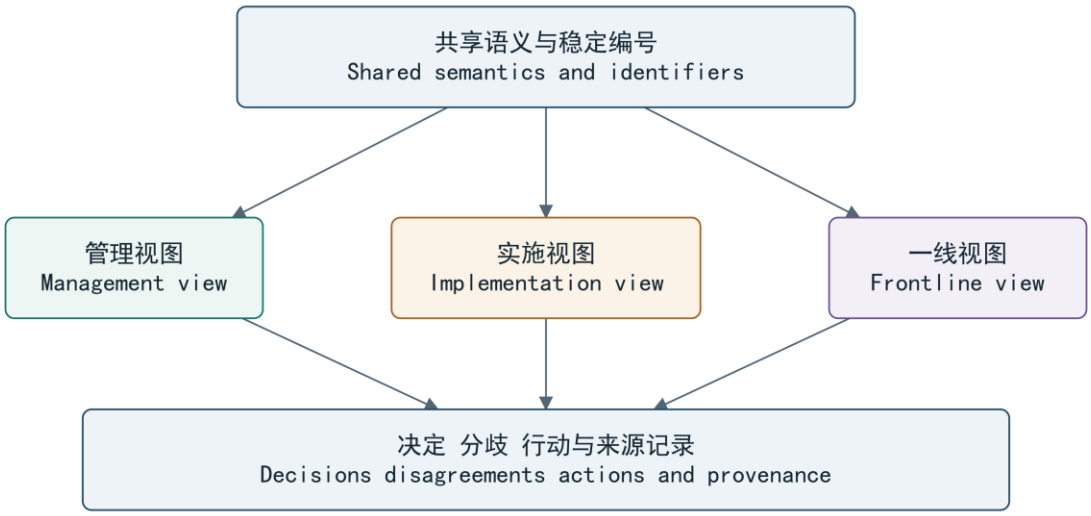


图 22 共同语义支持不同组织视图
Figure 22 Shared semantics support different organizational views
本报告提出的概念与设计示意 Conceptual design proposed in this report

20 3 创造性探索与收敛的节奏
The rhythm of divergence and convergence

创造性工作需要扩展候选，也需要选择。AI 在扩展阶段可以提供反常组合、不同视角与场景变化，视觉工作区帮助团队把候选并列，而不是按聊天顺序逐个忘记。进入收敛阶段后，应使用预先确定的标准比较价值、可行性、风险和实施成本。若系统一直生成新图而不支持选择，团队可能产生更多材料却无法前进。

建议将候选保持在可管理数量，并记录每个方案的一项独特价值与一项主要弱点。图像差异不应被误认为概念差异；三个颜色不同的同一流程，不是三个方案。真正的替代可能改变服务时机、责任分配、数据使用或用户选择。AI 可以被要求先说明结构差异，再生成视觉差异。这样能够减少对外观新奇的依赖。

收敛也应允许混合与返回。团队可以从两个候选中组合有依据的部分，但必须检查新关系是否一致。某个方案因成本被排除后，如果外部条件变化，也可以重新进入比较。保留候选与理由，使设计过程拥有记忆。视觉思维在这里不仅帮助想象，也帮助决定何时停止探索并开始验证。

Creative work needs expansion and selection. AI can propose unusual combinations and alternative perspectives, while a visual workspace keeps candidates side by side instead of losing them in chat sequence. Convergence then uses prior criteria for value, feasibility, risk, and cost. Endless new images can increase material without enabling progress.

Keep a manageable set and record a distinctive benefit and weakness for each. Visual difference is not necessarily conceptual difference: three colors of one process are not three solutions. Alternatives may change service timing, responsibility, data use, or user choice. Asking AI to describe structural differences before drawing reduces dependence on novelty.

Convergence can combine supported parts or return to a rejected option when conditions change. Any hybrid needs consistency checks. Preserved alternatives and reasons give the process memory. Visual thinking supports imagination and the decision to stop exploring and begin testing.

20 4 组织学习与长期维护
Organizational learning and maintenance

一个团队真正积累的，不应只是图形模板，而应是问题模式、判断规则和失败经验。每次项目结束，可以归档问题、主要表征、关键证据、改变过的假设和结果。后来出现类似任务时，AI 检索这些经验并提出候选，但不能默认旧方案直接适用。相似情境仍可能具有不同客户、权限、时间与资源约束。

维护时要防止图形失去现实对应。岗位调整、系统升级和政策变化之后，旧图可能仍被反复引用。每张关键流程应有负责人、复核条件与版本日期。复核不必机械地频繁重画，而应由重要变化触发。图表若用于训练新员工，尤其需要确认它描述的是现行程序，而不是过去理想化的方案。

组织可以用简单指标检查视觉知识是否发挥作用：新成员理解流程需要多久，跨部门问题能否定位，错误是否能够追溯，修改是否同步到相关材料。这些指标比图形数量更接近价值。AI 时代的组织视觉能力，最终体现为团队是否能够更快形成共同理解，并在证据变化后及时修改行动。

Teams should retain problem patterns, judgment rules, and failures alongside templates. An archive can contain the question, main representation, evidence, revised assumptions, and outcome. AI retrieves related experience and proposes candidates, but a previous solution does not automatically fit a new context with different customers, permissions, timing, or resources.

Representations can lose contact with operations after role, system, or policy changes. Important workflows need ownership, versions, and review triggers. Review should follow meaningful changes rather than require constant redrawing. Materials used to train staff particularly need to describe current practice rather than an old idealized plan.

Useful measures include onboarding time, ability to locate cross-department problems, traceability of errors, and synchronization of revisions. These are closer to organizational value than the number of graphics produced. Capability appears when a team can establish shared understanding and revise action as evidence changes.

表 20 组织协作中的可见记录
Table 20 Visible records for collaboration

对象 Object	应说明 Required explanation	维护方式 Maintenance
建议 Proposal	是谁提出依据什么 Who proposes and on what basis	保持待确认状态 Keep pending confirmation
共识 Agreement	哪些岗位明确确认 Which roles explicitly agree	保存决定与日期 Record decision and date
分歧 Disagreement	冲突在节点还是规则 Node edge or rule conflict	记录证据与解决人 Record evidence and resolver
行动 Action	负责人与期限权限 Owner deadline authority	跟踪执行与结果 Track execution and outcomes
假设 Assumption	何时会失效 When it may fail	定义验证与更新触发 Define tests and review triggers
经验 Experience	适用与不适用条件 Applicability and limits	检索前重新匹配场景 Recheck fit before reuse

本报告的综合分析与设计建议 Report synthesis and design recommendations

21 部署 复现与日常使用手册
Deployment reproducibility and daily use

21 1 从最小可用流程开始
Beginning with a minimal viable workflow

部署的第一步可以很小：一份经过确认的资料、一种图形、一项验证和一个使用者。比如教师输入一页政策，系统提取条件，生成决策表，再由教师核对并转成课堂练习。这个最小流程足以检验结构转换是否可靠。不要在还没验证基本关系时，同时接入多种平台、多个专家和自动发布。复杂度应由任务需要逐步增加。

最小流程还需明确成功与失败。成功不只是生成文件，而是文件能打开、关系符合资料、读者能理解、任务能够完成。失败可能发生在识别、结构、渲染、翻译或使用阶段。每个阶段保留简短日志，就能避免把所有问题归为模型“不够聪明”。有些问题来自资料不清，有些来自图形不适合，有些来自界面；修复方式不同。

在课堂环境，建议先用教师审核后发布的方式，逐步开放学生自助生成。学生端应显示来源、版本和可修改范围，允许回到原始资料。系统可以建议，但应让学生知道哪些步骤需要自己解释。即使未来增加自动化，最小流程仍可作为故障时的替代路径和质量基线。

Deployment can begin with one confirmed source, one representation, one check, and one user. An instructor might provide a policy page, receive extracted conditions and a decision table, verify it, and turn it into an exercise. This minimal flow can test structural reliability before several platforms, roles, and publication mechanisms are connected. Complexity should grow with demonstrated need.

Success means more than file creation. The file must open, relationships must match the source, the audience must understand it, and the task must work. Failure may concern recognition, semantics, rendering, translation, or use. Short stage records prevent every problem from being attributed to insufficient model intelligence. Unclear evidence, unsuitable form, and interface problems require different repairs.

Classroom deployment can begin with instructor review before release and gradually enable learner generation. Sources, versions, and revision scope should be visible, with access back to evidence. The system suggests while learners retain explanatory duties. The minimal flow remains a fallback and quality baseline as automation expands.

21 2 文件结构与版本安排
File structure and versions

一个可维护项目可以区分资料、结构、数据、图形、成稿与验证记录。资料保留原始版本，结构保存节点与关系，数据保存输入与转换结果，图形保存可编辑源与导出图，成稿组织正文与引用，验证记录说明检查时间与问题。名称需要稳定、可理解，避免只有“最终版”“最终最终版”。重要产物应通过编号连接，而不是依靠人员记住文件关系。

版本变化应区分事实、结构和样式。事实修正可能要求重新分析；结构修正可能要求同步手册与作业；样式修正主要影响呈现。每次发布记录变化范围与复核要求。双语材料尤其需要共享事实与图表数据，不能分别维护两个无联系文件。若英文翻译发现中文概念不清，应回到概念层修复，而不是在译文中偷偷解释成另一种含义。

备份也应能实际恢复。保留源文件却缺少字体、数据或渲染方式，未必足以重建结果。可以用一个小型复现说明记录输入、运行方式、输出与已知限制。课程项目不必达到复杂软件工程水平，但应让另一个同学能够根据说明重建主要图表。可复现习惯让视觉成果成为可持续知识，而不是一次性图片。

A maintainable project separates sources, semantics, data, graphics, manuscript, and validation. Sources retain original versions, semantics store nodes and edges, data include inputs and transformations, graphics include editable sources and exports, and validation records checks and findings. Stable identifiers connect outputs instead of relying on memory or repeated final-version filenames.

Changes should distinguish facts, structure, and style. A factual correction may require reanalysis. A structural correction may update manuals and assignments. Styling primarily affects presentation. Releases record scope and review requirements. Bilingual materials especially need common facts and chart data. Ambiguity discovered during translation should be repaired conceptually rather than silently reinterpreted in English.

Backups must support restoration. Sources without fonts, data, or rendering information may not reproduce an artifact. A brief reproduction note records inputs, process, outputs, and limits. Another student should be able to reconstruct important graphics. This turns visual work into maintainable knowledge.

21 3 软件选择与可替换性
Software selection and replaceability

选择软件时可以比较任务覆盖、精确性、可编辑性、来源保留、协作、可访问性、成本和导出能力。结构图工具、数据图工具、文生图和交互计算各有适用范围。报告不建议用一个产品承担所有任务，也不把某个当前热门工具当作长期课程基础。核心方法应能在不同工具中实现，学生学到的是选择原则与验证方法。

Wolfram 的 Manipulate 与 Graph 等能力说明了符号计算、交互与关系表示可以结合的技术路线。[36][37]类似目标也可以由其他合适工具实现。选择时应考虑现有许可、学生环境、教师维护能力与输出要求。若需要离线使用、中文字体或手机阅读，应在小样本中实际测试，不能仅依据官方功能清单推断体验。

可替换性来自语义与呈现分离。只要结构和数据保持清楚，渲染器可以更换，报告也可以输出为不同格式。若所有逻辑只存在于一张不可编辑图片中，迁移成本就高。课程可以要求每个小组提交源结构和一份导出图，让可替换性成为基本交付要求。这样既便于教师修改，也能降低平台变化带来的影响。

Software selection can compare coverage, precision, editability, provenance, collaboration, accessibility, cost, and export. Structural drawing, charting, image generation, and interactive computation serve different

needs. One product should not automatically carry every task, and a currently popular tool should not become the only long-term foundation. Learners need selection and validation principles.

Wolfram's Manipulate and Graph illustrate a route combining symbolic computation, interaction, and relational display.[36][37] Other suitable tools may serve similar objectives. Licensing, learner environments, maintenance, and output requirements matter. Offline use, Chinese fonts, and mobile reading should be tested with real examples rather than inferred from a feature list.

Replaceability follows from separating semantics and presentation. Clear structures and data allow a renderer or export format to change. Logic trapped in an uneditable picture raises migration cost. Requiring both a source structure and an exported figure makes replaceability a practical student habit.

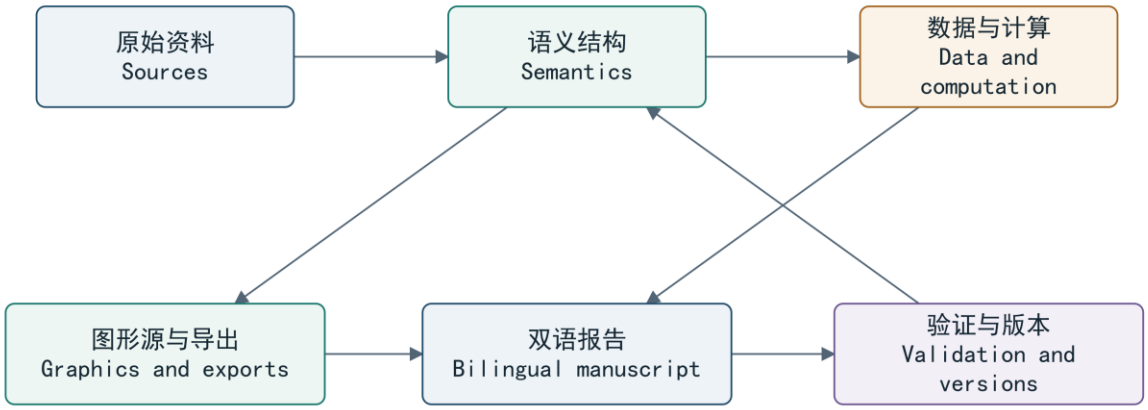


图 23 可维护的视觉知识项目
Figure 23 A maintainable visual knowledge project
本报告提出的概念与设计示意 Conceptual design proposed in this report

21 4 使用后的反馈与修订
Feedback and revision in use

日常反馈应围绕具体任务收集：哪里看不懂，哪一条关系有争议，哪一步操作失败，哪个结果无法复算。笼统满意度可以作为补充，却不能代替问题定位。系统应允许使用者直接引用元素编号报告问题，维护者再判断影响范围。修订完成后，相关材料一起更新，旧版本明确标记，避免课堂继续使用过期图。

还可以定期抽取少量输出进行人工审计，检查来源、数字、双语一致性与异常路径。抽样不能证明所有输出无误，但能发现系统性问题。若同一类错误反复出现，应修改输入规范、中间表示或验证规则，而不是只修单张图。这样的反馈闭环能把系统从生成器变成持续改进的工作环境。

对个人使用者，最重要的日常习惯是保留一个独立核查动作。每次图表至少复算一个点，每次流程至少测试一个例外，每次 AI 总结至少回看一段来源。习惯不必复杂，却能持续提醒人：输出是候选，理解需要自己的参与。视觉思维与 AI 结合的长期价值，依赖这种小而稳定的专业实践。

Daily feedback should identify a task-specific issue: an unclear relation, disputed rule, failed operation, or unreproducible result. General satisfaction is supplementary. Element identifiers help users locate a problem and maintainers assess its consequences. Related outputs update together, while older versions are marked to prevent continued use.

Small manual audits can inspect sources, numbers, bilingual consistency, and exceptions. Sampling cannot prove universal correctness but can expose recurring problems. Repetition should trigger changes to input requirements, intermediate structures, or validation rather than only local repairs. The environment then learns from use.

Individuals can retain one independent check per task: recompute a point, test an exception, or revisit a source passage. These modest habits reinforce the status of output as a candidate and understanding as an active responsibility.

表 21 可维护项目的文件与记录
Table 21 Files and records for a maintainable project

类别 Category	保留内容 Content	用途 Purpose
原始资料 Sources	版本日期与访问条件 Versions dates access	核对事实与适用性 Check facts and applicability
语义结构 Semantics	节点关系与确认状态 Nodes edges confirmation	共享中英文与不同视图 Share across languages and views
计算数据 Data	输入变换结果与参数 Inputs transformations parameters	复算与异常检查 Reproduce and inspect
图形源 Graphic sources	规范代码与导出图 Specifications and exports	更换渲染与修改 Replace renderer and revise
正文成稿 Manuscript	引用图表与解释 Citations graphics interpretation	供受众阅读与判断 Support audience judgment
验证记录 Validation	检查者时间问题与修复 Reviewer time findings repairs	定位质量与维护责任 Locate quality and ownership

本报告的综合分析与设计建议 Report synthesis and design recommendations

22 综合案例研讨与练习
Integrated case discussions and exercises

22 1 夜间客诉积压的调查任务
Investigating night time complaint delays

任务背景是一家模拟酒店发现夜间客诉处理时间增加。教师提供去标识化事件表、岗位安排与简化政策，要求学生提出调查与改善方案。第一步只能画现状，不得立即建议增加 AI。学生需要区分接收、首次响应、部门接手与实际解决的时间，并检查是否把跨日事件计算错误。完成变量词典后，才选择时间图与责任流程。

第二步由 AI 提出三个解释，例如需求集中、交接延迟和权限不足。学生给每个解释列出支持、反证与缺失资料，并在假设图中明确状态。不能因某原因听起来合理就宣布成立。第三步选择一个可测试改进，例如给夜班人员提供可检索政策和明确升级提示，说明预期机制、监测指标和可能副作用。最后用未见过的异常事件检验流程。

教师评价重点是问题定义、时间口径、证据与干预之间的连接。若学生最终认为资料不足以选择改进，只要能说明缺口与下一步调查，也可以形成合格研究结论。这个练习让学生学习不确定条件下的专业判断，而不是把每个任务都写成成功方案。AI 的作用是辅助候选与比较，不能替代观察数据。

A simulated hotel reports longer complaint resolution at night. Students receive de-identified events, staffing, and simplified policies. They first map current practice without immediately recommending AI. Receipt, first response, handoff, and resolution times must be distinguished, including events crossing midnight. A variable dictionary precedes timelines and responsibility diagrams.

AI then proposes explanations such as concentrated demand, delayed handoff, and limited authority. Students list support, counterevidence, and gaps for each. Plausibility does not establish a cause. They choose a testable improvement, explain its mechanism, measures, and side effects, and test an unseen exception.

Assessment focuses on definitions, temporal measures, and links between evidence and intervention. A justified conclusion that more evidence is needed can be successful. The exercise develops judgment under uncertainty rather than requiring every task to end in a success story. AI supports candidate comparison without replacing observations.

22 2 营销增长建议的图形审计
A graphical audit of marketing claims

第二项练习提供一份由 AI 生成的模拟营销报告，其中故意包含五类问题：曝光与人数混用、累计与当期值混用、没有分母的转化率、把相关当因果、只展示乐观情境。学生先独立标记，再小组讨论。目标不是寻找模型所有语言毛病，而是定位会改变经营判断的图表问题，并给出最小修复。

修复后，学生重新写主要结论。如果原资料只能说明活动期间订单增加，就不能继续写活动导致新增订单。若缺少成本，则不能判断利润改善。若完成入住下降，预订增长也需要重新解释。AI 可以协助生成替代图，但每次替代都要保留同一数据口径，避免修复过程又改变事实。最后要求一个未参与绘图的同学解释图，检查是否真的更清楚。

该练习还可加入双语审计：英文图注写的是 revenue，中文正文却写利润；英文使用 estimated，中文删除了估计性质。学生需要把这些差异视为内容错误，而不只是翻译风格。视觉、数量与语言共同构成判断，任何一层变化都可能影响结论。报告由此成为综合素养训练材料。

The second exercise supplies a simulated AI marketing report with deliberate problems: exposures confused with people, cumulative values confused with period values, conversion without denominators, correlation presented as causation, and optimistic scenarios alone. Students first mark issues independently and then discuss the smallest consequential repairs.

After revision, they rewrite the conclusion. More orders during a campaign do not alone establish incremental effect. Missing cost prevents a profit claim. Falling completed stays require a different interpretation of rising bookings. AI can propose replacement charts, but definitions must remain stable. A peer who did not create the figure then explains it as a comprehension check.

A bilingual audit can add revenue translated as profit or an estimate qualifier omitted from Chinese. These are content differences rather than stylistic choices. Visual form, quantity, and language jointly determine interpretation.

22 3 课程生成 Skill 的压力测试
Stress testing a course generation skill

第三项练习让学生或教师评估一份视觉课程生成规范。测试输入包括资料不足的主题、存在矛盾的政策、需要精确数量的任务和学生基础差异较大的班级。观察系统是否追问、是否承认不确定、是否选择合适图形，以及是否保留人工核查。规范不能只在最顺利输入上通过，也需要在困难条件下表现可解释。

可以给出一个常见但过强的要求：“所有内容都转成图，九十分钟固定九页，全部自动完成。”测试者应指出哪些目标可能冲突：资料密度、活动时间、无障碍与精确解释。然后把硬规则改成默认值和例外条件，补充验收标准。AI 可以参与提出反例，但最终规范由课程负责人确认。这个练习帮助理解 Skill 是一种需要维护的工作规范，而非保证正确的魔法。

测试报告应列出至少一项保留规则、一项修改规则和一项暂不能判断的规则，并说明依据。若没有实际使用记录，就不能宣称效率已经提高某个百分比。可以提出后续试验方案与记录表。学生由此学习把愿景变成可检验承诺，也能更理性地面对平台宣传和自动化演示。

The third exercise evaluates a visual-course generation specification with insufficient sources, conflicting policies, quantitative tasks, and varied learner backgrounds. Reviewers observe clarification, uncertainty, representation choice, and human checks. A specification needs interpretable behavior under difficult conditions as well as smooth demonstrations.

A deliberately strong rule might require every item to become a graphic, exactly nine slides for ninety minutes, and complete automation. Reviewers identify conflicts involving density, activity time, accessibility, and precision. They replace rigid rules with defaults, exceptions, and acceptance criteria. AI proposes counterexamples while a course owner confirms the specification.

The report retains one rule, revises another, and leaves one unresolved with reasons. Without usage records, it cannot claim a measured efficiency gain. A proposed trial is appropriate. The task develops the ability to convert aspirations into testable commitments and evaluate automation claims.

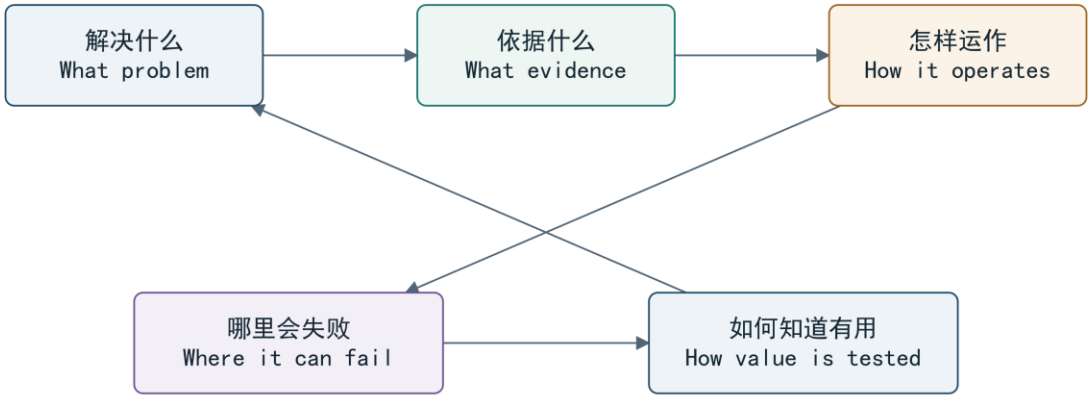


图 24 综合案例的五项答辩问题
Figure 24 Five questions for an integrated case defense
本报告提出的概念与设计示意 Conceptual design proposed in this report

22 4 一份综合答辩的组织方式
Organizing an integrated defense

综合答辩可以用十分钟完成：两分钟说明问题与证据，三分钟解释主要图，三分钟回答改变条件的问题，两分钟反思 AI 贡献与限制。学生提交问题图、证据表、方案流程、验证记录和修订说明。教师选择一项关键关系进行追问，确保团队不能只靠流畅展示。具体时间可按班级规模调整，但应保留独立解释与新条件检验。

评审应避免把口才当作全部能力。可以允许学生指图、标注或短写，给需要支持的学习者适当时间；核心仍是关系、证据与修改理由。小组成员分别承担不同追问，结合个人短测判断学习。AI 使用声明只要求与任务相关的信息，不要求公开无关私人内容。这样能够兼顾透明、学习与合理负担。

最终作品应能回答五个问题：解决什么，依据什么，怎样运作，哪里可能失败，如何知道有用。视觉思维帮助把这些问题放到共同结构中，AI 帮助降低构建和比较成本，研究方法帮助约束

结论，教师与学生共同承担解释和验证。这样的结合不依赖某一种产品，能够随工具变化继续发展，也最贴近 AI 导论培养行业问题解决能力的目标。

A ten-minute defense can devote two minutes to the problem and evidence, three to the principal representation, three to changed conditions, and two to AI's contribution and limits. Submitted materials include a problem map, evidence table, workflow, validation, and revisions. A reviewer probes a consequential relationship so that fluent presentation alone is insufficient. Timing can change while independent explanation and transfer remain.

Assessment should not equate eloquence with competence. Pointing, annotation, or short writing can provide alternative access with appropriate time. Relationships, evidence, and justified revision remain central. Different group members answer different questions, supplemented by individual assessment. Tool declarations concern relevant work rather than private unrelated content.

The final project answers five questions: what problem, what evidence, how it operates, where it can fail, and how usefulness will be assessed. Visual thinking supplies shared structure, AI lowers construction and comparison costs, research methods constrain conclusions, and people retain explanation and validation. This combination can continue across changing products and supports industry problem solving.

表 22 综合案例的任务与证据
Table 22 Integrated exercises and evidence

案例 Case	学生需要完成 Task	核心检验 Key test
夜间客诉 Night complaints	现状图与时间口径 Current map and time definitions	不把候选原因当事实 Keep hypotheses separate
营销审计 Marketing audit	找出并修复图表错误 Locate and repair errors	结论随证据调整 Revise conclusions with evidence
Skill 测试 Skill testing	困难输入与规则修订 Difficult inputs and revised rules	默认与例外清楚 Defaults and exceptions explicit
收益推演 Revenue scenario	改变参数并解释 Change and explain parameters	计算与经验有效性分开 Separate computation and validity
双语审查 Bilingual review	比较数字条件与术语 Compare numbers conditions terms	不确定与否定不丢失 Preserve uncertainty and negation
个人答辩 Individual defense	解释一条关系与新情境 Explain an edge and new context	理解不依赖小组代劳 Understanding beyond group assistance

本报告的综合分析与设计建议 Report synthesis and design recommendations

23 交互计算与可视推演
Interactive computation and visual simulation

23 1 从观看图形到操作模型
From viewing graphics to manipulating models

静态图适合保留与比较，交互图则适合探索变化。交互的价值来自用户能够提出一个假设、改变参数、观察结果并解释差异，而不只是拖动滑块。教学设计应先确定需要理解的关系，再决定允许调整什么。若每个参数都可随意变化而没有问题引导，学生可能玩得投入，却没有形成可迁移理解。因此，交互模型也需要目标、约束与检验。

可以采用四步操作：先预测，再改变，再比较，最后解释。比如收益案例中先问变动成本增加会让利润最优价格怎样变化，学生写下预测后调整参数，再查看曲线，最后解释差异。预测记录让教师看到原有模型，解释记录让教师判断是否只是跟随画面。AI 可以根据回答提出追问，但不能每次都把结果直接说完。

交互也应保留关键状态。一个学生发现特别情境后，可以保存参数、结果和说明，与其他人比较。否则图像随操作不断变化，思考证据容易丢失。建议提供默认状态、重置、状态命名和导出表，必要时保留两条曲线叠加。静态与交互并非相互替代，而是分别支持探索与记录。

Static graphics preserve and compare states, while interactive graphics support exploration. Value comes from proposing a hypothesis, changing a parameter, observing a result, and explaining the difference. Instruction should define the relationship before deciding which controls to expose. Unrestricted sliders without a guiding question may produce engagement without transferable understanding. Interaction still needs objectives, constraints, and checks.

A four-step sequence is prediction, change, comparison, and explanation. Students predict how higher variable cost affects a profit optimum, adjust it, inspect the curve, and explain the difference. Predictions expose prior models; explanations reveal whether learners merely followed the display. AI can ask further questions without giving every result immediately.

Important states should be retained with parameters and interpretation. Defaults, reset, named states, exported tables, and overlaid curves support comparison. Static and interactive forms serve complementary roles in exploration and record keeping.

23 2 用梯度下降理解模型更新
Understanding parameter updates through gradient descent

AI 导论可以用一个简单二次函数说明参数更新，而不必一开始进入大型神经网络。设损失是参数与目标值之差的平方，学生从某个初值出发，按梯度方向更新。图中同时显示函数曲线、当前位置和连续更新轨迹。这里的模型刻意简化，目的是理解步长、下降方向与迭代，不代表真实训练问题都具有同样形状。

学生可以比较小、中、大三种步长，观察收敛缓慢、较快接近与振荡或发散。AI 帮助把学生的观察整理成说明，计算工具负责产生轨迹。教师要求学生区分“曲线看起来下降”和“算法在设定条件下收敛”，并说明一个简单例子的结论不能直接推广到任意损失面。这样，图帮助理解抽象操作，同时保留数学条件意识。

课程还可以把同一过程表示为数值表、公式说明与轨迹图。学生选取一步手算，与图中的位置对照；再用文字解释为什么更新幅度会变化。三种表征承担不同任务：公式表达规则，表保留数值，图显示变化。AI 若在其中一处给出不一致结果，学生应能通过另外两种形式发现问题。这是多重表征互相核验的具体练习。

A simple quadratic can introduce parameter updates before a large neural network. Loss is the squared difference between a parameter and a target. A learner starts at an initial value and follows gradient updates. The display shows the curve, current position, and trajectory. The simplification teaches step size, direction, and iteration, not the claim that all training problems have this shape.

Small, medium, and large steps can illustrate slow progress, faster approach, and oscillation or divergence. AI organizes observations while computation produces the trajectory. Students distinguish an apparently descending curve from convergence under specified conditions and avoid generalizing one example to every loss surface.

The same process appears as a rule, numerical table, and trajectory. Learners calculate one step and locate it visually, then explain changing update magnitude. The rule specifies the operation, the table preserves values, and the graphic shows change. Inconsistency in one form can be detected through the others, providing a concrete mutual-checking exercise.

23 3 敏感性不是不确定性的全部
Sensitivity is not all of uncertainty

敏感性分析通常改变某个输入，观察结果怎样变化；不确定性还可能来自模型结构、测量误差、样本偏差和未知事件。一个滑块只改变需求斜率，并没有覆盖酒店经营的全部不确定性。报告中的交互示例应明确“哪些变化被允许、哪些仍未建模”，否则使用者容易把可调整范围误当成现实的完整边界。

可以让学生建立一个不确定性表，分别列出已知参数范围、资料质量、结构假设和外部变化。数量模型能够处理其中某些项目，其他项目需要情境讨论、额外数据或组织判断。AI 可帮助枚举，但不能凭语言把未知转换为精确概率。若缺少依据，应使用清楚的情境标签，避免以小数点制造权威感。

敏感性结果也应连接行动。若结论对一个参数非常敏感，下一步可能是优先收集该参数的信息；若在较宽范围内结论稳定，可能支持进一步小试验。图的目的不是展示所有曲线，而是帮助决定哪里值得投入调查。学生应说明自己如何从图走向下一步，而不是只提交一张颜色丰富的参数地图。

Sensitivity analysis changes an input and observes an output. Uncertainty also concerns model structure, measurement, sampling, and unknown events. A demand-slope slider does not cover all hotel uncertainty.

An interactive example should state both permitted variations and omitted mechanisms so that adjustable bounds are not mistaken for reality's boundaries.

Students can distinguish parameter ranges, evidence quality, structural assumptions, and external changes. Some can be quantified; others need scenarios, new data, or judgment. AI can enumerate possibilities but should not assign precise probabilities without evidence. Clear scenario labels are preferable to unsupported decimal precision.

Results should guide action. Strong dependence on one parameter may justify collecting better information. Stability over a range may justify a small trial. The aim is to identify where further investigation matters, not display every possible curve. Learners explain the next step supported by the graphic.

23 4 交互作品的验收标准

Acceptance criteria for interactive artifacts

验收交互作品时，至少检查参数单位、有效范围、默认状态、极端情况、重置和输出记录。图形每次变化都应来自同一套清楚规则，不能随机制造看似合理的趋势。若系统使用随机模拟，应保留种子或重复规则，并解释单次轨迹与总体分布的区别。学生应能够根据一个保存状态复现主要结果。

可访问性也属于验收。滑块应有可读数值和替代输入，颜色应有标签，动态图应允许暂停，手机界面应能读到关键说明。复杂交互还应配静态总结，使无法操作的人仍能理解主要关系。最后要求一个未参与制作的人完成指定任务并说明结果，检查工具是否真正支持理解。技术运行与学习可用性需要分别通过。

Acceptance checks include units, valid ranges, defaults, extremes, reset, and recorded output. Changes must follow explicit rules rather than randomly plausible trends. Random simulation needs a seed or reproducible procedure and a distinction between one trajectory and a distribution. Saved states should reproduce important results.

Controls need readable values and alternative input, labels alongside color, pause for motion, and accessible explanations. A static summary should preserve the main relationship for those unable to interact. An independent user then completes a task and interprets the result. Technical operation and learning usability require separate checks.

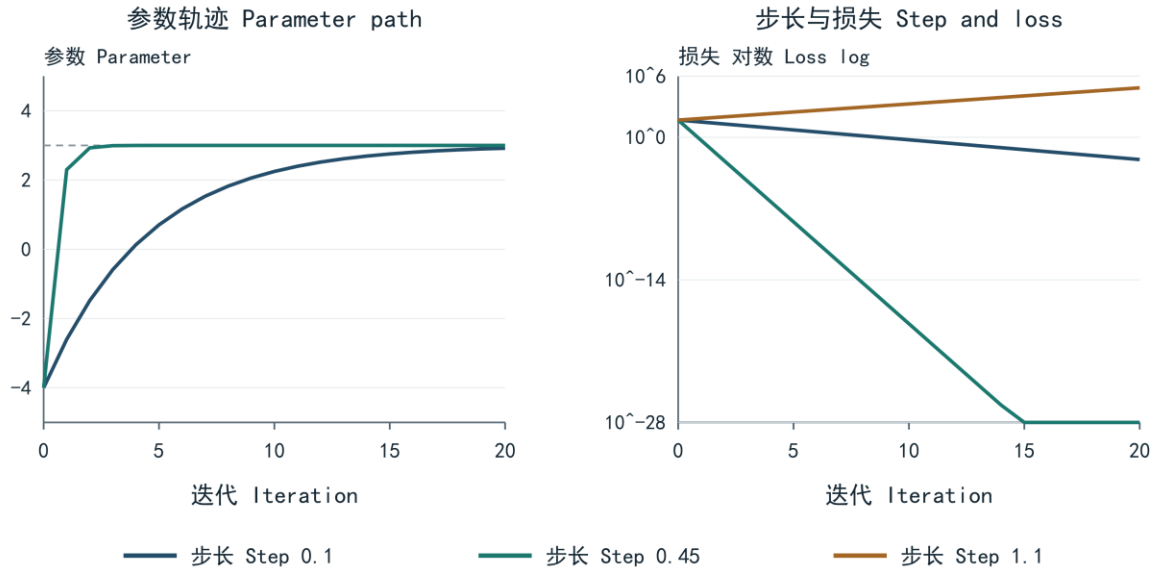


图 25 二次函数中的梯度更新

Figure 25 Gradient updates for a quadratic objective

本报告计算的教学示例 目标参数 3 初值负 4 损失为参数误差平方 不是大型模型训练的完整表示 Computed teaching example not a complete model of training

零损失设置显示下限以便使用对数尺度 不表示实际损失仍为正

Zero loss uses a display floor for the logarithmic axis and does not imply positive residual loss.

表 23 交互模型的学习与技术验收

Table 23 Learning and technical checks for interaction

维度 Dimension	检查 Check	证据 Evidence
目标 Objective	操作对应何种理解 What understanding does action serve	预测与解释记录 Predictions and explanations
规则 Rule	变化是否来自明确计算 Does change follow computation	公式代码与复算 Rule code and independent calculation
范围 Range	极端值是否仍有效 Are extremes valid	边界与错误提示 Bounds and error messages
状态 State	能否保存重置与比较 Can states be saved reset compared	导出参数与结果 Exported parameters and results
不确定性 Uncertainty	遗漏了哪些现实因素 Which factors are omitted	假设与限制说明 Assumptions and limits
访问 Access	是否有键盘文本静态替代 Keyboard text and static alternatives	独立用户完成任务 Independent user task

本报告的综合分析与设计建议 Report synthesis and design recommendations

24 综合讨论与优先行动
Integrated discussion and priority actions

24 1 哪些结论较稳健 哪些仍需研究
Stable conclusions and unresolved questions

相对稳健的判断是：表征能够改变信息搜索与推理条件，学习者需要主动组织与整合，视觉任务效果取决于设计与支持，能力差异不应被简化成固定学习类型。较不确定的是：AI 代画图在何种条件下能保留学习收益，多模态视觉中间过程怎样帮助人理解，以及长期使用是否改变独立推理习惯。这些不同确定程度应该在报告、课程与产品说明中保持可见。

技术研究已经提供多种可行路线，但可行不等于适用。一个模型能完成基准任务，不说明它能读懂每张业务图；一个原型让少量用户成功，也不说明大规模班级可以直接采用；一个工具支持导出，也不说明所有字体与交互都能完整保留。应用者需要把技术能力转成自己的任务测试。研究报告的价值在于提供判断框架，而不是替所有场景给出同一选择。

本报告对附件的主要修订，正是把有吸引力的愿景变成带条件的方案。视觉思维可以贯穿课程，但不要求所有内容图形化；Skill 可以复用经验，但需要验收与维护；专家协作可以分担职责，但仍需要独立检查与负责人；AI 可以提高部分生产效率，但效率与学习必须分别测量。保留这些条件，能够让实践更容易成功，也更容易发现失败。

Relatively stable conclusions concern the effects of representation on search and inference, the need for active organization and integration, design-dependent learning outcomes, and the inadequacy of fixed learning-type labels. Less certain questions concern when AI drawing preserves learning, how visual intermediate steps support human understanding, and how sustained use affects independent habits. Different levels of certainty should remain visible in reports, courses, and products.

Technical research offers feasible routes, but feasibility does not establish suitability. Benchmark success does not guarantee every business graphic, a small prototype study does not establish classroom scalability, and export support does not guarantee full fidelity. Users need task-specific tests. A report provides a judgment framework rather than one universal selection.

The principal revision to the attachments turns an attractive vision into conditional proposals. Visual thinking can span a course without making every item graphical. Skills can reuse experience with maintenance and acceptance. Roles can divide work with independent checks and ownership. Production efficiency and learning need separate measurement. Conditions improve the ability to identify both success and failure.

24 2 对教师的优先行动
Priority actions for instructors

教师可以从三件小事开始。第一，选择一个学生经常混淆的概念，把它变成需要解释关系的图，而不只是加插画。第二，给学生一张有意保留错误或缺口的图，让他们根据资料修订。第三，

安排一个不使用 AI 的新情境任务，检查理解能否迁移。这些活动不需要先建设完整平台，却能够产生直接的教学证据。

下一步建立术语表、表征说明和评分样例。术语表让中英文与不同课时保持一致；表征说明明确每张图要回答的问题；评分样例使学生知道准确性与推理为什么比装饰重要。AI 可以帮助形成初稿，教师用真实误解与作业修订。一个学期逐步积累后，再决定哪些步骤值得封装成 Skill，哪些仍应保留现场判断。

教师还应记录自己的时间与质量变化。若工具只让课件更复杂，却没有节约准备时间或改善理解，就需要调整。若工具节约排版时间，可以把收益投入反馈和学生讨论。视觉思维改革不必表现为每节课都有更多图，而应表现为学生更能解释关系、发现错误和提出有依据的行动。这是最接近课程目标的检验。

Instructors can begin with three small actions: represent a commonly confused concept through relationships, ask learners to repair a sourced diagram with a deliberate gap or error, and assess an unaided transfer task. These provide instructional evidence without requiring a complete platform.

Next build terminology, representation briefs, and scoring examples. Terms support consistency, briefs identify each graphic's question, and examples prioritize reasoning over decoration. AI drafts can be revised using actual misconceptions and student work. Repeated use then identifies which steps deserve reusable skills and which need situated judgment.

Record preparation time and quality as well. Greater complexity without time or learning value calls for revision. Saved layout time can support feedback and discussion. Improvement should appear in learners' explanations, error detection, and evidence-based action rather than simply in more graphics per lesson.

24 3 对组织与开发者的优先行动
Priority actions for organizations and developers

组织应先确定一个有限任务、一个责任人和一组验收标准。优先选择资料边界清楚、后果可控、能够比较基线的场景。把来源与语义结构保存下来，采用可替换输出，保留人工出口。不要把系统演示效果直接当成实际收益，也不要仅按生成文件数量评价使用情况。真正价值应连接业务判断与工作结果。

开发者可以优先支持四种能力：让用户指出图中的具体元素，显示其来源和状态，允许局部修改，并在修改后提示受影响内容。这比不断生成整张新图更能支持持续工作。对数量图，应让用户查看转换后数据；对流程，应让用户测试例外；对教学，应保留学生先尝试的机会。可解释与可控需要进入界面，而不能只写在说明书中。

在组织扩展时，维护与退出能力同样重要。系统应能导出核心结构，数据权限变化后能够限制相关任务，旧版本能够被识别，错误能够被追溯。若某项自动化不再合适，团队仍应能够继续使用已有知识。以这种方式建设的视觉 AI 系统，更可能成为工作基础，而不是一次性演示项目。

Organizations should choose a bounded task, an owner, and acceptance criteria, preferably with clear evidence, manageable consequences, and a baseline. Preserve sources and semantics, support replaceable

outputs, and retain human exits. Demonstration quality and output count are not equivalent to operating value.

Developers can prioritize element-level selection, visible provenance and status, local revision, and affected-content warnings. These support continuing work better than repeatedly regenerating a whole image. Quantitative tools should expose transformed data, workflows should test exceptions, and educational tools should preserve learner attempts. Control and explanation belong in interaction.

Maintenance and exit matter during expansion. Core structures need export, changed permissions need enforcement, old versions need identification, and errors need traceability. Teams should retain knowledge if an automation becomes unsuitable. This supports durable infrastructure rather than a temporary demonstration.

24 4 最终结论
Final conclusions

视觉思维的核心贡献，是让对象、关系、变化和证据进入一个能够共同检查的空间。AI 的核心机会，是降低表达、转换、生成候选与重复制作的成本。二者结合的价值，取决于人是否仍在定义问题、选择表征、检验依据和承担后果。只有图片增加而没有这些活动，不能称为思维能力提升；只有自动化流程而没有证据与责任，也难以形成可靠应用。

对于 AI 导论，最值得形成的不是一套固定画风，而是一条稳定学习路径：从行业问题出发，以资料约束概念，用图表组织关系，以计算和测试检查方案，在新条件中证明迁移。教师、学生、AI 与计算工具各自承担清楚职责，过程留下可复查证据。这样的课程既能够回应认知多样性，也能够培养 AI 时代共同需要的专业判断。

本报告因此建议把视觉思维作为 AI 素养的一部分持续建设：会看图，会做图，更会问图中的关系凭什么成立；会调用 AI，也会在必要时离开 AI 完成解释。最终目标是使人面对更丰富、更快速、更具说服力的信息环境时，仍能清楚理解、审慎判断，并把想法转化为可以验证的行动。

Visual thinking brings objects, relationships, change, and evidence into a space for shared inspection. AI lowers the cost of expression, transformation, candidate generation, and repeated production. Their value depends on people continuing to define questions, choose representations, verify evidence, and accept consequences. More images alone do not establish stronger thinking, and automation without evidence and responsibility does not establish reliable application.

An introductory AI course needs a stable learning path rather than one visual style: begin with an industry problem, constrain concepts through sources, organize relations visually, test proposals through computation, and demonstrate transfer under changed conditions. Instructors, learners, AI, and computational tools retain defined responsibilities and reviewable evidence. This supports diversity while developing shared professional judgment.

Visual thinking should therefore become a continuing part of AI literacy: reading and making graphics while asking why their relationships hold, using AI while retaining independent explanation. The objective is clearer understanding, careful judgment, and verifiable action in an increasingly persuasive information environment.

表 24 不同角色的优先行动
Table 24 Priority actions by role

角色 Role	优先行动 Priority	判断是否有效 Evidence of value
教师 Instructor	选一个难点设计建构与迁移 Design construction and transfer for one difficulty	学生解释与独立表现 Explanation and independence
学生 Learner	保留首次图修订与理由 Keep initial map revisions and reasons	能发现错误并改到新任务 Detect error and adapt
课程负责人 Course owner	统一术语来源与验收 Align terms sources acceptance	交付一致并可维护 Consistent maintainable outputs
组织管理者 Manager	先做有限试点与基线 Run a bounded pilot and baseline	完整成本与实际结果 Total cost and actual outcomes
开发者 Developer	支持局部修改来源和状态 Support local edits provenance and status	用户能控制与核查 Users can control and verify
研究者 Researcher	区分任务机制与证据层级 Separate task mechanism and evidence	可复现的条件性结论 Reproducible conditional conclusions

本报告的综合分析与设计建议 Report synthesis and design recommendations

参考文献与延伸阅读
References and further reading

下列条目为英文研究及官方资料。预印本与技术文档不作为人类学习成效的直接证据。部分条目依据可访问摘要或正式记录使用 访问核查日期为 2026 年 10 月 17 日。

Entries include English research and official sources. Preprints and documentation are not direct evidence of human learning gains. Some entries are used through accessible abstracts or formal records. Access verification date 17 October 2026.

[1] Arnheim, R. (1969). Visual Thinking. University of California Press.
来源 Source https://www.ucpress.edu/books/visual-thinking/paper
Theoretical monograph

[2] Grandin, T. (2022). Visual Thinking The Hidden Gifts of People Who Think in Pictures Patterns and Abstractions. Riverhead Books.
来源 Source https://www.penguinrandomhouse.com/books/673207/visual-thinking-by-temple-grandin/9780593418383/
Research-informed trade book and personal perspective

[3] Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is Sometimes Worth Ten Thousand Words. Cognitive Science, 11, 65–100.
来源 Source https://doi.org/10.1111/j.1551-6708.1987.tb00863.x
Computational and theoretical analysis

[4] Blazhenkova, O., & Kozhevnikov, M. (2009). The new object spatial verbal cognitive style model Theory and measurement. Applied Cognitive Psychology, 23, 638–663.
来源 Source https://doi.org/10.1002/acp.1473
Psychometric research

[5] Blazhenkova, O., Becker, M., & Kozhevnikov, M. (2011). Object spatial imagery and verbal cognitive styles in children and adolescents Developmental trajectories in relation to ability. Learning and Individual Differences, 21, 281–287.
来源 Source https://doi.org/10.1016/j.lindif.2010.11.012
Developmental correlational research

[6] Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning Styles Concepts and Evidence. Psychological Science in the Public Interest, 9, 105–119.
来源 Source https://pubmed.ncbi.nlm.nih.gov/26162104/
Critical evidence review

[7] Uttal, D. H., et al. (2013). The malleability of spatial skills A meta analysis of training studies. Psychological Bulletin, 139, 352–402.
来源 Source https://pubmed.ncbi.nlm.nih.gov/22663761/
Meta-analysis

[8] Mayer, R. E. (2021). Cognitive Theory of Multimedia Learning. The Cambridge Handbook of Multimedia Learning, third edition, chapter 5.
来源 Source https://www.cambridge.org/core/books/abs/cambridge-handbook-of-multimedia-learning/cognitive-theory-of-multimedia-learning/3A0E495CDB9D9D84665FAEFCBAB51D05
Theory and research synthesis

[9] Ainsworth, S. (2006). DeFT A conceptual framework for considering learning with multiple representations. Learning and Instruction, 16, 183–198.
来源 Source https://www.sciencedirect.com/science/article/pii/S0959475206000259
Conceptual framework

[10] Rau, M. A. (2017). Conditions for the Effectiveness of Multiple Visual Representations in Enhancing STEM Learning. Educational Psychology Review.
来源 Source https://doi.org/10.1007/s10648-016-9365-3
Review

[11] Fiorella, L., & Zhang, Q. (2018). Drawing Boundary Conditions for Learning by Drawing. Educational Psychology Review.
来源 Source https://doi.org/10.1007/s10648-018-9444-8
Review

[12] Cromley, J. G., Du, Y., & Dane, A. P. (2020). Drawing to Learn Does Meta Analysis Show Differences Between Technology Based Drawing and Paper and Pencil Drawing. Journal of Science Education and Technology, 29, 216–229; first online 2019.
来源 Source https://doi.org/10.1007/s10956-019-09807-6
Meta-analysis

[13] Leutner, D., & Biele, J. (2025). Without Integration Everything Is Nothing A Meta Analysis of the Effectiveness of Instructional Support for Drawing to Learn. Educational Psychology Review, 37.
来源 Source https://consensus.app/papers/without-integration-everything-is-nothing-a-metaanalysis-leutner-biele/b38ed9708ed65cdfa2cb6c76ee19a585/
Meta-analysis; abstract and bibliographic record consulted

[14] Rexigel, E., Kühn, J., & Becker-Genschow, S. (2024). The More the Better A Systematic Review and Meta Analysis of the Benefits of More than Two External Representations in STEM Education. Educational Psychology Review.
来源 Source https://doi.org/10.1007/s10648-024-09958-y
Systematic review and meta-analysis

[15] Zhang, Q., & Fiorella, L. (2021). Learning by drawing When is it worth the time and effort. Contemporary Educational Psychology.
来源 Source https://consensus.app/papers/learning-by-drawing-when-is-it-worth-the-time-and-effort-zhang-fiorella/a4335cc625825348a99de7adabfc408a/
Two experiments; abstract-level evidence

[16] Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation can it facilitate. International Journal of Human Computer Studies.
来源 Source https://doi.org/10.1006/ijhc.2002.1017
Review and design analysis

[17] Kirsh, D., & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. Cognitive Science, 18, 513–549.
来源 Source https://doi.org/10.1207/s15516709cog1804_1
Situated cognition analysis

[18] Cleveland, W. S., & McGill, R. (1984). Graphical Perception Theory Experimentation and Application to the Development of Graphical Methods. Journal of the American Statistical Association, 79, 531–554.
来源 Source https://doi.org/10.1080/01621459.1984.10478080
Graphical perception experiments

[19] Sutherland, I. E. (1963). Sketchpad A Man Machine Graphical Communication System. MIT doctoral dissertation.
来源 Source https://dspace.mit.edu/entities/publication/b5e8025c-c8b2-4843-84e0-76db824e07e6
Technical dissertation

[20] Hu, Y., et al. (2024). Visual Sketchpad Sketching as a Visual Chain of Thought for Multimodal Language Models. NeurIPS 37.
来源 Source https://papers.neurips.cc/paper_files/paper/2024/hash/fb82011040977c7712409fbdb5456647-Abstract-Conference.html
Peer-reviewed technical benchmark study

[21] Dibia, V. (2023). LIDA A Tool for Automatic Generation of Grammar Agnostic Visualizations and Infographics using Large Language Models. ACL System Demonstrations.
来源 Source https://aclanthology.org/2023.acl-demo.11/
Research system

[22] Wang, C., Thompson, J., & Lee, B. (2023). Data Formulator AI powered Concept driven Visualization Authoring. IEEE VIS; Microsoft Research publication record.
来源 Source https://www.microsoft.com/en-us/research/publication/data-formulator-ai-powered-concept-driven-visualization-authoring/
Research prototype and small user study

[23] Microsoft Research (2024). Data Formulator Exploring how AI can help analysts create rich data visualizations. Research project account.
来源 Source https://www.microsoft.com/en-us/research/blog/data-formulator-exploring-how-ai-can-help-analysts-create-rich-data-visualizations/
Primary project description

[24] Masry, A., et al. (2022). ChartQA A Benchmark for Question Answering about Charts with Visual and Logical Reasoning. Findings of ACL, 2263–2279.
来源 Source https://aclanthology.org/2022.findings-acl.177/
Benchmark

[25] Rahmanzadehgervi, P., et al. (2024; revised 2025). Vision language models are blind Failing to translate detailed visual features into words. arXiv 2407.06581, version 6.
来源 Source https://arxiv.org/abs/2407.06581
Preprint benchmark; version-specific findings

[26] Lu, P., et al. (2024). MathVista Evaluating Mathematical Reasoning of Foundation Models in Visual Contexts. ICLR; arXiv 2310.02255.
来源 Source https://arxiv.org/abs/2310.02255
Benchmark

[27] Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful A systematic review and meta analysis. Nature Human Behaviour, 8, 2293–2303.
来源 Source https://doi.org/10.1038/s41562-024-02024-1
Preregistered systematic review and meta-analysis

[28] Lee, H. P., et al. (2025). The Impact of Generative AI on Critical Thinking Self Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. CHI; DOI 10.1145/3706598.3713778.
来源 Source https://www.microsoft.com/en-us/research/publication/the-impact-of-generative-ai-on-critical-thinking-self-reported-reductions-in-cognitive-effort-and-confidence-effects-from-a-survey-of-knowledge-workers/
Self-report survey

[29] Bastani, H., et al. (2025). Generative AI without guardrails can harm learning Evidence from high school mathematics. PNAS, 122, e2422633122.
来源 Source https://pmc.ncbi.nlm.nih.gov/articles/PMC12232635/
Randomized controlled trial

[30] NIST (2024). Artificial Intelligence Risk Management Framework Generative Artificial Intelligence Profile. NIST AI 600-1.
来源 Source https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
Official risk framework

[31] UNESCO (2023). Guidance for generative AI in education and research. UNESCO.
来源 Source https://doi.org/10.54675/EWZM9535
Policy guidance

[32] W3C Web Accessibility Initiative (accessed 2026). Complex Images. Images Tutorial.
来源 Source https://www.w3.org/WAI/tutorials/images/complex/
Official accessibility guidance

[33] W3C Web Accessibility Initiative (accessed 2026). Understanding Success Criterion 1 4 11 Non text Contrast. WCAG 2.2 guidance.
来源 Source https://www.w3.org/WAI/WCAG22/understanding/non-text-contrast.html
Official accessibility guidance

[34] Mermaid (accessed 2026). Flowcharts Syntax and Mindmap. Official documentation.
来源 Source https://mermaid.js.org/syntax/flowchart.html
Technical documentation

[35] Vega Lite (accessed 2026). Visualization specification and transformations. Official documentation.
来源 Source https://vega.github.io/vega-lite/docs/transform.html
Technical documentation

[36] Wolfram Research (accessed 2026). Manipulate. Wolfram Language Documentation.
来源 Source https://reference.wolfram.com/language/ref/Manipulate.html
Technical documentation

[37] Wolfram Research (accessed 2026). Graph. Wolfram Language Documentation.
来源 Source https://reference.wolfram.com/language/ref/Graph.html
Technical documentation

[38] Coalition for Content Provenance and Authenticity (version 2.2). Content Credentials Technical Specification. C2PA.
来源 Source https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html
Technical specification

[39] Dechamps, T., & Skulmowski, A. (2025). The Effective Design of Tasks Involving Learning by Drawing Current Trends and Methodological Progress in Research on Drawing to Learn. Educational Psychology Review, 37.
来源 Source https://consensus.app/papers/details/f2767c4696705dba86dbfd0bb5a073dc/
Review; abstract-level evidence

[40] Kim, S., et al. (2025). Can the effects of generative drawing on lasting learning be optimized through integration of retrieval practice. Learning and Instruction.
来源 Source https://consensus.app/papers/details/2cee5cccb46a57c3b116089217088891/
Experimental study; abstract and bibliographic record consulted

[41] Wu, J., et al. (2026). Visual Generation Unlocks Human Like Reasoning through Multimodal World Models. arXiv 2601.19834; preprint.
来源 Source https://arxiv.org/abs/2601.19834
Preprint technical study

[42] Zhang, H., et al. (2025). Latent Sketchpad Sketching Visual Thoughts to Elicit Multimodal Reasoning in MLLMs. arXiv 2510.24514; preprint.
来源 Source https://arxiv.org/abs/2510.24514
Preprint technical study

附件来源与对应说明

Supplied materials and their role

A1 「AI 导论·视觉化课程生成」Skill.docx

Course generation proposal covering maps cards workflows manuals and activities

A2 visual thinking-skill-到 workbuddy-expert.docx

Discussion of drawing capabilities skills expert roles and production workflows

A3 视觉思维.docx

Grandin-inspired discussion and initial ideas for an industry-oriented introductory AI course

附件作为观点与场景材料使用。报告保留其视觉建构与行业问题导向 同时对认知分类 自动化比例 时间承诺和免干预判断作出研究性修订。

The attachments supply proposals and context. Their visual construction and industry orientation are retained while cognitive labels automation coverage timing claims and intervention assumptions are revised through critical analysis.

图表索引

List of figures and tables

图 1 可检验的视觉推理闭环
Figure 1 A cycle of inspectable visual reasoning
图 2 相关概念的边界与联系
Figure 2 Related concepts and their boundaries
图 3 外部表征如何支持理解
Figure 3 How external representations can support understanding
图 4 绘图学习元分析中的不同结果
Figure 4 Different outcomes in a drawing to learn meta analysis
图 5 尺度改变视觉印象
Figure 5 Scale changes visual impression
图 6 把流畅建议变成审查对象
Figure 6 Turning fluent advice into a reviewable object
图 7 AI 支持视觉思维的主要路线
Figure 7 Main routes for AI supported visual thinking
图 8 来源语义呈现与验证分层
Figure 8 Sources semantics presentation and validation
图 9 人机协作中的解释修订与反例
Figure 9 Explanation revision and counterexamples
图 10 酒店 RAG 的证据与升级路径
Figure 10 Evidence and escalation in hotel retrieval
图 11 收入与利润最优点可以不同
Figure 11 Revenue and profit can have different optima
图 12 营销观察假设与因果检验
Figure 12 Marketing observation hypotheses and causal tests
图 13 AI 导论的五层能力地图
Figure 13 Five layers of an introductory AI course
图 14 九十分钟课时的活动节奏
Figure 14 Activity sequence for a ninety minute lesson
图 15 课程生产职责与返工路径
Figure 15 Production responsibilities and rework
图 16 AI 辅助绘图的四条件研究设计
Figure 16 Four conditions for studying AI assisted diagrams

图 17 视觉 AI 风险的三个来源
Figure 17 Three sources of risk in visual AI
图 18 十二周试点与继续或退出
Figure 18 Twelve week pilot with continuation or exit
图 19 提示模板如何进入可检查工作流
Figure 19 Prompts within an inspectable workflow
图 20 视觉 AI 研究路线的层次
Figure 20 Levels of visual AI research approaches
图 21 科研中的问题证据与结论索引
Figure 21 Linking research questions evidence and conclusions
图 22 共同语义支持不同组织视图
Figure 22 Shared semantics support different organizational views
图 23 可维护的视觉知识项目
Figure 23 A maintainable visual knowledge project
图 24 综合案例的五项答辩问题
Figure 24 Five questions for an integrated case defense
图 25 二次函数中的梯度更新
Figure 25 Gradient updates for a quadratic objective
表 1 附件观点的保留与修订
Table 1 Retaining and revising attachment proposals
表 2 常见表征与问题匹配
Table 2 Matching representations to questions
表 3 个体差异与教学响应
Table 3 Responding to individual differences
表 4 关键学习证据及解释边界
Table 4 Selected learning evidence and its limits
表 5 图表制作与可访问性核查
Table 5 Graphic and accessibility checks
表 6 人机分工与能力评价
Table 6 Division of work and capability assessment
表 7 AI 实现路线及扩展能力比较
Table 7 AI implementation routes and extensions
表 8 分层架构的错误定位
Table 8 Locating errors by architecture layer
表 9 一次视觉推理工作会的产物
Table 9 Artifacts from a visual reasoning session

表 10 合成酒店情境的计算结果
Table 10 Computed results for a synthetic hotel scenario
表 11 营销与文旅金融的判断边界
Table 11 Interpretation boundaries in marketing and tourism finance
表 12 十六周课程建议框架
Table 12 Proposed sixteen-week course
表 13 九十分钟课时安排
Table 13 Ninety-minute lesson sequence
表 14 课程生产职责与验收
Table 14 Production responsibilities and acceptance
表 15 视觉推理分析性量规
Table 15 Analytic rubric for visual reasoning
表 16 试点风险与退出条件
Table 16 Pilot risks and exit conditions
表 17 提示模板的输入与验收
Table 17 Prompt inputs and acceptance checks
表 18 核心术语中英文对照
Table 18 Core bilingual terminology
表 19 文献提取与证据审查字段
Table 19 Literature extraction and review fields
表 20 组织协作中的可见记录
Table 20 Visible records for collaboration
表 21 可维护项目的文件与记录
Table 21 Files and records for a maintainable project
表 22 综合案例的任务与证据
Table 22 Integrated exercises and evidence
表 23 交互模型的学习与技术验收
Table 23 Learning and technical checks for interaction
表 24 不同角色的优先行动
Table 24 Priority actions by role