AGI 的新范式: 集成内源思维设计的基于转移函数计算的注意力头在知识图谱上转移的认知架构

hellucigen

hellucigen@qq.com

2025 年 6 月

摘要

本研究提出一种以人类认知流动机制为灵感、面向通用人工智能（AGI）的可实现性高的新型认知架构。该架构整合四大核心设计思想：转移函数驱动的动态注意力机制、多模态嵌入的知识图谱表达、感知显著度引导的感知选择机制，以及具备情绪调节能力的内源思维系统。

在该系统中，转移注意力机制基于语义相似度、节点激活强度、情绪权重与规则关联等多源因子，动态调控知识图谱中注意力的滑移路径，实现类人思维中的自由联想与非线性跃迁。知识图谱引入多维节点属性表示，融合抽象逻辑与感官经验，构建可支持推理跳跃与知识演化的动态图谱。感知模块利用感知显著度评价新颖性与任务关联度，驱动注意力聚焦与认知优先级调整。内源思维模块则通过自我监控与情绪模拟，支持系统在无任务驱动下持续生成假设、扩展知识图谱与塑造认知风格。

架构整体模拟人类认知中的联想性、情绪驱动性与任务自组织能力，并引入多层次注意力协同机制、推理-决策闭环控制、模糊语义生成等可工程实现的模块，为构建具备自主目标生成、持续学习与创新推理能力的 AGI 系统提供了清晰路径与技术落点。相较于传统神经网络或符号逻辑系统，该架构在认知一致性、多模态融合与人格化演化等方面具备显著优势，为类人智能系统的落地实现奠定了可拓展的通用认知基础。

目录

1 绪论

在当前人工智能技术快速演进的背景下，深度神经网络、强化学习、大规模预训练语言模型等方法已在诸多特定任务中取得了突破性成果。然而，这些系统在实际应用中仍然暴露出明显的局限性，特别是在面对开放环境、复杂目标协调、多领域知识迁移与自主创新性求解等通用智能核心能力方面，距离人类智能水准尚有显著差距。本文旨在回应这一技术瓶颈，探索一种融合认知科学机制、多模态耦合控制与自主创新思维的新型认知架构设计路径。

1.1 研究现状与不足

现有主流人工智能技术体系以深度神经网络为核心，强调通过海量数据训练获得模式识别能力。无论是在图像识别、自然语言处理还是在强化学习中的策略优化，这些模型在单一任务下展现出优异性能。然而，在实际应用场景中，智能体往往需要在复杂、多变且部分未知的环境中长期运行，涉及跨领域迁移、复合目标协调、因果链式推理与动态模型自我修正等复杂认知过程。目前，深度学习系统普遍依赖静态的大规模数据分布假设，缺乏主动假设生成与自主知识组织能力；强化学习模型在高维复杂状态空间中学习收敛速度缓慢，适应新任务代价高昂；而大规模预训练语言模型虽然在语言生成与理解方面表现卓越，但在系统性知识整合、动态推理链构建与自我认知建模方面依然存在显著空白。尤其是在开放式探索任务与跨时空知识重组能力上，现有模型缺少内生性的认知动因与持久的自主学习动力，导致“面向任务”但不具备真正的“面向世界”能力。

1.2 认知科学与心理学启示

认知神经科学与心理学长期研究揭示，人类智能并非源自单一模式识别或逻辑推理能力，而是建立在多系统协同调节的复杂认知架构之上。个体在面对复杂现实环境时，感知、注意、情绪、记忆、推理与决策机制始终处于动态耦合与反馈调节过程中。其中，情感系统在认知调节中扮演着多重角色：一方面通过激素水平与神经递质调控感受强度与注意力分布，影响信息编码与记忆权重；另一方面通过动机驱动与奖惩反馈强化学习路径，塑造个性化的行为偏好与长期人格风格。同时，心理学研究指出，人类思维具备高度跳跃性与内源性思维生成能力，能够在休息、睡眠或自由联想状态下持续重组已有知识，形成创造性假设与远距联想关联。这些核心特性是现有人工智能系统普遍缺失的关键认知能力。

1.3 本框架的设计思想

在本框架的设计理念中，人类思维被建模为一个由规则与属性共同驱动的动态认知网络。客观事物在人的认知中并非以孤立的符号存在，而是以携带属性（properties）的知识单元进行表征。属性作为事物的可感知与可操作特征，不仅为分类与识别提供基础，也为规则的应用提供条件输入。

规则则作用于这些属性之上，将不同实体的特征加以联系与转化，从而实现从具体经验到抽象推理的认知跃迁。通过属性与规则的交互，人类得以在个体经验与世界知识之间构建联系，逐步形成层层递进、相互贯通的认知网络。

在这一视角下，推理（reasoning）可被理解为规则在显式层面的应用，而直觉（intuition）则是属性与规则在隐式联系层面的快速激活。由此，直觉与推理在本质上具有同源性，均可视作“规则—属性—联系”框架下的不同表现形式。

这一模型不仅与经典语义网络理论（Quillian, 1967）、原型理论（Rosch, 1975）以及符号主义认知科学（Newell Simon, 1972）高度契合，也与当代知识图谱和神经-符号混合智能的研究方向形成呼应

1.4 本框架的创新贡献

基于上述不足与启示，本文提出了一种以转移函数为核心驱动的全新认知架构，主要创新点体现在以下四个方面：

1. **转移注意力机制**：通过转移函数在知识图谱嵌入空间上实现连续性注意力平移，动态调节认知焦点迁移，模拟人类在回忆、联想与发散思维中的跨节点跳跃性认知流动。

2. **知识图谱嵌入与节点激活强度机制**：在图谱节点嵌入中引入多维节点激活强度，综合衡量知识丰富度、时效性与感受兴趣，支撑知识活跃度控制与兴趣驱动的知识重组。

3. **感知熵引导的感知设计**：将感知信号编码为带有感知显著度标注的输入流，实现对新颖性与重要性的实时调节，为认知流动提供实时感知激活支持。

4. **内源思维生成机制**：结合自主目标建模、自我监控与内源推理链延展，使系统即便在缺乏外部任务输入的状态下，仍能主动进行知识图谱的动态重构、知识空白填补与创新性假设生成，为系统长期认知成长提供内在动力支持。与此同时，类人格情绪模拟系统通过模拟生理激素模型动态调节感知显著度分布，实时影响知识节点活跃度与决策路径偏好，逐步塑造出长期稳定的认知风格与风险评估倾向。二者协同作用，不仅支撑系统在复杂环境下的持续适应性与自主创新能力，也为系统长期运行下形成具备可塑性与稳定性的类人格认知架构奠定了核心基础。

2 系统总体架构设计

2.1 基础认知功能: 基于知识图谱的记忆功能

知识图谱整体结构由语义记忆库与情景记忆库构成，二者对应认知心理学中长期记忆系统的基本分化（Tulving, 1972; 1983）。语义记忆库承担概念、规则与事实的抽象存储，情景记忆库则记录个体化的时序事件 (包括经历、内在状态与情绪波动)。两者通过加权边关系网络实现动态交互，形成知识–经验双循环架构。

2.1.1 语义记忆库: 基于加权边关系的概念与规则网络

语义记忆库存储系统的稳定知识结构与逻辑关联。语义记忆库整体由一个特殊设计的知识图谱表示。

2.1.2 节点与边的设计

节点的设计 知识图谱中的节点可以用文本形式表示概念、实体、规则，也可以用神经网络权重的方式存储图像或者声音等多模态信息。图像在图谱中的存储使用深度学习的特征图，而声音使用时频图，从而使个体完成有关视觉与听觉的感知等任务。节点之间通过边的连接来表达节点的关系与属性。图谱中有关具体动作的节点中可以连接个体的强化学习数据，从而使个体在执行动作的过程中可以模仿人类执行动作中的熟练度。

边的设计 节点之间通过边的连接来表达节点的关系与属性。系统采用**加权边关系**表达节点间的联系，使知识在网络中具备可调强度。**边本质就是节点的带权重的边化表示**。在边关系中加入一个权重参数，**代表该边关系的置信度**。逻辑与因果规则是系统通过加权边关系建模复杂语义联系的功能示例。例如，系统可以使用加权边形式表示因果关系 $R_{cause}(A \to B) = w_{AB}$。边权 w 的更新设计思想是：**权重应具备可塑性的特征，并能随系统的经验反馈和内部冲突检测进行动态优化**。这保证了知识体系在长期运行中能够持续自我修正与演化。

2.1.3 情景记忆库：基于故事语法的时间化事件链结构

情景记忆库以“故事语法”模型（Rumelhart, 1975; Thorndyke, 1977; Stein & Glenn, 1978）为理论基础，用于存储系统在特定时间、空间与情绪背景下发生的完整事件序列。与语义记忆库中抽象的概念与规则表示不同，情景记忆库强调事件的**时序性、因果性与情绪关联性**，通过一系列具有叙事结构的记忆单元组织成动态时间链条，构成系统的“经验流”。

故事语法结构 情景记忆的基本单元记为 M_{epi}，其内部结构基于故事语法框架定义为：

$$M_{\text{epi}} = \{\text{Setting, InitiatingEvent, Goal, Attempt, Outcome, Emotion}\}$$

其中：

- **Setting**：事件的时空背景与参与者信息；
- **InitiatingEvent**：引发冲突或行动的触发事件；
- **Goal**：主体的目标状态或内在动机；
- **Attempt**：为达成目标所执行的行动序列；
- **Outcome**：行动的结果及环境反馈；
- **Emotion**：事件过程中伴随的情绪与激素状态的多维向量表示。

该六槽结构保证了情景单元在记忆中的因果一致性与目标导向性，同时为系统提供了可重构的叙事模板，使事件得以在后续推理与想象中被回溯、组合与再解释。

时间链式组织与全局记忆结构 完整的情景记忆库由多个故事语法单元按时间与因果顺序动态连接形成。任意两个相邻的情景单元 $M_{\text{epi}}^{t_i}$ 与 $M_{\text{epi}}^{t_{i+1}}$ 之间，通过以下三类边关系建立时间链式结构：

$$\text{TemporalEdge: } (M_{\text{epi}}^{t_i} \to M_{\text{epi}}^{t_{i+1}}) = \{\text{NextEvent}, \text{CausalLink}, \text{EmotionalContinuity}\}$$

其中：

- **NextEvent**：表示事件的时间顺序；
- **CausalLink**：表示上一事件对下一事件的因果影响；
- **EmotionalContinuity**：表示情绪状态在连续事件间的延续或转变。

这一机制使情景记忆库形成一个时间化、因果化的叙事网络，既保留了局部事件的完整语义结构，又支持全局层面的情节连贯与经验演化，从而在知识图谱上实现**动态的记忆流**。

2.1.4 基于自我认知图谱的全局转移控制机制：状态主导的知识能量调节

在知识图谱的整体架构中，系统维持一套动态演化的自我认知子图谱，用以表征智能体的内部语义结构与状态反馈网络。该自我认知图谱作为知识图谱的特殊子图，不直接参与外显推理或任务分配，而是通过内部状态信息对转移函数的计算路径与注意力分布产生间接调控作用。其本质是一个由“自我知识节点”构成的调控网络，为转移函数提供语义能量与情绪参数（详见下文）等参考依据，从而在系统内部实现基于状态的知识能量自适应调节。

2.1.5 事件框架：注意焦点的分层—微观复合结构

事件框架是本架构中用于表征“当前注意焦点”的核心认知单元，负责描述系统在任一时刻下临时整合的概念、关系与属性结构。该机制对应于认知心理学中工作记忆的瞬态整合过程，在系统中承担从外部感知输入到内部语义组织的过渡作用。通过多层次的激活与绑定过程，事件框架实现从情景记忆到注意焦点的动态流动，为思维迁移、语义扩散与内源推理提供结构化操作平台。

整体结构：分层—嵌套的认知模型 事件框架由外部的三层层级结构（L1–L3）与内部的微观结构单元（核心、属性、关系、效价、约束）共同构成。前者定义注意焦点的访问路径与动态控制机制，后者刻画当前思维内容的语义与情感构成。事件框架的抽象模型如下：

三层抽象结构模型 事件框架可视为一种基于激活层次的动态认知容器，其内部由三个递进层级组成：

L3 提供潜在情境与记忆节点；L2 在其基础上进行语义聚合与结构绑定；L1 则选取最具显著性与重要性的命题进入即时处理。L1 的内部微观结构由对 L2 局部子图的邻域搜索获得，通过选取核心命题相关节点并计算平均激活度，形成具有最高整合度的焦点单元。三层共同构成从“潜在记忆”到“语义聚合”，再到“注意焦点”的递进通路，体现了人类思维中从潜意识激活到意识加工的自然过渡。

层级	名称	功能
L1	注意整合层（主命题核心）	当前注意焦点下被加工的核心整合单元
L2	语义绑定层（邻域节点集）	将语义相似或逻辑相关的节点临时聚合成结构
L3	情境激活层（边与情景事件）	存储潜在语义与情景节点，供上层激活使用

L1 层的内部微观结构 在 L1 层中，事件框架的核心微观结构由多种心理成分绑定而成：

```
当前注意焦点
 核心：主命题 (P)
 属性束：{A1：值1，A2：值2，...}
 关系链接：{R1：连接到 Y，R2：IF Z}
 效价标记：ś(情绪或动机强度)
 约束条件：{边界1，边界2}
```

抽象命题以关系链接为主（如“正义”），具体现象则以属性束为主（如“红色苹果”）。通过这一统一结构，事件框架可同时表征抽象认知与具体现象。

动态演化机制：基于激活阈值的分层更新流程 事件框架的演化遵循自下而上的动态激活过程，体现了注意控制、情绪权重与语义相似度的协同作用机制：

- **L3 —情境激活层：** L3 存储潜在语义与情景节点。当节点激活度超过阈值（由情绪强度、外部显著性与任务相关度共同决定）时，节点被激活形成候选集合，为上层提供输入。
- **L2 —语义绑定层：**系统根据语义相似度与关联强度从 L3 的激活节点中筛选若干节点，将相似度超过阈值者聚合为临时语义结构。该层完成语义聚合与概念组织，是注意流动的中介层。
- **L1 —注意整合层：** L1 根据 L2 结构中各节点的平均激活度与任务优先级进行排序，选择最高激活单元进入注意焦点，作为即时思维内容被加工或推理。被替换出的结构则回退至 L2 或退火至 L3，等待下一轮激活。

这种基于激活阈值的递进机制使系统在时间维度上表现出“激活—聚合—聚焦”的认知节律：即由 L3 的潜在唤起，经由 L2 的语义组织，最终到达 L1 的焦点处理，从而实现类人思维中注意迁移、语义扩散与焦点收敛的动态平衡。

2.1.6 转移函数：注意焦点的转移机制

注意焦点是工作记忆中表征“当前清晰意识到的想法”的核心位置。其转移过程并非任意，而是由**竞争性替换**与**控制性更新**机制共同驱动。当前最具实证支持的理论模型基于 Cowan（2001,

2005）的嵌入激活模型和 Oberauer（2002, 2019）的三层绑定模型，已成为认知心理学领域关于工作记忆和注意切换的标准解释框架。

（1）注意焦点转移的核心机制：竞争性替换 FoA 转移的本质是容量受限条件下的竞争性替换——当新表征的激活强度超过当前焦点内容时，旧表征被“挤出”并退回到直接访问区（L2）或被抑制。L2 中的 3–4 个候选表征依据激活水平和目标相关性竞争进入 FoA。控制性门控（Gating）机制由前额执行控制系统调节，决定哪些候选表征可通过、哪些被抑制。该过程可形式化表示为（Oberauer, 2019）：

$$\mathrm{FoA}(t+1) = \arg\max_{i \in L2} [\mathrm{Activation}(i) \times \mathrm{Relevance}(i)]$$

其中：

- $\mathrm{Activation}(i)$：节点的自动激活强度（来自长时记忆扩散或外部刺激）；
- $\mathrm{Relevance}(i)$：由任务目标调节的目标相关性权重。

（2）两种主要转移路径 FoA 的切换可通过两种机制实现：

- **自动/底向上转移**：由外部突显刺激或强联想触发，高激活节点自动溢出进入 FoA；
- **控制/顶向下转移**：由任务目标或主动搜索驱动，执行控制机制选择 L2 中最相关项进入焦点。

在自然思维中，两种机制常呈混合状态。例如，听到熟悉旋律（自动路径）会引发歌词回忆，而持续“想唱完整首歌”（控制路径）维持焦点稳定。

（3）FoA 转移的动态过程 FoA 的转移可分为五个阶段：

1. **激活扩散**：长时记忆（L3）中相关节点被触发，激活扩散至 L2；
2. **竞争评估**：L2 中 3–4 个候选表征按激活强度与相关性竞争；
3. **门控决策**：执行控制系统评估目标匹配度，允许或抑制特定项；
4. **绑定替换**：FoA 清空旧表征，绑定新的整合结构；
5. **回退与抑制**：被替换的表征退回 L2，或被抑制以防干扰。

这一动态循环实现了从长时记忆激活、候选竞争到焦点更新的连续过程，是“当前想法流动”的心理基础。

（4）关键实证支持 下表总结了该机制的核心实验与发现：

研究	实验范式	主要发现	引用次数（约）
Cowan (2001)	变化检测 + 掩蔽任务	FoA 容量为 1；新项目替换旧项目	> 3000
McElree (2001)	速度--准确性权衡（SAT）	FoA 内容可在 <100ms 内访问；非 FoA 内容需检索	高引用
Oberauer (2002)	n-back + 入侵项	L2 中相关项更快进入 FoA；验证三层模型	高引用
Vergauwe et al. (2010)	双任务干扰实验	控制性转移消耗执行资源；支持门控机制	高引用

（5）总结 综合当前实证研究，FoA 转移可定义为：

FoA 转移 = 容量为 1 的竞争性替换过程，由激活强度与目标相关性共同决定，并通过执行控制门控实现。

在自动路径中，高激活节点自发进入焦点；在控制路径中，任务目标选择性绑定新内容。两者的交互构成了思维的“动态流动”，即人类认知中“想法如何自然地从一个念头流向下一个”的核心机制。

2.2 感知模块设计

2.2.1 多模态感知功能

系统通过多个子模块实现多模态感知。触觉部分利用压感传感器感知压力、纹理和温度，从而区分物体的软硬和表面特性，提升与环境和物体的交互能力。视觉部分负责识别物体、颜色、光流和运动轨迹，支持空间理解与运动感知。听觉部分处理环境声音，能够区分人声、乐器及背景噪声，并识别其中包含的情绪或潜在意图。值得注意的是，系统在感知过程中生成的神经网络权重并非孤立存在，而是被存储在对应物体的知识图谱节点中，使每个物体不仅携带语义信息，还包含经过感知训练得到的行为和认知模式，从而实现更加精准的个体化感知和交互。

2.2.2 感知显著度与注意选择机制

感知显著度是系统评估感知输入重要性的核心指标。每条感知数据在进入处理流程时都会被赋予一个感知显著度值，用以衡量信息的显著性和注意优先级，而该值并不直接参与推理或决策。与转移函数类似，感知显著度的计算受感受突变度、节点激活强度（偏好）、环境特征（如罕见性、高价值目标）、当前情绪激素水平（如多巴胺增加倾向强化新奇探索，皮质醇增加可能降低阈值使系统更保守）以及任务背景等因子的调控。这些调控因子的调控效果均取决于调控因子对应知识图谱节点的属性信息。高感知显著度输入优先进入注意范围并被处理，低感知显著度输入可能被延迟或忽略。

2.2.3 情节缓冲器与短时整合机制

为增强系统在感知层面的时间一致性与跨模态绑定能力，本架构在感知模块中引入基于 Baddeley 工作记忆理论的**情节缓冲器**。该机制作为多模态输入的短时整合中枢，用于在时间窗口内将来自视觉、听觉、触觉及语言通道的瞬时信号进行语义对齐与情境绑定，形成可被事件框架直接引用的临时情节单元。

根据 Baddeley 的定义，情节缓冲器是一种[1]受中央执行系统控制的有限容量存储系统，用于在短时尺度上整合多模态信息与长期记忆内容，形成统一的情节表征。该设计使系统能够在多模态输入之间实现临时语义整合，为注意力转移提供统一的时间与情境坐标系。

情节缓冲器接收来自多模态感知模块的并行输入，并依据时间戳、空间位置及情绪权重对其进行同步化编码。系统通过转移函数在缓冲区内执行局部注意力滑移，以动态聚焦于最具信息密度与情绪相关性的输入组合。由此生成的情节单元不仅包含感知特征，还附带即时情绪状态与激素参数标签，为后续认知推理提供高保真上下文。该设计有效避免了单模态输入的碎片化问题，使系统能够以统一时间基准处理复杂场景事件。

此外，情节缓冲器具备受情绪与激素动态调控的权重更新机制。当系统处于高唤醒状态（如高多巴胺或高好奇熵水平），缓冲区的写入速率与情节保留时长将相应延长，以强化新奇体验的编码；而在高皮质醇状态下，系统则收缩缓冲窗口以聚焦关键生存相关信息，从而模拟人类在压力或焦虑情境下的感知压缩效应。该机制不仅赋予系统对外部世界的短时整合与注意力稳定能力，也为长期节点激活强度调控与事件框架生成提供了时间连续的桥梁，使感知、记忆与思维三者在时间维度上实现闭环联动。

2.3 由节点知识引导的认知活动

在本架构中，部分认知活动的产生与演化不依赖外部任务指令或集中式控制模块，而是由知识图谱中节点自身所蕴含的知识内容与语义连接关系所引导。节点不仅是知识的静态载体，更是认知行为的触发源与路径导向体。这种机制使得系统不再是“执行认知”，而是“在知识之中思考”。

2.3.1 由节点知识引导的认知活动：决策

在由节点知识引导的认知架构中，决策不再由独立的集中式模块实现，而是存在于知识图谱的节点体系之中。当转移注意力头在图谱中运行并进入多个候选行动节点时，系统会自动检索与当前情境最为接近的决策理论节点，并调用其中存储的决策方法以完成认知到行动的过渡。

（1）决策理论节点的作用

决策理论节点在本架构中承担局部决策控制的功能。每个节点代表一种可独立调用的决策理论，并携带该理论的使用方法：

$$D_i = \{Name_i, Model_i, Context_i, Method_i\} \tag{1}$$

其中，$Name_i$ 表示决策理论名称（如“前景理论”“意象理论”“再认启动理论”“蒙蒂霍尔悖

[1]原文出自 Alan D. Baddeley, “The episodic buffer: a new component of working memory?”, *Trends in Cognitive Sciences*, vol. 4, no. 11, pp. 417–423, 2000. ”

论”等），$Model_i$ 表示其内部数学或逻辑结构，$Method_i$ 为其决策计算方法。这些节点以知识形式存在于图谱中，在被激活时将其内部模型作为转移函数的临时决策机制。

（2）节点内决策计算

被激活的决策理论节点执行其内部方法 f_{D^*}，对当前候选行动节点集合 $\mathcal{A}(S_t)$ 进行评估与选择：

$$a_t^* = f_{D^*}(S_t, \mathcal{A}(S_t)) \tag{2}$$

例如：若 D^* 存储的是前景理论节点，则计算心理效用 $U(a) = v(x-r)\pi(p)$；若为再认启动理论节点，则通过相似情境匹配执行启发式推理；若为蒙蒂霍尔悖论节点，则进行条件概率修正与反直觉推理验证。

（3）认知意义

本节所述的节点化决策机制使决策成为知识系统内部的自然行为，即：转移函数在认知路径上“遇见”决策，而非“调用”决策。系统不再依赖单一效用模型，而是能够根据情境自动选择最合适的理论，实现了由多种认知决策理论共同支撑的分布式理性。

2.3.2 由节点知识引导的认知活动：情绪与性格

在本架构中，引入了情绪与性格作为调控智能体认知活动的重要因素。在自我认知图谱中，情绪与性格均以节点知识的形式存在，其中情绪节点对应一组模拟激素参数的知识单元，而性格节点则对应这些激素受体的长期调节机制。通过节点间的转移函数计算，系统在认知流动过程中实现了对注意力、推理范围与思维风格的持续动态调控。

（1）情绪节点：基于激素参数的多维知识建模

情绪节点以激素为核心建模元素，每个节点代表一种激素相关的知识单元，并携带描述该激素主导情绪的语义向量。系统通过预定义知识将生理机制映射为语义状态，例如：当“多巴胺”节点激活度上升时，系统在知识图谱中表现为“愉悦”与“探索性”倾向的增强；当“皮质醇”节点权重升高时，转移函数会偏向收缩路径，聚焦于低风险、高确定性的区域。由此，复杂的多维情绪可由多种激素节点的联合激活模式表示。

（2）性格节点：激素受体的长期调控

性格节点用于模拟激素受体的敏感性特征，其参数反映系统对各类激素节点信号的响应强度。在长期的转移与经验积累过程中，这些节点会根据任务反馈与情绪波动自动调整权重，形成稳定的个体化响应模式。因此，性格可被视为一种由节点知识长期演化形成的“受体分布图”，定义了系统在不同认知状态下的稳定风格与偏好。这种结构化的受体网络为认知过程提供了持续的调性约束与长期行为一致性。

（3）情绪—转移函数耦合机制

在转移计算中，转移函数不仅考虑节点间的语义相似度，还同步参考当前情绪节点的状态。系统通过计算下一个候选节点与活跃情绪节点在图谱中的距离，动态调整转移方向与强度：

$$P_{slide}(n_{t+1}) \propto \exp\big(-\lambda \cdot Dist(n_{t+1}, E_{emotion})\big) \tag{3}$$

其中，$Dist(\cdot)$ 为节点间的语义距离，$E_{emotion}$ 表示当前活跃的情绪节点嵌入向量，λ 为情绪调节系数。该设计使情绪节点成为转移过程中的“能量场”，通过影响节点选择概率，间接决定认

知迁移路径。同时，激素强度与节点激活度直接绑定，通过上文的激活度反馈机制，形成对情绪动态的实时模拟。

（4）认知意义
情绪与性格节点的引入，使认知过程具备了内在的动态调节能力。情绪提供即时的能量偏向与思维拓扑扰动，性格提供长期的调控基调与风格一致性。两者在节点层面共同作用，使系统在知识图谱中表现出兼具灵活性与稳定性的思维流动特征。因此，认知活动不再是纯粹的逻辑计算，而成为由节点知识内部能量动态所引导的复杂自组织过程。

2.3.3 由节点知识引导的认知活动：预期

在人类认知体系中，对动作结果的预期是决策与学习的核心驱动力之一。用于预测潜在反馈、奖惩后果及环境变化。这种预测性认知不仅存在于显性决策阶段，还贯穿于感知、情绪调节与思维生成等多层次认知过程中。基于此，本架构设计在知识图谱的动作节点中连接一个有关引入动作预期的节点以实现类人式的预测性思维与前瞻性行为控制。

(1) 动作节点的预期建模 在系统的语义记忆层中，每个动作节点不仅记录动作的定义与执行特征，还连接一个或多个**预期结果节点**。这些向量描述该动作在不同情境下可能导致的多维后果，包括:

- 环境变化
- 情绪反馈
- 奖励或惩罚信号
- 社会反应与信任变化

在描述更复杂的预期时，预期节点可以连接更多次级结果与远期影响，并在边关系上指导转移注意力头探索次级节点

在转移到相关预期节点后, 系统可根据节点提供的信息套用出当前场景中该动作可能产生的预期，从而输出决策节点所需要的有关动作节点的决策信息。

(2) 预期的适用范围与阶段性作用 (Applicability and Stage Function) 动作预期不仅用于决策阶段。在本架构中，其作用范围覆盖三个主要阶段:

- **感知阶段：**预期用于预测即将到来的感知输入，调整感知显著度分布，使系统提前聚焦于高价值感知通道;
- **决策阶段：**预期模型作为动作选择的主要参考依据，与情绪节点、奖励模型共同参与期望效用计算;

(4) 认知意义 动作预期节点的引入，使知识图谱能够动态模拟未来情境。这种机制强化了系统的主动性与前瞻性：在面对未知环境时，系统可通过预期节点在内部先行模拟未来可能，以最小的代价探索高收益路径;

2.3.4 由节点知识引导的认知活动：程序化行动与逻辑推理

在人类的认知中，当行动目标具备明确逻辑与可预测条件时，思维常会自发地组织出具有程序化特征的行动结构。这种“程序性思维”体现为：根据条件生成分支、设定终止规则、调用内在推算。为模拟这一理性化规划能力，本架构提出 **代码模板驱动的行动机制**，即：行动节点内部可以保存**代码模板**；系统在具体情境中根据当前信息即时填充模板，生成实例化的逻辑结构并执行，从而实现语义—逻辑—计算三层的动态统一。

(1) 行动节点的模板结构 用 T_{code} 表示该行动的逻辑模板，包含参数化占位符：

$$T_{code} = \text{“IF \{\$condition\} THEN \{\$action\} ELSE \{\$alt\}”} \tag{4}$$

系统在认知执行阶段根据当前上下文状态 $S_{context}$，填充模板参数：

$$L_{inst} = \text{Instantiate}(T_{code}, S_{context}) \tag{5}$$

从而得到一段临时生成的逻辑脚本 L_{inst}。

(2) 模板实例化与执行 模板的实例化遵循**上下文映射规则**：

$$\$condition \Leftarrow f_{cond}(S_{context}), \quad \$action \Leftarrow f_{act}(A_i), \quad \$alt \Leftarrow f_{alt}(S_{context}) \tag{6}$$

例如，当节点“避开障碍物”被激活时，模板内容可能为：

```
IF distance(object, agent) < threshold THEN change_path()
```

系统根据当前感知数据自动替换参数，并在逻辑层执行对应函数调用。

(3) 逻辑推理通路:符号化规则演算与知识整合 为确保程序化行动的严谨性与可验证性,CTAM 内置 **逻辑推理通路（LRP）**，当事件框架中检测到主语—谓语—宾语槽间存在逻辑完备性特征时，系统调用大语言模型（LLM）执行逻辑结构识别：

$$Q_{logic} = \text{LLM}(E_{frame}) \Rightarrow \{Predicate(x, y),\ Rule(a \to b)\} \tag{7}$$

识别出的逻辑形式输入推理引擎（如 Prolog、DLV、PyReasoner）执行符号推演：

$$Inference:\ \{Premises\}[\text{LogicEngine}]\{Conclusions\} \tag{8}$$

结论以高置信度节点形式回写至语义记忆库，并更新边权 $w_{edge} \uparrow \beta$，以强化因果与条件规则的稳定连接；同时，逻辑结论不仅作为事实节点存储，还作为**逻辑证据边**标注于相关节点间，使后续转移推理可自动参考逻辑约束，在“语义扩散”与“逻辑收敛”之间实现动态平衡。该通路与 CTAM 深度耦合：推理结论可直接填充模板参数（如 $condition、$action），确保生成的 L_{inst} 具备形式化可证明性。

(4) 认知意义 代码模板驱动的行动机制将程序语言的形式严谨性融入类人思维的动态语义网络，并通过内置的逻辑推理通路实现**语义—逻辑—计算—验证**的闭环统一，使得每个行动节点既可作为语义实体，又可在必要时生成**可执行、可验证的逻辑实例**。

3 局限性与未来研究方向

尽管本架构在理论设计上具备较强的通用智能潜力，但在实际实现与应用中仍面临多方面挑战与局限，未来研究需重点突破以下几个方面：

3.1 感知模块的局限性与特征引导神经网络的设想

当前主流的感知模块多依赖卷积神经网络（CNN）结构，虽在静态图像识别等标准任务中取得显著成果，但其训练高度依赖封闭标签数据集与预定义的特征提取模板，导致在通用人工智能（AGI）所需的动态、开放式环境中表现出明显局限。CNN 难以有效应对新奇场景、罕见事件及未标注样本，缺乏灵活的特征迁移与实时适应能力，无法实现类人类的感知主动性与深层语义理解。

为突破上述瓶颈，预提出一种融合元学习机制的特征引导神经网络设想，作为未来感知模块演化的重要方向。具体而言，在每个感觉通道（如视觉、听觉、触觉等）中引入基于 MAML 算法的元学习器，用于快速适应新任务与未见特征。这些元学习器通过实时提取输入信号中的低阶感知特征（如颜色、边缘、纹理、频率等），结合其频次、置信度与历史关联经验，动态生成特征偏好向量。与此同时，该机制可进一步结合复杂物体分块策略（如分解为形状、纹理），借鉴 AIMA 中基于结构的物体识别方法，并融合点云技术以支持三维场景的真实感知与多模态信息整合，从而在物理及场景几何结构层面实现更高保真的感知。

在空间定位方面，系统可自动基于视觉中心构建动态坐标系以处理物体位置信息。当智能体发生移动时，坐标系随之偏移以保持空间一致性，并在必要时引入 GPS 定位数据作为补充参考，从而提升大尺度场景中的位置精度与环境稳定性。

特征引导神经网络据此调整感知通道内训练样本的关注度与局部学习率，使模型优先聚焦于高频、高价值特征，抑制低相关性噪声信息，从而提升学习效率、特征表达质量与泛化能力。同时，为实现跨模态感知的一致性，该架构在更高层次引入“元元学习器”结构，对多个感觉通道的特征偏好进行整合抽象，提取诸如视觉与触觉中的物体属性一致性、听觉与情感信号中的情绪表达关联等通用模态内结构，进一步强化感知系统的迁移能力与认知对齐能力。

该自适应特征引导感知机制旨在摆脱传统模型僵化特征模板的限制，构建具备开放式、实时自适应、跨模态与空间一致性感知能力的系统，为 AGI 在复杂多变环境中的持续感知演化与认知自主成长提供支撑。

3.2 大规模图谱运算的优化策略

随着知识图谱规模的持续扩展，节点数量和边的复杂度呈指数增长，导致图谱查询、推理与更新的计算成本迅速攀升。当前图谱动态扩展与实时推理机制在大规模场景下性能受限，容易出现响应延迟和资源瓶颈。未来研究需探索图谱分布式存储、图神经网络加速、近似推理算法以及图谱压缩与知识蒸馏技术，实现在保持推理精度的前提下，提升系统的时效性与可扩展性。

3.3 内源情绪与激素模型的生理合理性与计算复杂度

情绪调节机制依赖于模拟多种激素与神经递质的参数变化，虽能丰富系统行为表现，但其生理机制的简化模型尚无法完美捕捉人类情感的多样性和微妙变化。过于复杂的激素交互模型

又会带来较高计算负担，限制实时响应能力。未来应结合神经科学最新成果，设计更高效且生理拟合度更高的情绪模型，并研究多尺度情绪状态与认知行为的耦合机制。

3.4 自我反思与元认知机制的深度实现

如何使其真正实现深度的自我监控、自我纠错及自我优化，仍处于探索阶段。有效的自我反思机制需要整合长短期记忆、情绪状态和推理过程的反馈，形成闭环改进体系。未来研究方向包括引入强化学习中的自我监督机制，以及结合心理学和认知科学的元认知模型，实现更具人类特质的自我意识能力。

3.5 空间感知与想象能力的缺失

尽管当前内源思维模块已初步实现类人化的认知模拟与创造性思维生成，但在空间感知与空间想象能力方面仍存在明显不足。系统尚未具备对三维空间关系、物体布局及动态环境变化的深度建模与推理能力，限制了其在涉及空间因果关系、场景构建与具象推理等任务中的表现。同时，当前的认知模拟多以语义维度为主，缺乏对空间结构的象征性编码与生成机制，无法有效支持复杂的假想场景构建与具身认知模拟。因此，未来研究可进一步引入基于空间嵌入和视觉—语义联合建模的机制，构建具有空间构型理解、虚拟空间想象与环境重构能力的空间感知子模块，强化内源思维中关于场景推理、任务规划与具身交互的认知能力，从而推动 AGI 系统在具象思维与具身智能方向上的进一步发展。

3.6 神经网络化的知识图谱与基于经验的公理可塑性建模

传统知识图谱通常以静态的图结构表达固定事实与规则，缺乏基于主观感受、经验积累或情境变化的动态自我重构能力。而在面向通用人工智能的系统中，知识不应被视为恒定不变的，而是应能随着感受、反馈与环境交互经验而持续调整，包括对原有事实关系乃至“公理性”知识的权重、适用性与结构的重新评估。这一理念促使知识图谱向神经网络化结构演化，即通过引入端到端可训练的神经机制，在节点与边的表示中嵌入可微分的记忆强度、信任度与情感权重，使图谱具备对新输入和多模态刺激的自适应重构能力。与此同时，这种神经网络支撑下的图谱结构，也为表达潜意识提供了更拟人化且灵活的通道。具体而言，系统可将潜意识内容以低显性度的向量权重持续编码于图谱中，借助梯度驱动的方式不断影响显性认知路径的生成与偏好调整。这种机制不仅增强了知识图谱的“认知弹性”与表达精度，也更接近人类以模糊情绪、模态记忆和非逻辑性联想方式处理信息的心理现实，为构建具备自我成长、自我修正与思维多样性的 AGI 认知结构提供了坚实基础。

4 结语

本文围绕通用人工智能的发展需求，提出了一种以人类思维本质为启发的新型认知架构，融合转移函数注意力机制、多维知识图谱嵌入、感知驱动控制系统与内源思维模块，构建具备类人认知流动性、自主性与稳定性的推理体系。该架构不仅从机制层面重构了联想性思维、非线性跳跃、情绪调节与自我监控的动态过程，更在结构设计上强调模块化实现路径与落地可行性，为构建具备持续学习、情境适应与创造性生成能力的 AGI 系统提供了基础支撑。

未来研究将进一步拓展该架构在多模态对齐、长程记忆调控、结构性元反思、以及人格演化机制等方向的表达能力，并探索其在自主任务建构、人机协同交互与真实世界复杂推理场景中的应用潜力。面向类人智能系统的演进目标，本文所提出的认知结构为具备可解释性、可拓展性与可迁移性的通用智能系统探索提供了新的理论支点与工程方向。

参考文献

[1] Baddeley, A. D. (2000). The episodic buffer: a new component of working memory?. *Trends in Cognitive Sciences*, 4(11), 417–423.

[2] Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29.

[3] Anderson, J. R. (1983). *The Architecture of Cognition*. Harvard University Press.

[4] Anderson, J. R., & Lebiere, C. (1996). ACT-R: A theory of higher level cognition and its relation to visual attention. *Cognitive Systems Research*.

[5] Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.

[6] Minsky, M. (1974). *A Framework for Representing Knowledge*. MIT AI Memo 306.

[7] Schank, R. C., & Abelson, R. P. (1977). *Scripts, Plans, Goals, and Understanding*. Lawrence Erlbaum Associates.

[8] Tulving, E. (1972). *Episodic and semantic memory*. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381–403). Academic Press.

[9] Tulving, E. (1983). *Elements of Episodic Memory*. Oxford University Press.

[10] Rumelhart, D. E. (1975). *Notes on a schema for stories*. In D. G. Bobrow & A. Collins (Eds.), *Representation and Understanding: Studies in Cognitive Science* (pp. 211–236). Academic Press.

[11] Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9(1), 77–110.

[12] Stein, N. L., & Glenn, C. G. (1978). Understanding and remembering stories: A developmental analysis. *Discourse Processes*, 1(2), 283–311.

[13] Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–185. https://doi.org/10.1017/S0140525X01003922

[14] Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 411–421. https://doi.org/10.1037/0278-7393.28.3.411

[15] McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3), 817–835. https://doi.org/10.1037/0278-7393.27.3.817

[16] Oberauer, K. (2019). Working memory and attention –A conceptual analysis and review. *Journal of Cognition*, 2(1), 36. https://doi.org/10.5334/joc.58

[17] Vergauwe, E., Barrouillet, P., Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science*, 21(3), 384–390. https://doi.org/10.1177/0956797610361340