# 钥匙孔效应：为什么聊天界面不适用于数据分析——人机混合交互的认知科学框架

**Mohan Reddy**
首席科学家，Cornerstone AI
Labs mreddy@csod.com

arXiv:2602.00947v1 [cs.AI] 1 Feb 2026

**摘要**

聊天已成为AI辅助数据分析的默认界面。但对于多步骤、依赖状态的分析任务而言，这种做法是错误的。基于 Woods（1984）提出的钥匙孔效应（透过狭窄视口查看大型信息空间所产生的认知成本），本文指出，聊天界面会通过五种机制系统性地降低分析表现：（1）内容持续位移，破坏海马空间记忆系统；（2）隐藏状态变量超出工作记忆容量（负荷下为 ∼4组块）；（3）强制言语表达引发言语遮蔽，削弱视觉模式识别；（4）线性文本流阻碍认知行动和认知卸载；（5）序列化代价随数据维度增加而上升。本文将认知过载形式化为 $O = \max(0,\ m - v - W)$，其中 $m$ 表示与任务相关的项目数，$v$ 表示可见项目数，$W$ 表示工作记忆容量。当 $O > 0$时，错误概率会上升，分析偏差（锚定效应、确认偏差、变化盲视）也会加剧。本文提出八种混合式设计模式，以应对这些问题：生成式用户界面、无限画布、指示性交互、状态轨道、幽灵图层、备料式布局、语义缩放和概率式用户界面。每种模式都针对特定的认知瓶颈，同时保留自然语言，用于意图指定和综合。对于有指导的任务，经过良好搭建、能够编码专家先验的对话系统或可降低认知负荷；这一框架最适用于开放式探索。文章最后提出可证伪假设和实验范式，以供实证验证。

**关键词：** 对话式界面；认知负荷；空间认知；工作记忆；双重编码；言语遮蔽；认知行动；分布式认知；生成式 UI；形态界面

## 1 引言

ChatGPT 于 2022年11月推出，聊天突然成了一切事务的界面：写作、编程、分析、研究。其隐含假设很简单：如果人工智能能够理解自然语言，那么自然语言必定是与人工智能交互的最佳方式。

这一假设忽略了 40 年的人机交互研究和一个世纪的认知心理学研究。

我们正处于人工智能交互的拟物化阶段。早期网站看起来像报纸的整版版面。早期电子书会模拟翻页。新技术通常会先模仿自己所取代的事物的形式，然后才找到自身原生的表达方式。基于聊天的人工智能模仿人类对话，这是一种熟悉而舒适的隐喻，但对话并非探索数据的原生表达方式。这种错位并非审美问题，而是认知问题。

从命令行转向图形界面在20世纪80年代并非审美上的转变，而是认知上的转变。Shneiderman（1983年）指出，直接操作界面利用人类的空间认知来降低记忆负荷。识别胜过回忆。指点胜过输入。看见胜过记住。基于聊天的人工智能却逆转了这些优势。

我认为，聊天适用于分析工作的两个阶段：**表达意图** 和**综合结果**。但它在构成大部分实际工作的三个阶段中表现不佳：**探索**（搜寻、筛选、假设显现）、**比较**（并列评估备选方案）以及**下钻**（迭代精炼和溯源追踪）。

该钥匙孔效应由Woods（1984）在过程控制显示器的上下文中首次识别，概括了这一失败。Woods观察到，当虚拟信息空间超出显示器的视口容量时，用户一次只能看到一小部分，从而导致导航负担以及

“迷失”现象。Bennett和Flach（2012）后来将视觉动量正式确立为一种设计原则，通过空间专用性、感知地标和显示重叠来抵消该效应。

用于数据分析的聊天界面是钥匙孔问题的一种极端情形，在某些方面甚至是一种倒退。与 Woods 所分析的显示系统不同——那些系统至少允许导航和空间组织——聊天强加了一种根本性的线性结构，从而加剧了原有问题：内容不断被挤出、空间信息被迫言语化，以及认知行动受阻。你试图通过一个狭窄的滚动窗口去理解庞大的多维数据集。你失去了空间上下文。你无法并排比较。先前的发现从屏幕顶部消失。你忘记哪些筛选器处于激活状态。你在本应点击时却在打字。

本文将界面与认知的失配形式化为一种过载函数，综合六种认知科学框架（空间认知、工作记忆、双重编码、言语遮蔽、认知行动、分布式认知）来解释聊天为何无法胜任分析工作，梳理了八种混合式设计模式及其认知理据，并追溯了一条形态演变轨迹：从当前带有制品的聊天，经由一次性即时界面，走向环境式面向对象人工智能。该研究议程提出了可供实证验证的可证伪假设。

## 2 形式化钥匙孔效应 t

设一个分析时刻需要维持一组同时相关的项目：状态变量（筛选器、时间窗口、队列）、制品（图表、表格、模型输出）以及假设（候选解释、竞争性叙事）。

**定义 1**（认知过载：基础模型）。定义：

- $m$ = 正确推理所需的相关项目数量
- $v$ =外部可见的项目数量（持久、可一瞥）
- $W$ = 以组块计的工作记忆容量（$\approx 4$负荷下）（*Cowan, 2001*）

内部维持负荷为 $L_{internal}$ = $\max(0, m - v)$，认知过载为：

$$\boxed{O = \max(0, m - v - W)} \tag{1}$$

这个基础模型将项目视为同质，并将可见性视为二元变量。现实系统则更加复杂：固定图表对 $v$ 的贡献大于最小化面板；核心假设的权重高于次要筛选条件。扩展模型将纳入项目权重 $w_i$、反映显著性和可访问性的分级可见性函数 $\sigma_i \in [0, 1]$，以及受专业知识和组块化影响的动态 $W$：

$$O_{\text{ext}} = \max\left(0, \sum_i w_i - \sum_i w_i \sigma_i - W(e, c)\right) \tag{2}$$

其中 $e$表示专业知识，$c$则表示外部组块辅助工具。基础模型（公式 1）是其中 $w_i = 1$ 且 $\sigma_i \in \{0, 1\}$的退化情形。为便于处理，我在全文中使用基础模型，将其视为一阶近似；它能够捕捉定性动态，但同时也承认，实证研究需要采用扩展公式。

**对 $m$ 和 $v$进行操作化。** 出于实验目的，$m$ 可以通过以下项目的数量来衡量：活动筛选器、比较所需的并行数据视图、正在考虑的假设卡片，以及状态变量（时间范围、群组定义、聚合层级）。$v$ 可以通过这些项目中无需滚动或检索操作即可同时看到的项目数量来衡量。在纯聊天中，$v \approx 1$（当前回复）；使用显示 4 个活动筛选器的状态轨道时，$v = 5$；使用完整仪表板时，$v$ 趋近于 $m$。

一种从过载到错误概率的候选映射呈指数形式：

$$P(\text{error}) \approx 1 - e^{-\lambda O}, \quad \lambda > 0 \tag{3}$$

这种函数形式的提出，借鉴了噪声环境下信号检测的类比：随着负荷增加，性能会呈指数级下降。但这应被视为一种假设，而非推导得出的结果。其他候选形式包括逻辑函数或扩散至边界模型；$\lambda$ 的适当形式和参数化方式应通过经验研究确定。

聊天界面迫使 $v \to 1$ （采用单一滚动流），而分析会使 $m$上升，因此 $O$会随之可预见地增加。混合界面通过将状态和制品外置来提高 $v$ ，从而缩小 $O$。

### 2.1 序列化惩罚

信息本质上具有维度；文本本质上是串行的。要通过聊天传达数字矩阵、关系网络或空间分布，就必须将其序列化 。这种压缩会造成信息损失。

作为初步近似，定义序列化惩罚 $S(d)$，用于衡量将$d$维结构展平为一维流的代价：

$$S(d) = \alpha(d-1), \quad d \geq 1,\ \alpha > 0 \qquad (4)$$

这种线性形式是一种简化；实际代价可能取决于结构（稀疏还是稠密）、冗余（可压缩的模式）以及先验知识（熟悉的模式所需的描述较少）。信息论论证表明，$S(d)$的变化可能与结构的熵相关，而非原始维度。线性近似所体现的定性洞见是：随着结构复杂度增加，序列化代价也会增加。

在图形界面中，可以直接呈现 $d$维数据： $S(1) = 0$。在聊天中，即使是二维表格也需要逐行列举： $S(2) = \alpha$。对于关系结构或多面板比较， $S$会随着复杂度增加而扩大。

扩展公式1：

$$O' = O + S(d) = \max(0, m - v - W) + \alpha(d-1) \quad (5)$$

这解释了为什么随着分析任务的结构日益复杂，聊天会让人感到越来越疲惫。这种代价不仅与信息的数量有关，也与结构的维度有关。

## 3 空间认知的神经科学

### 3.1 位置细胞与认知地图

1971年，John O’Keefe 和 Jonathan Dostro-vsky 发现了一种现象，这一发现最终为奥基夫赢得了2014年诺贝尔生理学或医学奖。他们记录了自由活动的大鼠海马体中的神经元活动，发现有些细胞

表1：对于包含 $m = 8$个相关项目的典型分析任务，不同界面类型的过载比较。数值仅供说明；仍需进行实证验证。

| 界面类型 | $v$ | $L_{\text{int}}$ | $O$ |
|---|---|---|---|
| 纯聊天 | 1 | 7 | **3** |
| 聊天 + 状态轨道 | 4 | 4 | 0 |
| 完整混合式仪表板 | 7 | 1 | 0 |

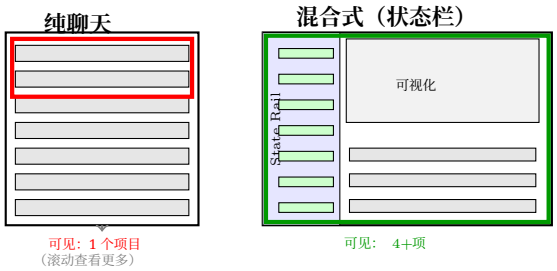


图 1：钥匙孔效应示意图。**左侧：**纯聊天界面一次只显示一项；用户必须滚动浏览并记住内容。**右侧：**带有状态轨道的混合式界面将筛选状态外化，提高了 $v$ 并消除了认知过载。

只有当动物处于其环境中的特定位置时才会放电。他们将这些细胞称为“位置细胞”（O’Keefe andDostrovsky，1971）。
其影响不止于导航。海马体并非仅仅参与记忆，它还在构建环境的内部空间地图。

O’Keefe 和 Lynn Nadel 在 1978 年出版的著作中对此进行了系统阐述：《海马体：认知地图》(O’Keefe 和 Nadel，1978）。他们提出，大脑天生具备将经验置于三维欧几里得框架中的能力。这并非后天习得，而是大脑的内在结构。海马体提供了他们所称的“绝对的、非自我中心的空间框架”，使地标和目标能够根据其异我中心（以世界为中心）的位置进行编码。

位置细胞具有令人惊叹的特性。在恒定环境中，它们的放电模式数月保持稳定。一项研究发现，位置场在153天内始终没有变化（Thompson 和 Best，1990）。大脑维持着持久的空间表征，使其能够在漫长的时间内被可靠地访问。

后续发现进一步拓展了这一图景。莫泽夫妇（与奥基夫共同获得2014年诺贝尔奖）在内嗅皮层中发现了网格细胞，它们以规律的六边形模式放电，提供了一种度量坐标系统（诺贝尔奖委员会，2014年）。头部方向细胞编码朝向。边界细胞在环境边界附近放电。这些细胞类型共同构成了神经GPS。

但认知地图的功能不止于导航。它还为情景记忆、规划和想象提供基础。当你回忆起钥匙放在哪里、规划上班路线，或想象重新摆放家具时，你都在调动海马空间回路。

### 3.2 对界面设计的启示

专门负责空间表征的神经机制的存在，对界面设计具有启示意义，不过应将这种联系视为相互印证的证据，而非直接的因果机制。位置细胞和网格细胞研究表明，大脑具有专门编码空间关系的系统，这些系统有别于处理序列性言语信息的系统。

这一假设关乎架构：当你看到仪表板左侧有一张图表、右侧有一张表格时，你可能正在调动支持持久、基于位置的检索的空间编码系统。你知道事物在哪里，也能回到它们所在的位置，并根据空间邻近性推理事物之间的关系。

阅读对话记录时，你会调用不同的系统。工作记忆中的语音回路负责顺序性的语言处理（Baddeley and Hitch，1974年）。它是串行的，衰减很快，且容量限制非常严格。

二维屏幕布局是否会直接调动海马位置细胞（而非在认知层面激活类似的空间加工），仍是一个有待实证研究的问题。这里的论点属于认知架构层面：人类拥有彼此独立的空间加工系统和语言加工系统，界面应将任务分配给合适的系统。聊天界面迫使空间问题进入顺序通道。这是任务需求与认知架构之间的不匹配，与具体的神经实现无关。

### 3.3 空间记忆具有层级结构

空间记忆研究揭示了一种层级结构，界面既可以利用这种结构，也可能破坏它。人们首先记住空间的整体布局，然后借助这一布局定位特定的目标位置。

Cockburn（2004年）证明，用户会对界面元素的位置形成空间记忆，并以此快速访问信息，且无需回忆即可识别。当界面元素的位置保持稳定时，用户会像记住物理环境一样，在数字工作空间中构建认知地图。

聊天会破坏这种层级结构。每条新消息都会改变所有先前内容的空间位置，因此没有稳定的布局可供学习。用户无法形成“那张图表在左上角”这样的直觉，因为下一条消息一出现，图表的位置就变了。

### 3.4 记忆宫殿作为界面隐喻

位置法（“记忆宫殿”）将事物编码在想象中的空间位置里，并通过在脑海中沿路线行走来提取记忆（Bower，1970年；Yates，1966年）。这并非派对把戏，而是利用空间支架来实现持久回忆。

二维分析画布是数据工作中共享的外部记忆宫殿：假设各有其位；比较化为房间；数据溯源体现在邻接关系中。聊天则是一条失忆的走廊：你可以向前走，但地标会从身后滑走。

## 4 工作记忆：聊天忽视的瓶颈

### 4.1 米勒组块与容量限制

1956年，乔治·米勒发表了《神奇的数字七，再加或减二》，这是心理学领域被引用次数最多的论文之一（Miller，1956年）。米勒指出，人类的即时记忆大约能容纳七个组块的信息。

“组块”这一概念至关重要。棋艺大师看着棋盘时，不会看到 32 个独立棋子，而是看到一个个熟悉的布局，每个布局都作为一个组块发挥作用。专业知识在一定程度上

就是不断形成越来越复杂的组块，将更多信息装入容量固定的存储空间。

后续研究下调了米勒的估计。Cowan (2001) 提出，在控制复述和长期记忆支持后，真实的上限更接近四个组块。关键的一点没有争议：工作记忆存在严格的容量限制。

聊天界面没有针对这些限制做出任何调整。一次复杂的分析会话可能需要追踪多个筛选器、若干可视化内容、比较条件以及中间发现。这些内容都在争夺有限的工作记忆容量。

### 4.2 组块化与外部世界

米勒关于组块化的洞见还带来一个推论，界面设计师早已明白这一点：环境可以替你完成组块化。当信息以空间方式组织在持久显示界面上时，用户无需将所有内容都保存在工作记忆中。他们可以把存储任务交给界面。

包含六张图表的仪表板，并不意味着六个组块在争夺工作记忆。它呈现的是六个空间上彼此区分的区域，用户可以通过眼动按需访问。与维持六个心理表征相比，这种方式的认知负荷要低得多。

聊天不提供组块化支持。之前的消息会滚出屏幕。用户必须记住其中的内容，或者滚回去查找。无论哪种方式，都会消耗原本可以用于分析思考的认知资源。

### 4.3 视觉空间画板

Baddeley and Hitch 的工作记忆模型将视觉空间画板与语音回路区分开来（Baddeley and Hitch，1974年）。画板维持视觉和空间信息。语音回路维持言语和听觉信息。这是两个容量各自独立的系统。

聊天界面主要调动语音回路：阅读文本、编写查询、解析回复。即使任务涉及空间信息，visuospatial sketchpad 也仍未得到充分利用。这浪费了认知资源。视觉空间系统拥有独立于言语工作记忆的容量。优秀的界面会同时调动这两个系统。用户可以一边查看可视化内容（视觉空间），一边思考如何用语言提问。这就是双通道处理，比强迫所有信息都通过单一通道处理更高效。

Logie （1986）表明，未受注意的视觉材料能够优先进入空间加工机制。视觉表征会自动且迅速地编码，而言语描述则不然。当你看到一张图表时，会立即感知到模式、异常值和趋势。用文本描述这些相同的模式，则需要序列加工和有意识的努力。

### 4.4 系列位置效应

自由回忆呈现出稳定的系列位置效应：与中间位置的项目相比，人们对开头和结尾附近项目的记忆更好（Murdock，1962）。长时间的聊天会话实际上就是一串按序排列的认知制品，容易导致：

- **首因偏差**：早期框架会锚定后续探索
- **近因偏差**：最新的图表或解释占据主导
- **中间遗忘**：中间环节的注意事项会被遗忘

结果是，连贯性由叙事脉络驱动，而非谨慎的比较。

## 5 双重编码：为什么图片胜过文字

### 5.1 Paivio理论

Allan Paivio 的双重编码理论形成于20世纪60年代和20世纪70年代，该理论提出，心智通过两类截然不同的心理表征运作：言语代码（词语生成器）和意象代码（意象生成器）（Paivio，1971，1986）。这两类表征由不同的系统处理，这些系统既可独立运作，也能相互作用。

关键发现：同时调动两个系统的具体信息，比仅调动言语系统的抽象信息记得更好。这就是为什么像“卡车”这样的名词比像“真理”这样的名词更容易记住。这个

具体词会同时激活言语表征和心理意象。

Paivio 证明了图片优越效应：当人们学习同时包含图片和文字的列表时，回忆出的图片数量是文字的两倍。图像会自动进行双重编码。文字是否会唤起意象，则取决于其具体程度。

### 5.2 言语遮蔽：语言如何削弱视觉

Jonathan Schooler 关于言语遮蔽（Schooler 和 Engstler-Schooler，1990）的研究揭示了一个更令人担忧的发现：参与者看过一张面孔后，若用语言描述它，之后辨认这张面孔的准确率就会更低，低于完成无关任务的参与者。将视觉体验转化为语言这一行为，会主动损害视觉记忆。

这一发现已得到广泛重复验证。一项涉及多个独立实验室的注册重复研究报告证实了这一核心现象（Alogna 等人，2014年）。不过，边界条件很重要：当言语描述紧接着编码发生、测试刺激与目标刺激相似（需要进行精细辨别），以及视觉信息是整体性的而非特征性的时，这种效应最强（Meissner 和 Brigham，2001）。对于高度独特的刺激，或者参与者事先获知可能存在干扰时，这种效应可能会减弱。

尽管存在这些边界条件，界面设计的核心含义依然成立：迫使用户用言语描述视觉模式（聊天界面所要求的方式），可能会损害这些视觉表征。言语编码是一种有损压缩。当你描述一张图表“呈上升趋势，并在五月左右出现峰值”时，你是在将精确的视觉信号转换为不精确的语义近似。

聊天界面持续要求进行这种转换。“给我看右上角的异常值”需要从视觉感知转码为言语描述。这种转码在认知上代价高昂（为完成转换而消耗工作记忆），并且在表征上有损（言语标签缺乏视觉编码的精确性）。言语描述可能会干扰随后对同一模式的视觉识别。

其含义是：聊天界面并非仅仅隐藏数据；它们可能通过强迫将非言语结构言语化，从而主动削弱用户的心智模型。

### 5.3 对数据界面的启示

数据分析关注的是模式、关系和结构。这些都是受益于视觉表征的空间概念。展示客户细分群体的散点图，与对这些相同细分群体的言语描述相比，会以不同方式被处理。视觉版本能够调动意象系统。它的编码更加持久，也更容易被提取。

聊天界面默认输出言语内容。即使其中包含图像，这些图像也嵌入在文本流中，用户必须滚动才能访问。因此，基于聊天的分析所获得的编码弱于通过视觉界面执行的等效分析。其结果不那么令人难忘，模式也不那么显著。

### 5.4 顺序处理与并行处理

言语表征是按顺序处理的。阅读句子时，你会逐字阅读。你无法一眼看完整个段落。

视觉表征是并行处理的。你看到一张图表后，会立即感知其整体形态、分布和异常值。细节会在集中注意时逐渐显现，但整体形态能够瞬间被把握。

分析洞见往往来自对多个数据点之间关系的感知，而这正是并行视觉处理所捕捉的内容。当你强行通过顺序性的言语通道进行分析时，就会把整体形态割裂成若干部分，必须在工作记忆中费力地重新组合起来。

## 6 分布式认知

### 6.1 野外认知

Edwin Hutchins 任职于加州大学圣地亚哥分校，他通过对海军舰艇上导航团队开展民族志研究，提出了分布式认知理论（Hutchins，1995）。他的核心洞见是：认知过程并非局限于个体的大脑，而是分布于人、工具和环境之中。

在 Hutchins 的著名例子中，一艘船的导航

系统包括船长、领航员、各类船员、图表、仪器和操作流程。没有任何一个人掌握安全导航所需的完整表征。知识分布于整个系统之中。执行导航的“心智”是整个社会技术组合体。

### 6.2 认知卸载

分布式认知的一个关键概念是认知卸载：利用外部资源来减轻内部记忆和处理负担。写购物清单时，你卸载了记忆负担；使用计算器时，你卸载了算术运算；把文件摆放在桌面上以表示项目状态时，你卸载了规划任务。

设计精良的界面能够支持认知卸载。仪表板将数据关系外化。状态指示器将筛选设置外化。空间布局将分析结构外化。每一种外化方式都能减轻用户的认知负担。

聊天界面不利于认知卸载。对话并不是稳定的外部表征，而是转瞬即逝的信息流。若要借助它卸载信息，你就得滚动回看并进行搜索。检索成本削弱了外化带来的益处。

Kirsh 和 Maglio （1994年）证明，俄罗斯方块高手会通过实际旋转下落的方块，减少想象旋转过程所需的心理努力。他们将空间推理卸载到界面上。聊天界面没有为分析任务提供类似的操作支持。没有可供操控或摆放的对象，也没有任何东西可以用来外化分析状态。

### 6.3 认识性行动：用双手思考

Kirsh 和 Maglio 关于俄罗斯方块的研究揭示了实用性行动 与认知行动之间更深层的区别。实用性行动直接让你朝目标更进一步：输入查询、点击提交。认知行动则改变外部世界，让思考变得更轻松：拨弄拼字游戏的字母牌，触发词语识别；转动地图，使其与朝向一致；把纸张摊在桌面上，以看清各种关系。

认知行动是人类将认知延伸到环境中的方式。它们将内部计算转化为外部操作。它们并非思维的偶然附属，而是思维的构成部分。

聊天会阻碍认知行动。你无法拨弄人工智能的回复。你无法在空间上将相关发现归为一组。你无法把备选方案叠放起来进行比较。你无法转动可视化内容，从不同角度观察它。媒介抗拒操控。你被迫对着 界面思考，而不是借助 界面思考。

因此，钥匙孔效应不仅仅是一个可见性问题，更是一个能动性 问题。聊天界面阻止用户用双手思考。

### 6.4 界面作为认知伙伴

Hutchins 认为，我们应当分析认知系统，而非个体心智。飞行员与其驾驶舱构成一个认知系统。系统的边界由信息的流动决定，而非由皮肤与颅骨决定（Hollan 等人，2000）。

这一视角重新框定了界面设计。问题不在于“我们如何向用户呈现信息？”，而在于“我们如何构建一个包含用户的认知系统？”

聊天界面是分析任务中糟糕的认知伙伴。它们不维持状态。它们不表征空间关系。它们不支持并行比较。它们把用户当作孤立的处理器，必须把一切都记在脑子里。

## 7 交互速度与语言摩擦

### 7.1 菲茨定律与指向速度

菲茨定律对目标定位任务中的人体运动进行建模（Fitts，1954）：

$$MT = a + b \log_2 \left( \frac{D}{W} + 1 \right) \tag{6}$$

其中 $MT$ 表示移动时间， $D$ 表示距离，$W$ 表示目标宽度。Card 等人（1978）将这一模型应用于计算机输入设备。鼠标的吞吐量为每秒 4 到 10 比特。目标获取需要数百毫秒。点击标签清晰的按钮很快。

表2：不同模态下的估计交互时间比较（基于菲茨定律和打字速度模型；仍需实证验证）。

| 操作 | 聊天（秒） | 图形用户界面（秒） |
|---|---|---|
| 简单筛选 | 5.2 | 0.5 |
| 日期范围选择 | 8.5 | 1.2 |
| 多选下拉菜单 | 12.3 | 2.1 |
| 下钻导航 | 7.8 | 0.8 |
| 比较视图 | 15.2 | 1.5 |
| **50 次操作** | **~8 分钟** | **~1 分钟** |

### 7.2 输入瓶颈

熟练打字员每分钟可输入 40-80 个单词。但拟定一条分析命令，不只是单纯地打字。先组织查询内容，确定措辞，再输入，最后修改以使表达清晰。像“筛选企业用户”这样简单的细化操作，需要 5-10 秒。

点击标有“企业”的复选框需要 500 毫秒。

这相差一个数量级。在一次分析会话中进行 50 次细化操作，累积的时间成本就会相当可观。更糟的是，每次输入命令都会打断分析流程。你会从思考数据切换到斟酌措辞。

### 7.3 歧义问题

自然语言天生具有歧义。“右上象限里的那个异常值是什么？”依赖空间指示（这个、那个、这里）。面对面交流时，你可以指给对方看；在图形用户界面中，你可以点击。

聊天缺乏指示能力。你必须把指点转化为描述：“留存率约为 87%、收入为 45 万美元，且看起来与主簇分离的数据点。”这样做既缓慢又容易出错，而且对方的理解可能与你的本意不同。

## 8 认知负荷与心理偏差

### 8.1 斯威勒框架

认知负荷理论区分了内在负荷（任务固有的负荷）、外在负荷（糟糕的界面设计施加的负荷）和促进性负荷（用于图式建构和学习的负荷）（Sweller，1988）。目标是尽量减少额外的负荷，以便将认知资源用于内在加工和促进性加工。

聊天界面会给分析任务带来外在负荷：明明点击一下就够了，却要组织语言查询；为了找到先前的结果而不断滚动；在脑中追踪隐藏状态；根据文字描述重建空间关系。这些都无助于推进分析任务，只是额外负担。

### 8.2 实证证据

Nguyen 等人(2022)开展了一项对照实验，比较聊天机器人与基于菜单的界面在执行相同任务时的表现。结果毫无歧义：聊天机器人带来了较低的感知自主性和较高的认知负荷。用户表示，尽管对话显得“自然”，聊天机器人仍需要他们付出更多心理努力。问题根源在于不确定性。开放式对话要求用户在缺乏视觉支架的情况下，组织出精准的提示。

这一发现与理论预测一致。识别比回忆更省认知资源。图形用户界面支持识别，而聊天则要求回忆。

### 8.3 偏差放大

钥匙孔效应也是一个判断问题：界面的形态会改变哪些假设看起来更可信。

**锚定效应**。流畅的早期解释会成为锚点，后续证据必须克服这些锚点（特沃斯基和卡尼曼，1974年）。聊天会让这一过程代价高昂，因为用户必须检索并比较依次出现的片段。

**确认偏误**。假设检验往往会变成保护假设：用户请求进一步阐述，以寻求确认而非证伪（瓦森，1960）。将相互竞争的假设外显呈现的界面，能让证伪变得轻而易举。

**解释深度错觉**。人们经常高估自己对因果系统的理解程度；流畅的解释可能会加深这种错觉，却不必迫使人们将其操作化（罗森布利特和凯尔，2002）。

**变化盲视**。 在基于滚动的文字记录中比较两段相似的文本输出，会干扰注意，并使人容易遗漏差异（Rensink 等人， 1997年； Simons 和 Chabris，1999年）。

## 9 混合式界面的八种模式

解决方案不是放弃对话式人工智能，而是认识到对话应当在交互生态中的何处发挥作用。

### 9.1 生成式用户界面：短暂应用

不要让人工智能以文本或静态图像进行响应，而是让它按需生成交互式软件。查询“分析五月份的客户流失激增情况”后，会生成一个仪表板组件，其中包含图表及相关控件：流失概率滑块、区域下拉菜单和日期范围选择器。你通过对话表达意图，通过直接操控进行细化。

这同时调动了两种编码系统。文字查询触发视觉显示。视觉显示为视空间画板提供稳定的空间结构。用户可以将状态卸载到界面上。认知负荷大幅下降。

最新的实现证明了其可行性。谷歌的生成式用户界面系统能够根据任意提示生成完整的交互体验，人工评估者更偏好这些输出，而不是标准 Markdown（Leviathan 和 Valevski，2025年）。Vercel 的 AI SDK 为基于 React 的生成式界面提供基础设施，LangChain 则通过其生成式用户界面框架提供类似能力。这些系统表明，该模式不仅停留在理论层面，而且能够大规模实现。

### 9.2 无限画布：用于分析的空间记忆

用二维工作区取代线性聊天信息流。你提出一个问题，一张带有可视化内容的卡片便会出现。你从这张卡片向空白处拖出一支箭头，并输入“按产品细分”。随后，一张与之相连的新卡片便会出现。空间布局编码了分析关系，而这些关系在顺序式对话中会丢失。

这利用了空间记忆的层级结构。整体布局提供上下文。在——

单张 Card 可以在该布局中定位。用户构建起对其分析的认知地图。

### 9.3 直指交互：指向提问

将直接操控的精准性与自然语言的表达力结合起来。按住某个键以进入指向模式。在散点图上圈出一个数据簇，然后问：“这些客户有什么共同点？”系统会同时收到所选内容（具体的数据点 ID）和查询（分析意图）。无需冗长描述空间位置。

这就是人类的沟通方式。我们会一边指，一边说。

### 9.4 状态栏：持久上下文显示

隐藏状态会损害分析准确率。十条消息前，你筛选了欧洲。十五条消息前，你排除了一个产品类别。这两个筛选条件仍然生效，却不可见。

状态轨道会在持久侧边栏中显示当前分析状态。每个筛选器、分组、日期范围和排除项都会以可见标签的形式呈现。随着对话进行，状态轨道会随之更新。点击任意标签上的 X，即可直接移除相应限制。这将状态外化，并实现认知卸载。这会提高 $v$ 方程1中的值。

### 9.5 幽灵图层：主动引导

用户会陷入空白页瘫痪。他们看到一张图表，却缺乏直觉，不知道该提出什么问题。

幽灵图层颠覆了查询—响应模式。人工智能会分析每个可视化内容，并在具有统计学意义的特征上叠加细微标记。收入异常下降时，会出现一个小小的指示标记。将鼠标悬停即可查看上下文。点击即可自动下钻分析。这减轻了回忆负担。在分析策略中应用识别优于回忆的原则。

**9.6 备料式布局：准备就绪的工作区**聊天会形成一种“取数任务”模式：用户提出请求，人工智能检索数据，用户再次提出请求。每个周期都是一次穿过钥孔的往返交互。

备料式布局 协议将这一点反转。你陈述自己的分析意图：“我想分析客户流失。”人工智能不会以

答案作为回应。相反，它生成一个工作区：四项最相关的可视化图表已按空间排布，三项最可能的筛选器已预先配置，异常也已标出。副厨把食材准备在台面上；分析师负责烹饪。

这将交互从顺序检索转变为并行探索。用户浏览准备就绪的工作区，发现意料之外的模式，并开始直接操作。通过对话表达意图；通过空间交互进行细化。

**9.7 语义缩放：分辨率即交互**钥匙孔效应部分源于分辨率问题。在任何时刻，你看到的要么是摘要，要么是细节，无法同时看到两者之间的关联。

语义缩放重新定义了滚动。鼠标滚轮不会移动视口，而是改变抽象层级。向上滚动：图表收拢为一句摘要。滚动到中间：这句话展开为一个仪表板。向下滚动：仪表板中的柱形图展开为底层数据行。

每个缩放层级都是完整且连贯的呈现。用户浏览的是抽象层级，而非消息堆栈。上下文不会丢失，因为只需一个手势，较高层级的内容便始终能在认知上触及。

**9.8 概率式用户界面：让不确定性清晰可见**

传统图形用户界面暗示着确定性。切换开关非开即关。筛选条件要么已应用，要么未应用。当系统确切知道用户想要什么时，这种方式行之有效。
大语言模型具有概率性。它们推断意图。它们猜测你想要哪个筛选条件。传统界面的确定性视觉语言无法表达这种不确定性。

概率式用户界面元素通过视觉方式呈现置信度。带有发光边框的筛选器标签表示“这是我的推断，但需要验证。”半透明面板表示“这种解释并不确定。”幽灵副本则展示系统考虑过的替代解释。

这弥合了确定性软件与概率式人工智能之间的鸿沟。用户看到的不只是系统做出的决定，还有系统对此有多大把握。

表 3：混合模式的认知益处。

| **模式** | **认知益处** |
|---|---|
| 生成式用户界面 | 双重编码；　　认知卸载 |
| 无限画布 | 空间记忆；层级回忆 |
| 指示性交互 | 减少 $m$ （无需进行语言重构） |
| 状态轨道 | 提升 $v$ （状态可见性） |
| 幽灵图层 | 识别优于回忆 |
| 备料式布局 | 并行探索；减少往返操作 |
| 语义缩放 | 分辨率层级；上下文保留 |
| 概率式用户界面 | 不确定性清晰可读；验证-通过识别完成 |

验证变成了识别，而非重建。

一项设计顾虑：如果每个元素都显示置信度等级，呈现不确定性可能会增加认知负荷。有效的实现应对可见性设置阈值（仅在高于或低于有意义的置信度界限时显示不确定性标记），采用渐进式披露（悬停或按需查看细节），并建立一致的视觉语言（例如，用饱和度表示置信度，用边框处理表示验证需求）。目标是在不造成负担的前提下保持可读性：在不确定性对用户决策至关重要时使其可见，在系统确信时则将其隐藏。

## 10 分析界面的帕累托原则

这些模式表明了一条实用启发式。
**聊天处理 20% 的交互：**
• 初始意图指定（“显示客户流失随时间的变化”）

• 复杂的一次性查询（“比较首次使用 feat ureX 后一周内发现它的客户与后来才发现它的客户的留存率”）

• 最终综合（“总结关键发现”）
**图形用户界面处理 80% 的交互：**
•所有细化操作：筛选、排序、分组、下钻

• 全部比较：并排查看备选方案

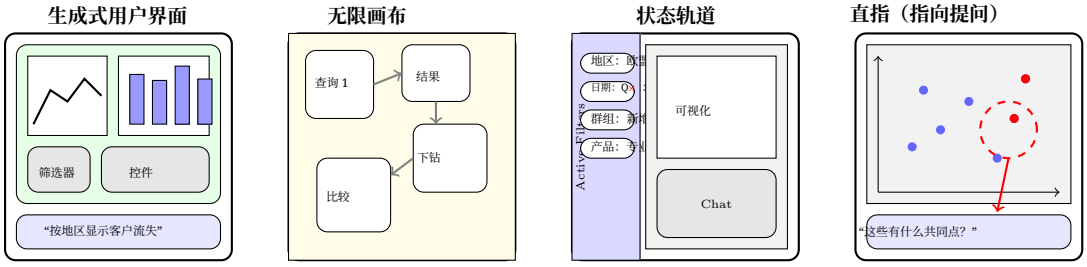


图 2：四种混合式设计模式。**生成式用户界面：** 聊天输入可生成交互式仪表盘。**无限画布：** 二维工作空间保留查询之间的空间关系。**状态轨道：** 持久侧边栏将筛选状态外显。**指示性交互：** 圈选数据点，并针对所选内容提问。

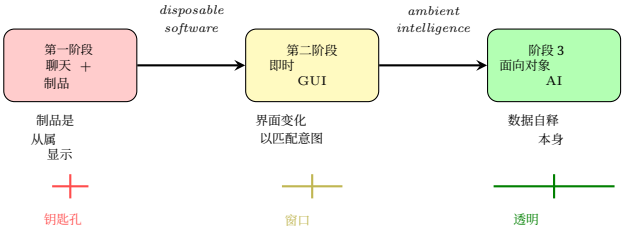


图 3：从与聊天制品交互到环境智能的形态演变轨迹。钥匙孔逐渐变宽，直至消失。

- 状态管理：理解并修改活动筛选器

•迭代探索：沿着一连串相关问题展开

使用公式 1：如果 $O = 0$，纯聊天尚可接受（简单任务、状态少）；如果 $O \geq 1$，则将状态外置（提高 $v$）；如果 $O \gg 1$，纯聊天就会变得容易出错，而且这种情况可以预见。

由此可见：最适合数据分析的聊天界面，是由语音和文本控制的动态仪表板构建器。聊天负责启动，生成的图形用户界面则是执行载体。

## 11 形态演变轨迹：从聊天走向环境智能

上述模式并非终点，而是脱离聊天范式的演变轨迹上的路标。理解这一演变轨迹，有助于预测界面设计的发展方向及其原因。图 3展示了这三个阶段。

### 11.1 第一阶段：制品

当前的实现将聊天和视觉输出并排放置。Claude 在制品中生成内容；用户则在聊天窗口中交谈。这是一种进步：视觉输出具有持续性、可交互性，并且在空间上与对话区分开来。

但制品仍然处于从属地位。要修改它，你必须返回聊天。视觉面板只是一个显示；它还不是一个工作区。当需要细化时，用户仍然要通过钥匙孔进行操作。

### 11.2 第二阶段：一次性软件

下一阶段将制品融入界面本身。“应用程序”变成一个能力数据库，加上一组数据连接。界面 完全基于上下文和意图实时生成。

你提到时间序列数据，软件就变成时间轴工具。你提到地理信息，它就变形成地图。你提到比较，它就生成并排布局。这里没有“地图选项卡”或“图表视图”。系统会为当下的工作构建适当的界面，然后在任务发生变化时将其解散。

这就是即时生成图形用户界面：软件围绕你的需求自行组装，而不是强迫你的需求适应预先设定的形式。聊天窗口缩小为命令行，或完全消失，取而代之的是对所生成界面的直接操控。

### 11.3 第三阶段：面向对象人工智能

最终阶段彻底颠倒了智能与数据之间的关系。人工智能不再是一个与你交谈的独立实体，而是成为数据本身的一种属性。

你打开的不是聊天窗口，而是数据集。你右键点击一列，这一列便开口说道：“我有 40% 的缺失值，主要集中在第二季度的欧盟记录中。”你选中一个点簇，它便解释道：“这些客户的参与度较高，但转化较低；共同因素是仅限移动设备访问。”

智能变得无处不在。每个界面元素都能回答有关自身的问题。“聊天界面”随之消解，因为已经没有什么可供聊天的了。剩下的只有能够自我认知，并能按需自我解释的数据。

这不是科幻小说。这是将语言模型视为通用推理引擎，而非对话伙伴的逻辑终点。钥匙孔消失了，因为你与信息之间不再有一堵墙。房间变得透明。

## 12 研究议程

### 12.1 可证伪假设

**H1（负荷）：**
对于具有隐藏状态的多步骤分析任务，纯聊天界面相比具有持久状态和交互式视图的混合式界面，会产生更高的认知负荷。

**H2（准确率）：**
与带有状态轨道的界面相比，纯聊天分析会产生更多状态一致性错误（在被遗忘的筛选条件下得出结论）。

**H3（偏差）：**
与支持并排比较反假设的界面相比，顺序聊天会加剧锚定效应和确认偏差。

**H4（比较）：**
与基于滚动浏览记录的比较相比，并排可视化能减少遗漏差异（变化盲视）。

### 12.2 实验范式

**被试内界面对比**。条件：纯聊天与混合式（聊天 + 状态轨道 + 交互式视图）。测量指标：任务时间、准确率、NASA-TLX、回退次数、遗忘筛选条件错误。分析：以参与者为随机效应的混合效应模型。

**锚定操纵**。提供一个初始的合理解释；测量在不同界面条件下面对矛盾证据时的坚持程度。

**指示效率研究**。比较“描述该点”与“指向提问”在离群值识别和子组特征刻画上的速度与准确率。

## 13 讨论

### 13.1 与先前工作的关系

本文提出的框架扩展并综合了若干研究传统。Woods 最初的钥匙孔效应（1984）针对的是复杂显示系统中的导航问题；我已论证，大语言模型聊天界面构成了一种极端的、在某些方面甚至是倒退的情形。Woods 的操作者至少还能在不同视图之间导航，而聊天用户则面临持续的内容位移，且没有稳定的空间结构可供导航。

### 13.2 混合式与基于分解的系统

近期研究展示了从纯聊天到完整图形用户界面之间的一系列方法，其中有几种部分解决了本文提出的问题。Phasewise 和 Stepwise 分解式用户界面通过渐进式披露，在以聊天为中心的容器中外化假设、计划和代码，在无需完整生成式图形用户界面的情况下，实现了状态外化带来的一些益处。InsightLens 表明，自动提取并以多视图方式组织洞见（通过 Insight Gallery、Minimap、Topic Canvas），可以通过辅助窗格呈现 $v$，减轻聊天记录中的钥匙孔效应。Talk2Data 展示了在叙述与可执行代码之间进行路由，并以符合本文所提出的“聊天用于表达意图，图形用户界面用于细化”这一理念的方式返回图和表格。

ProactiveVA 通过主动、感知上下文的建议和界面注释，与幽灵图层和备料式布局模式相呼应。DASHBOARD QA 突显了多步图形界面推理和状态遍历的难度，进一步说明了明确设置状态轨道和计划跟踪机制的必要性。

这些系统表明，纯聊天与混合式之间的区分更宜被理解为一个连续谱。在聊天容器中将可编辑假设外化的分解式用户界面，代表了可能降低 $O$的中间设计

而无需在空间意义上完全“抬高 $v$”。

### 13.3 引导式对话何时有效：StatZ 的反例

StatZ 报告称，与成熟的图形用户界面统计工具相比，引导式对话代理在准确率和耗时方面优势显著，尤其对新手而言。这一发现值得仔细考量，因为它似乎挑战了本文的核心论点。

要化解这一矛盾，关键在于认识到 StatZ 的对话代理提供了什么：编码在系统中的专家先验、防止常见错误的防护机制，以及缩小假设空间的针对性建议。用本文的术语来说，StatZ 提升了有效的 $v$，并非通过空间持久性，而是通过对话式脚手架将专家知识外化到对话结构中。用户无需在工作记忆中维持统计决策树，因为代理会呈现相关选项。

这表明存在重要的边界条件：(1) 对于可以通过对话方式编码并呈现专家先验的任务，引导式聊天可能比要求用户自行发现功能的复杂图形用户界面更能减轻负担；(2) 面对陌生工具的新手可能会受益于对专家而言反而显得束缚的对话式引导；(3) 本文提出的批评最适用于开放式多步骤分析探索，在这类探索中，用户必须同时维持并比较多个假设、状态和制品。

此处提出的模式并非反对对话，而是支持混合式：对话式脚手架（如 StatZ）与视觉状态外化（如 State Rails）可以结合使用。80/20 启发式应被理解为依任务而定的分配，而非僵化的规定。

### 13.4 框架的操作化

此处提出的设计模式将以人为中心的人工智能（Shnei-derman，2022）的原则付诸实践：状态轨道通过外显隐藏状态来提升透明性；指示性交互和生成式用户界面恢复直接操控；帕累托启发式则承认不同任务需要不同程度的自动化。近期关于混合界面的研究（DirectGPT 基于指点的交互（Masson 等人，2024）、谷歌的生成式用户界面相较于 Markdown 获得用户的强烈偏好（Leviathan 和 Valevski，2025年）），为这些模式提供了初步的实证支持。

意义建构研究文献（Pirolli 与 Card，2005）将信息觅食（信息收集）与综合（结构创建）区分开来。聊天擅长综合（总结、解释、生成文字），但在信息觅食方面表现不佳，而信息觅食需要快速筛选、比较和假设检验。混合模式在保留聊天综合优势的同时，恢复了信息觅食能力。

**数据溯源与可复现性**。生成式用户界面和无限画布面临的一个问题是，短暂的、动态生成的界面可能缺乏传统仪表板的可审计性。未来的实现应维护数据溯源日志，将生成的用户界面元素与产生它们的查询和数据状态关联起来，从而支持分析人员之间的回放、共享和验证。

### 13.5 对实践的启示

对于构建由大语言模型驱动的分析工具的实践者而言，该框架提出了具体的优先事项：以可见方式外化状态（状态轨道），同时支持指向和键入（直指交互），生成交互式制品而非静态文本（生成式用户界面），并为分析线程提供空间持久性（无限画布）。80/20 启发式（使用聊天来表达意图和进行综合，使用图形用户界面进行细化和比较）提供了一项实用的分配原则，但应根据任务复杂度和用户专业知识进行调整。

对于研究人员而言，过载函数（方程 1）提供了可检验的预测：认知负荷应随隐藏状态（$m-v$）增加；当 $O > 0$时，错误应增多；偏差放大应与序列化惩罚相关$S(d)$。第 12 节中的假设旨在通过受控实验进行证伪。实验设计应纳入受分解式用户界面启发的条件，以检验聊天容器中结构化、可编辑的假设是否能够在没有空间外化的情况下减少 $O$。

### 13.6 拟物化时刻

我们可能正处于一个转折点，类似于从命令行界面向图形界面过渡的时期

（20世纪80年代）。聊天是人工智能的命令行：在表达和规范需求方面灵活自如，却与空间视觉任务格格不入。本文所述的模式，代表了人们对适用于人工智能辅助分析的原生界面模态所做的早期探索。历史经验表明，未来的主导范式不会是经过渐进式改进的聊天，而会是一种截然不同的交互模式。

## 14 局限性

本文主要是理论性的，综合现有《认知科学》研究，以推导关于界面设计的预测。应当承认其中存在若干局限性。

**模型简化**。 基础过载模型（方程 1）将项目视为同质，将可见性视为二元状态，并将工作记忆容量视为固定值。现实系统需要对项目加权（筛选器与假设的贡献各不相同）、采用分级可见性（固定与最小化的元素），并考虑动态因素 $W$，这些因素会受到专业知识、组块化和外部辅助工具的影响。扩展模型（方程 2）勾勒了这些扩展，但仍未经检验。指数误差映射和线性序列化惩罚是候选函数，而非推导出的结果；应通过经验研究评估其他函数形式。

**实证验证**。 认知负荷估算和交互时间比较（表2）是根据理论模型（菲茨定律、打字速度研究）推导得出的，并非对聊天界面与混合界面用于分析任务时进行直接实验测量的结果。第 12节中的假设可以证伪，但尚未针对现代带有制品的聊天系统进行检验。

**引导式对话系统**。 如第13节所述，经过充分支架设计、能够编码专家先验的对话系统（如 StatZ）可以降低认知负荷，相较于复杂图形用户界面尤其能帮助新手。本文提出的批评最适用于开放式、多步骤的分析探索；引导式对话何时足以满足需求、何时必须进行空间外化，其边界条件仍有待通过实证厘清。

**中间设计**。 许多现代“聊天”工具已经是混合式工具（固定图表、侧边面板、代码单元）。该框架的预测取决于系统在聊天到混合式的连续谱中所处的位置。在聊天容器中将可编辑假设外化的分解式用户界面，或许无需完整的图形用户界面生成，就能获得状态外化带来的一些好处。

**言语遮蔽效应的适用范围**。 尽管面孔识别中的言语遮蔽效应已得到充分证实，但将其应用于数据可视化仍属推演。该效应在特定条件下最为显著（即时言语描述、相似干扰项、整体性刺激），而这些条件是否适用于典型的分析工作流程尚不确定。预计需要对相似分布进行精细辨别的任务会受到最强影响。

**实现挑战**。 本文提出的混合式模式假定具备某些技术能力（实时用户界面生成、模态切换、概率渲染），但这些能力目前仍面临工程挑战。在可运行的实现得到评估之前，其认知益处仍停留在理论层面。

**个体差异**。 用户在空间能力、言语流畅性和界面偏好方面各不相同。该框架尚未考虑个体差异可能如何调节钥匙孔效应，或影响混合式模式的有效性。

**引文可获取性**。 部分被引文献（例如，Leviathan 和 Valevski （2025年））属于近期行业出版物，可能未收录于传统出版渠道；其作为同行评审研究的证据地位应据此理解。

## 15 结论

不同的认知任务需要不同的界面模态。这不是观点。这是架构。

大脑拥有以海马体为中心的专门空间加工系统。它拥有独立的言语处理系统。其工作记忆容量限制约为四个组块。它受益于同时调动言语系统与意象系统的双重编码。它通过分布式认知和认知卸载延伸至环境之中。言语遮蔽研究表明，迫使视觉信息经由语言通道，不仅无法调动这些系统，还会主动劣化它所触及的

表征。

聊天界面忽略了这一切。它们强迫用户通过语言渠道完成空间任务，无法为 visuospatial sketchpad 提供稳定的空间结构。聊天界面不断变化，阻碍了认知卸载，也不支持双重编码。当识别就足够时，它们却要求用户回忆。

钥匙孔效应由 Woods 于四十年前首次提出，在对话式人工智能时代显得愈发紧迫。通过聊天窗口分析多维数据，迫使用户透过狭窄且不断移动的视口，查看广阔的信息空间。空间语境、并排比较，以及先前发现的对象恒常性都随之消失。每次滚动都会扰乱认知地图。每次输入细化内容都会打断分析流程。

解决方案是采用混合界面，让每种模态各展所长：通过聊天表达复杂、新颖的意图；通过图形界面快速、精准地进行细化。本文介绍的八种模式（生成式用户界面、无限画布、指示性交互、状态轨道、幽灵图层、备料式布局、语义缩放、概率式用户界面）提供了让它们协同工作的具体机制。

发展方向是构建能够根据任务需求动态分配交互方式的界面：通过对话进行需求说明与综合，通过生成式可视化界面进行探索与比较。本文提出的模式和假设为通过经验研究验证这些主张提供了一套研究方案。

**参考文献**

V. K. Alogna、M. K. Attaya、P. Aucoin、S. Bahnik 等。《注册重复研究报告：Schooler 和 Engstler-Schooler（1990）》。《心理科学观点》，9(5):556–578，2014年。Alan D. Baddeley 和 Graham Hitch。工作记忆。载于 Gordon H. Bower（编），《学习与动机心理学》，第8卷，第47–89页。学术出版社，1974年。Kevin B. Bennett 和 John M. Flach。视觉动量再探。《国际期刊》

人机交互研究期刊，70(6):399–414，2012。Gordon H. Bower。《记忆辅助工具分析》。《美国科学家》，58(5):496–510，1970。Stuart K. Card、William K. English 和 Betty J. Burr。评估鼠标、速率控制等距操纵杆、步进键和文本键在阴极射线管上进行文本选择时的表现。《人体工学》，21(8):601–613，1978。Andy Cockburn。重新探讨二维与三维对空间记忆的影响。载于第五届澳大拉西亚用户界面会议论文集，第28卷，第25–31页，2004。Nelson Cowan。短时记忆中的神奇数字4：重新审视心理存储容量。《行为与脑科学》，24(1):87–114，2001。Paul M. Fitts。人类运动系统在控制运动幅度时的信息容量。《实验心理学杂志》，47(6):381–391，1954。James Hollan、Edwin Hutchins 和 David Kirsh。分布式认知：为人机交互研究奠定新的基础。《*ACM*计算机与人类交互汇刊》，7(2):174–196，2000。Edwin Hutchins。《野外认知》。麻省理工学院出版社，1995。David Kirsh 和 Paul Maglio。论如何区分认知性行动与实用性行动。《认知科学》，18(4):513–549，1994。Yaniv Leviathan 和 Dani Valevski。生成式用户界面：大语言模型是有效的用户界面生成器。技术报告，谷歌研究院，2025年。参见 generativeui.github.io。Robert H. Logie。工作记忆中的视空间加工。《实验心理学季刊》，38(2):229–247，1986。Damien Masson、Sylvain Malacria、G´ery Casiez 和 Daniel Vogel。DirectGPT：一种用于与大语言模型交互的直接操作界面

模型。载于*2024*年计算机系统中的人为因素 *CHI* 会议论文集，美国纽约州纽约市，2024年。ACM。

Christian A. Meissner 和 John C. Brigham。人脸识别中言语遮蔽效应的荟萃分析。《应用认知心理学》，15(6):603–616，2001年。

George A. Miller。神奇的数字七，加减二：我们处理信息能力的某些局限。《心理学评论》，63(2):81–97，1956年。

Bennet B. Murdock。自由回忆的系列位置效应。《实验心理学杂志》，64(5):482–488，1962年。

Quynh N. Nguyen、Anna Sidorova 和 Russell Torres。用户与聊天机器人界面和基于菜单的界面的交互：一项实证研究。《人类行为中的计算机》，128：107093，2022年。

诺贝尔奖委员会。2014年诺贝尔生理学或医学奖，2014年。John O’Keefe、May-Britt Moser、Edvard I. Moser。

John O’Keefe 和 Jonathan Dostrovsky。《海马体作为空间地图：来自自由活动大鼠单元活动的初步证据》。《脑研究》，34(1):171–175，1971。

John O’Keefe 和 Lynn Nadel。《海马体：认知地图》。牛津大学出版社，1978。

Allan Paivio。《意象与言语过程》。霍尔特、莱因哈特与温斯顿出版社，1971。

Allan Paivio。《心理表征：双重编码方法》。牛津大学出版社，1986。Peter Pirolli 和 Stuart Card。通过认知任务分析确定的意义建构过程及分析师技术的杠杆点。《国际情报分析会议论文集》，5:2–4，2005。

Ronald A. Rensink、J. Kevin O’Regan与James J. Clark。看见还是看不见：感知场景变化对注意的需要。《心理科学》，8(5):368–373，1997年。

Leonid Rozenblit与Frank Keil。民间科学被误解的限度：一种解释深度的错觉。《认知科学》，26(5):521–562，2002年。

Jonathan W. Schooler 和 Tonya Y. Engstler-Schooler。视觉记忆的言语遮蔽：有些事情还是不说为好。《认知心理学》，22(1):36–71，1990。

Ben Shneiderman。直接操控：超越编程语言的一步。《计算机》，16 (8):57–69，1983年。

Ben Shneiderman。《以人为中心的人工智能》。牛津大学出版社，英国牛津，2022年。

Daniel J. Simons 和 Christopher F. Chabris。我们身边的大猩猩：动态事件中的持续非注意盲视。《知觉》，28 (9):1059–1074，1999年。

John Sweller。问题解决过程中的认知负荷：对学习的影响。《认知科学》，12 (2):257–285，1988年。

Lonnie T. Thompson 和 Phillip J. Best。自由活动的大鼠背侧海马中记录的单个神经元位置场活动的长期稳定性。《脑研究》，509(2): 299–308，1990。

Amos Tversky 和 Daniel Kahneman。不确定性下的判断：启发式与偏差。《科学》，185(4157):1124–1131，1974年。

Peter C. Wason。关于在概念任务中未能排除假设。实验心理学季刊，12(3)：129–140，1960年。

David D. Woods。视觉动量：改善人与《计算机》之间认知耦合的概念。国际人机研究期刊，21(3)：229–244，1984。

FrancesA. Yates。《记忆的艺术》。芝加哥大学出版社，1966年。