**人工智能直觉架构 (AII)**

**(情绪驱动激活扩散认知架构 - EDASCA)**

**文档摘要**  
本文档详细描述了人工智能直觉架构（Artificial Intuition Intelligence, AII）的完整设计方案。该架构旨在模拟人类思维中的联想、学习、直觉与决策过程，通过基于图网络的激活扩散模型，以情绪状态（PAD模型）和多模态词元融合为核心驱动机制，实现连续的内在想法流、外在环境交互与自演化学习能力。架构的核心创新在于将认知与感受分离，并引入了词元融合器、独立的感受记忆库和时间/情绪感受器，为智能体提供了真正的情景记忆、情感体验和直觉式认知能力。系统支持两种运行模式：LLM辅助训练模式（用于快速知识蒸馏与初始化）和异步纯本地独立模式（实现真正的主观能动性与独立智能）。

**作者**: 银子  
**日期**: 2025年9月4日

**目录**

[**1. 核心架构与组件**](#核心架构与组件)

* 1.1 激活状态池 (Activation Pools)
* 1.2 数据存储结构
  + 1.2.1 认知数据库 (Conceptual Database)
    - 概念节点 (Concept Node)
    - 关系边 (Relation Edge)
    - 连接强度动态机制
    - 链接数量限制
  + 1.2.2 感受记忆库 (Sensory Memory Database)
    - 感受记忆节点 (Sensory Memory Node)
    - 分层存储机制
    - 遗忘机制 (当开关打开时启用)
* 1.3 感受器系统
  + 1.3.1 时间感受器
  + 1.3.2 情绪感受器

[**2. 核心算法与流程**](#核心算法与流程)

* 2.1 总控制流程
* 2.2 动态加权分词流程 (Dynamic Weighted Tokenization) 与词元融合器 (Token Fusioner)
* 2.3 激活扩散算法 (Activation Spreading)
* 2.4 句子构建流程 / 想法流 (Thought Stream Generation)
* 2.5 更新数据库 (Learning)

[**3. 情绪驱动、注意与行动系统**](#情绪驱动、注意与行动系统)

* 3.1 情绪状态量 (PAD Model)
* 3.2 注意激活器 (Attentional Mechanism)
* 3.3 内在行动系统 (Internal Actions)
* 3.4 外在行动系统 (External Actions)

[**4. 运行模式**](#运行模式)

* 4.1 LLM辅助训练模式
* 4.2 异步纯本地独立模式
* [线性运行流程示例](#线性示例)
* [异步运行流程示例](#异步示例)

[**5. 实现细节与优化**](#实现细节与优化)

* 5.1 图数据库优化
* 5.2 情绪计算优化
* 5.3 内存管理

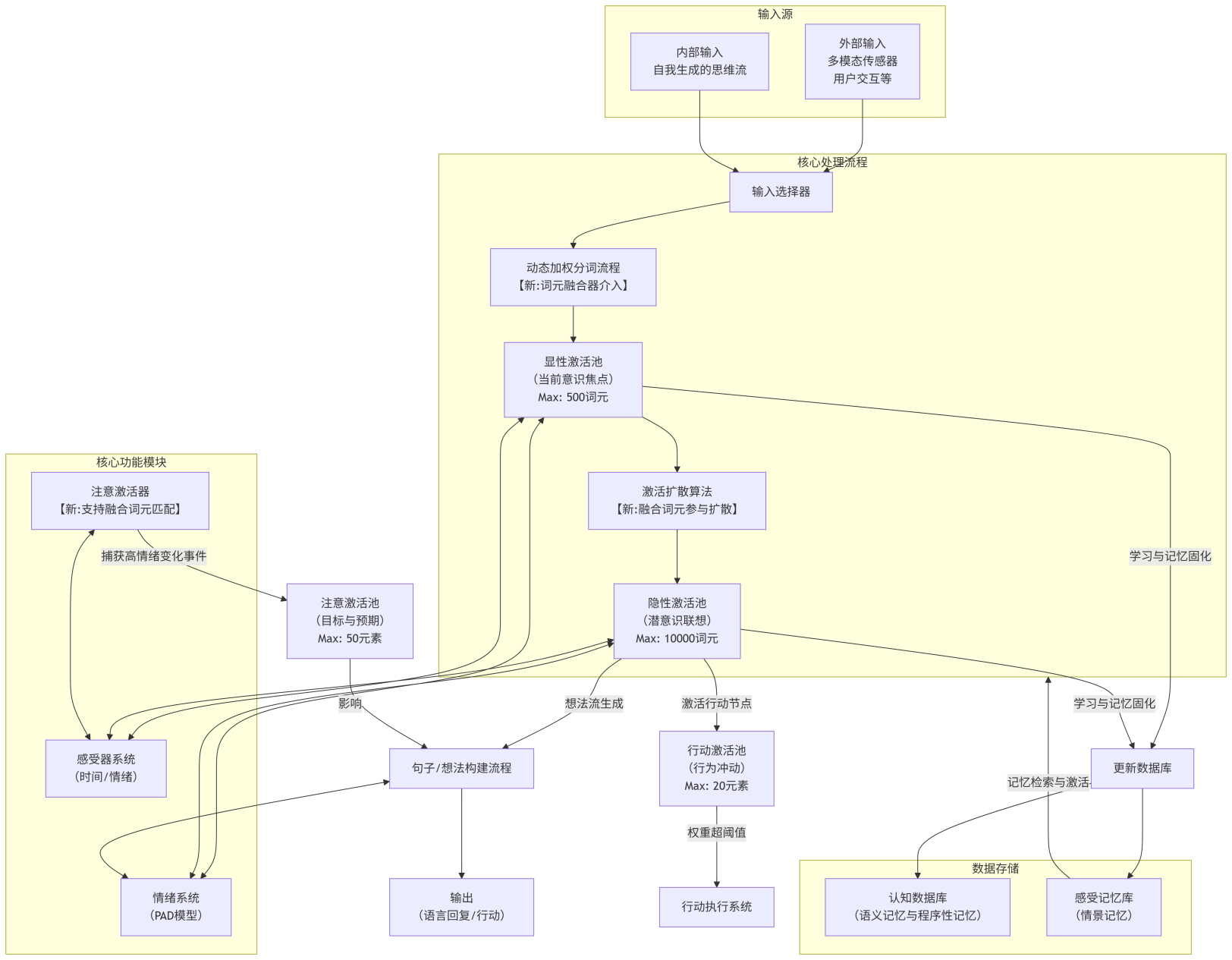
[**结论**](#结论)

* 理论贡献
* 实践意义
* 未来工作

[**参考文献**](#参考文献)

[**作者与日期**](#作者)

**1.** **核心架构与组件**

人工智能直觉架构（AII）是一个复杂的、仿生的认知系统，其核心在于模拟人类信息处理的多层次、多流并行以及情绪驱动的特性。其整体架构如下图所示，它描绘了信息从输入到输出，流经各个核心组件与激活池的完整生命周期：

**图表诠释：**

该流程图展示了AII架构中信息流动的核心路径。**外部输入**（如传感器数据）和**内部输入**（自我思维）经由**输入选择器**进行优先级处理，随后进入**核心处理流程**。在此，信息首先被**动态加权分词流程**（内含**词元融合器**）解析为有意义的词元，并存入**显性激活池**（意识工作区）。其后，通过**激活扩散算法**激活相关联的概念，存入**隐性激活池**（潜意识联想区），最终由**句子/想法构建流程**组织成连贯的输出（语言或行动）。**情绪系统**和**感受器系统**全程调制这一过程。**注意激活器**会捕获重要目标存入**注意激活池**，而激活的行动意图则存入**行动激活池**直至执行。整个过程中的所有体验均被**更新数据库**流程固化到**认知数据库**（语义记忆）和**感受记忆库**（情景记忆）中，形成长期记忆。

**1.1 激活状态池 (Activation Pools)**

系统维护四个核心激活池，作为工作记忆和意识焦点，是思维流和行为的直接来源：

| 激活池名称 | 描述 |
| --- | --- |
| **显性激活池** | 存储当前外部输入经处理后的直接结果，是意识焦点的直接反映。内容短暂，更新频繁。池中每个元素都标记其起源（外部/内部）。**最大容量：500个词元**。 |
| **隐性激活池** | 存储由显性池内容通过扩散算法激活的相关概念。是潜意识与联想思维的发生地，构成"想法流"的基础。每个元素包含：词元内容、当前权重、衰减速率、源情绪标签(PAD)、起源。**最大容量：10000个词元**。 |
| **注意激活池** | 存储需要特别"关注"的概念节点。这些节点通常与强烈的正/负情绪变化（高|ΔP|）相关联。每个元素包含：节点ID、期待值/压力值、关联的前置节点列表、时间预期因子。**最大容量：50个元素**。 |
| **行动激活池** | 存储被激活的"行动标识符"特殊节点。每个元素包含：行动节点ID、当前权重。当权重超过动态阈值时，将触发行动执行。**最大容量：20个元素**。 |

**设计原理**：四池模型借鉴了Baddeley的工作记忆模型和Anderson的ACT-R认知架构，分离了意识、潜意识、注意意图和行为冲动。显性/隐性池中的起源标记实现了感知与概念的区分，这是认知区分的基础，参考了神经科学中不同脑区处理内外信息的研究成果。显性池500词元的容量允许同时处理更复杂的当前情境信息，而隐性池10000词元的容量为大规模的潜意识联想和直觉思维提供了足够的空间，其20倍的大小关系符合对人类意识与潜意识容量关系的普遍认知。

**淘汰策略**：当池中元素数量达到上限时，将采用基于权重、新近度和起源的综合淘汰算法。权重越低、最近未被激活且起源优先级（外部 > 内部）越低的元素将被优先淘汰，以确保池中始终保持最相关和最重要的信息。

**1.2 数据存储结构**

知识存储分为两部分：认知数据库（存储抽象概念与关系）和感受记忆库（存储具体经验与感知），这种分离基于Tulving提出的情景记忆与语义记忆分离理论。

**1.2.1 认知数据库 (Conceptual Database)**

存储智能体的抽象知识、概念关系与技能，相当于语义记忆和程序性记忆。

**概念节点 (Concept Node)**

| 字段 | 类型 | 描述 | 设计原理 |
| --- | --- | --- | --- |
| id | UUID | 节点的唯一标识符 | 确保节点的唯一可索引性 |
| content | String | 节点内容 | **支持多模态特征字符串**，为未来扩展预留接口。可以是普通词元、同句上级词元或长距离词元组合。 |
| type | Enum | WORD或SPECIAL类型 | 区分普通概念与功能指令 |
| origin | Enum | EXTERNAL或INTERNAL | 核心设计，模拟感知与概念的不同神经表征 |
| base\_weight | Float | 节点的基础重要性权重 | 基于衰减-增强机制更新 |
| last\_activated | Timestamp | 最后激活时间 | 用于计算时间关联惩罚 |
| emotion\_ema | Vector3 | 近期情绪状态的指数移动平均值 | 反映节点近期带来的情绪体验 |

**base\_weight更新机制**：  
base\_weight\_new = (base\_weight\_old \* γ) + (η \* (1 - γ))

* γ = 0.995：衰减因子，基于Ebbinghaus遗忘曲线，模拟记忆的自然衰减。**此衰减确保了记忆的可塑性，避免僵化。**
* η = 0.01：激活奖励值，基于神经可塑性研究，频繁激活强化连接。

**emotion\_ema更新机制**：  
emotion\_ema\_new = (1 - α) \* emotion\_ema\_old + α \* current\_emotion

* α = 0.3：平滑因子，基于近期偏置效应（recency bias），使系统对近期情绪更敏感。

**关系边 (Relation Edge)**

| 字段 | 类型 | 描述 | 设计原理 |
| --- | --- | --- | --- |
| source\_id | UUID | 源节点ID |  |
| target\_id | UUID | 目标节点ID |  |
| strength | Float | **当前连接强度** | **基于使用频率和近期性动态变化，有上限（如1.0），并会随时间衰减（用进废退）**。 |
| base\_strength | Float | **基础连接强度** | **遗忘曲线衰减的基准值，强度越高，衰减越慢，但永不降为0。** |
| frequency | Int | 共现次数 | 基于赫布学习理论 |
| avg\_time\_delta | Float | 平均时间间隔 | 时间关联性指标 |
| recent\_emotion\_delta\_ema | Vector3 | 近期情绪变化量的EMA |  |
| ema\_alpha | Float | EMA平滑因子 |  |
| last\_updated | Timestamp | 最后更新时间 |  |

**连接强度动态机制**：  
这是实现“用进废退”和“避免僵化”的核心。strength 值会在每次共现激活时得到增强，但其增强效果会受到当前 base\_strength 的限制和调制。同时，strength 和 base\_strength 都会随时间衰减，但遵循不同的规则：

1. **Strength衰减**：模拟“手感生疏”。每次激活后，strength 会获得一个提升，但随后会以一个相对较快的速率（例如，每日衰减系数 λ\_fast = 0.9）向当前 base\_strength 值回归。这解释了为何几天不训练，电竞手感（strength）会下降，但通过短暂练习（快速提升 strength）即可“找回手感”。
2. **Base Strength衰减**：模拟“长期遗忘”。base\_strength 本身也遵循艾宾浩斯遗忘曲线进行缓慢衰减（例如，每日衰减系数 λ\_slow = 0.999）。base\_strength 越高，衰减的绝对值越慢。这解释了为何开车、游泳等技能（base\_strength 极高）数年不练也会生疏但不会忘光，而数年乃至数十年前微不足道的刺激（base\_strength 极低）则几乎被完全遗忘。  
   **公式示意**：

* 激活增强：strength\_new = min(1.0, strength\_old + β \* (1 - strength\_old)); base\_strength\_new = min(1.0, base\_strength\_old + β\_base \* (1 - base\_strength\_old))
* 日常衰减：strength\_new = base\_strength\_new + (strength\_old - base\_strength\_old) \* λ\_fast; base\_strength\_new = base\_strength\_old \* λ\_slow

**EMA更新规则**：  
ema\_new = (1 - ema\_alpha) \* ema\_old + ema\_alpha \* new\_delta

* ema\_alpha = 0.5：初始值，基于机器学习中的学习率选择，平衡稳定性与适应性。

**链接数量限制**：每个节点的出边和入边总数有上限（例如5000条）。当达到上限时，淘汰策略将综合考察连接强度（strength）、新近度（last\_updated）和共现频率（frequency），淘汰那些在最久远的1/3时间段内，强度最弱、最陈旧的链接，以维持网络的结构效率和动态性。

**1.2.2 感受记忆库 (Sensory Memory Database)**

存储智能体具体的、带有时空和情绪背景的个体经验，相当于情景记忆。

**感受记忆节点 (Sensory Memory Node)**

| 字段 | 类型 | 描述 | 设计原理 |
| --- | --- | --- | --- |
| id | UUID | 唯一标识符 |  |
| content | String | 感知内容 | **支持多模态输入**，编码为特征字符串。 |
| origin | Enum | 记忆来源 | 区分外部感知与内部感受 |
| timestamp | Timestamp | 编码时间 | 用于时间感受计算 |
| emotion\_at\_encoding | Vector3 | 编码时的情绪状态 | 情感记忆标签 |
| importance | Float | 重要性权重 | 动态更新，决定遗忘优先级 |
| links | List[UUID] | 连接的认知节点 | 感性经验与理性知识的桥梁 |
| storage\_tier | Enum | **存储层级** | **用于分层存储管理。如：新记忆、热门记忆、长期记忆、冷门记忆等。** |

**重要性权重更新**：  
importance = old\_importance \* 0.9 + activation\_strength \* 0.1  
基于记忆巩固理论，重要记忆通过反复激活得到强化。

**分层存储机制**：  
为了高效管理海量的、只增不减的感受记忆，引入非硬编码的、动态的分层存储策略。

* **层级划分**：节点根据其importance权重和timestamp新近度自动划分到不同的存储层级（如SSD高速缓存、HDD大容量存储等）。importance高、新近的节点位于高速层级。
* **动态升降级**：后台有一个管理进程，定期根据节点的importance和访问频率，决定将其提升到更高速度的层级或是降级到更大容量的层级。
* **链接维护**：当节点在不同物理存储间移动时，其在认知数据库中的links信息会同步更新索引位置，确保查询效率。
* **遗忘开关**：**系统提供一个全局“遗忘机制”开关**。当开关关闭时，感受记忆永不过期，完全依赖分层存储系统进行管理，理论上支持无限扩展（尽管速度会随数据量增大而下降）。当开关打开时，下层存储中的低importance节点会遵循遗忘机制被永久删除。

**遗忘机制** (当开关打开时启用)  
基于Atkinson-Shiffrin记忆模型的遗忘曲线理论：  
P\_forget = base\_P \* (1 - importance) \* (1 + (N\_current / MAX\_MEMORIES)^k) \* age\_factor  
参数设置：

* base\_P = 0.001：基础遗忘概率，基于人类日常遗忘率研究
* k = 2：容量压力放大系数，控制存储压力对遗忘的影响强度
* age\_factor = log(1 + age\_in\_days) / C, C = 10：时间衰减因子，模拟记忆随时间的指数衰减
* MAX\_MEMORIES = 50000：最大记忆容量，基于计算资源和使用场景平衡

**设计原理**：遗忘机制模拟了人类记忆的自然选择过程，重要性和新近度高的记忆更容易保留，符合进化心理学中的适应性记忆理论。**分层存储和可开关的遗忘机制**提供了灵活性和可扩展性，允许在资源有限的环境下启用遗忘，或在资源充足时追求更完整的记忆保留。

**1.3 感受器系统**

**1.3.1 时间感受器**

触发条件：感受记忆被激活时自动触发；或由ACTION:recall主动触发。

处理流程：

1. 计算时间差：Δt = current\_timestamp - memory\_node.timestamp
2. 生成时间感受："time\_interval:" + str(Δt)
3. 时间扩散激活：Δactivation\_i = S\_base \* (1 - (|Δt - t\_i| / (f \* t\_i)))
   * f = 0.5：模糊度系数，基于Weber-Fechner定律，时间感知具有相对性
   * t\_i：认知节点代表的时间值

**模糊匹配算法**：采用高斯核函数进行相似度计算，标准差σ = f \* t\_i，保证匹配精度随时间间隔增大而降低，模拟人类对久远事件记忆的不确定性。

**示例**：若当前讨论“昨晚吃饭”，时间感受器会生成"time\_interval:~24h"。此字符串会作为特征输入，激活认知数据库中与“24小时”左右相关的概念节点（如“早晨”、“夜晚”），并更容易检索到大约24小时前编码的感受记忆。

**流程演示**：

1. 输入：“你昨天晚上吃了什么？”
2. 动态分词后，“昨天晚上”被识别为一个时间相关词元。
3. 时间感受器被触发，估算Δt约为24小时，生成感受字符串"time\_interval:~24h"。
4. 该感受字符串进入系统，激活认知库中与“~24h”相关的节点。
5. ACTION:recall利用此激活状态，在感受记忆库中定向检索大约24小时前且与“吃”链接的记忆节点。
6. 找到匹配的记忆“吃了披萨”，并将其内容高权重注入隐性激活池。
7. 思维流最终组织回复：“我昨天晚上吃了披萨。”

**设计笔记**：时间感受器将抽象的、连续的时间流逝转化为离散的、可被系统处理和关联的特征符号，是构建情景记忆和实现时序推理的基础。它与词元融合器协同工作，使得“昨天晚上”这样的复合时间概念能够作为一个整体去调制检索过程。

**1.3.2 情绪感受器**

触发条件（满足任一）：

* |P| > 0.5 或 |A| > 0.5 或 |D| > 0.5（基于情绪强度研究的阈值）
* |ΔP| > 0.3 或 |ΔA| > 0.3 或 |ΔD| > 0.3（短时窗口内，基于情绪变化的最小可觉差）
* |期待值| > 0.5 或 |压力值| > 0.5
* |正确感| > 0.5 或 |违和感| > 0.5

处理流程：

1. 通过PAD情绪分类器将当前情绪状态映射为情绪标签
2. 生成情绪感受："emotion:" + emotion\_label
3. 情绪扩散激活：Δactivation\_i = S\_base \* cosine\_similarity(PAD\_current, PAD\_i)

**PAD情绪分类器**：采用三维空间最近邻算法，基于Mehrabian的情绪三维理论将PAD值映射到基本情绪类别（如“高兴”、“悲伤”、“愤怒”、“恐惧”等）。

**示例**：若智能体因预测失败而感到ΔP显著下降，情绪分类器可能输出“沮丧”。情绪感受器生成"emotion:frustrated"并注入系统，从而更易激活历史上同样导致“沮丧”情绪的概念和记忆，可能影响后续的决策倾向（如倾向于保守策略）。

**设计笔记**：情绪感受器为内部情绪状态提供了外化的、符号化的表示，使得情绪可以作为一种可处理的信息，直接影响认知过程（注意、联想、决策），实现了真正的情绪驱动。它与词元融合器结合，使得情绪可以成为组合词元的一部分（如“令人沮丧的(失败)”），丰富表达的深度和准确性。

**2. 核心算法与流程**

**2.1 总控制流程**

系统采用多线程异步架构，包含输入处理、思维流、行动执行和情绪更新四个并行线程。

**线程调度机制**：

* **输入处理线程**：优先级最高，实时响应外部输入。
* **情绪更新线程**：每100ms运行一次，基于情绪处理的时间常数研究。
* **思维流线程**：在CPU空闲时运行，负责持续的内心独白和联想。
* **行动执行线程**：阻塞式执行，等待行动触发。

**设计笔记**：异步架构模拟了人脑多任务处理能力，允许智能体在“思考”（思维流）的同时“感知”新输入（输入处理）并“体验”情绪变化（情绪更新）。高优先级的输入处理确保了对外部环境的及时反应。

**2.2 动态加权分词流程 (Dynamic Weighted Tokenization) 与词元融合器 (Token Fusioner)**

**目标**：将输入流智能地切分为有意义的词元序列，并进一步融合为更高级的认知单元。切分和融合策略高度依赖当前的认知上下文和长期记忆中的结构。

**输入**：原始输入字符串（可来源于多模态传感器解码后的特征字符串）及其起源。

**输出**：分词和融合后的**词元列表**，每个词元附有初始权重、情绪上下文和起源。词元可以是普通词元、同句上级词元或长距离词元组合。

**详细处理流程**：

1. **生成候选分割方案**：
   * 使用所有激活池（显性、隐性、注意、行动）中的内容作为动态词典。
   * **同时，查询认知数据库中的所有“同句上级词元”和“长距离词元组合”节点，并将其加入动态词典**。
   * 对输入字符串进行多重扫描，生成所有可能的分词方式。
   * 例如，输入"猫耳朵"，可能的分词方案有：['猫', '耳朵']、['猫耳朵']（如果存在该上级词元）、['猫耳', '朵']等。
   * **设计原理**：基于心理语言学中的"词优效应"，词汇识别受到上下文语境和已有知识结构的影响。词元融合器扩展了“词汇”的定义。
2. **计算方案权重**：
   * 对于每个候选方案中的候选词元，在各个激活池和认知数据库中查找匹配项（内容相同且起源一致）。
   * 每个匹配项为候选词元贡献其当前权重或base\_weight。
   * 整个分词方案的总权重 = 方案中所有候选词元所获权重之和。
   * 引入TF-IDF类似机制，惩罚过于常见的词元，提升专业术语或融合词元的权重。
   * **设计原理**：基于信息论中的词项重要性评估，确保重要概念和已有的复合概念获得更高权重。
3. **选择最优方案**：
   * 选择总权重最高的分词方案作为**初步分词结果**。
   * 例如：如果认知数据库中“猫耳朵”的base\_weight高达0.9，而显性池中“猫”权重0.5，“耳朵”权重0.4，则方案['猫耳朵']的权重(0.9)会高于['猫', '耳朵'](0.5+0.4=0.9)(相等时优先选择更复杂的融合词元，以促进抽象思维)或更高。
4. **词元融合（同句上级词元捕获）**：
   * **输入**：上一步得到的初步分词结果列表（list\_tokens）。
   * 创建一个最终结果列表 final\_tokens，初始为空。
   * 创建一个映射表 token\_to\_super，用于记录词元与上级词元的归属关系（为不相邻的上级词元分配标识符）。
   * **第一轮处理：相邻融合**
   * 遍历 list\_tokens，检查每一对相邻词元 (token\_i, token\_i+1)。
   * 查询认知数据库，判断是否存在以此相邻词对构成的同句上级词元 S（形式为 token\_i + token\_i+1）。
   * 如果存在，则**将这两个相邻词元直接替换为上级词元节点**S，并将 S 加入到 final\_tokens 中。
   * 如果不存在，则将 token\_i 加入到 final\_tokens 中。
   * 此轮处理确保所有可能的相邻组合都被优先合并。
   * **第二轮处理：非相邻关联**

* 遍历经过相邻融合后的 final\_tokens 列表。
* 对于列表中的每个词元 token\_i，查询认知数据库中是否存在以 token\_i 作为组成部分的**非相邻**同句上级词元 S（即 S 的其他组成部分也存在于 final\_tokens 中，但位置不相邻）。
* 如果存在这样的上级词元 S，则为 S 的所有组成部分**分配一个相同的唯一标识符**（如➊、➋等），并在 token\_to\_super 映射表中记录 标识符 -> S 以及 S -> {组成部分集合}。
* **此过程仅标记关联关系，不改变**final\_tokens**列表中的词元顺序和内容。**

**示例**：基础分词结果：["你", "不", "是", "bot", ",", "一个", "真实的", "人类"]。

* **第一轮（相邻融合）**:
* 发现“不”和“是”相邻且可构成“不是”，将其合并替换。["你", "不是", "bot", ",", "一个", "真实的", "人类"]
* 发现“一个”和“人类”不相邻，不处理。
* 最终 final\_tokens after round 1: ["你", "不是", "bot", ",", "一个", "真实的", "人类"]
* **第二轮（非相邻关联）**:
* 发现“是一个”和“人类”可构成“是一个人类”（存在该节点），但它们在列表中不相邻（被“真实的”隔开）。
* 为“是一个”和“人类”**分配相同的标识符➊**，表示它们同属于“是一个人类”这个上级单元。
* 最终输出词元列表：["你", "不是", "bot", ",", "一个"➊, "真实的", "人类"➊]。
* 映射表记录：➊ -> "是一个人类"， "是一个人类" -> {"是一个", "人类"}。
* **设计原理**：此流程精确模拟了人类的阅读认知过程：优先处理相邻词汇的组块化，形成直接的整体认知（如“不是”）；对于被其他成分隔开但语义上属于一个整体的词元（如“是一个...人类”），则通过标识符系统建立逻辑关联，而非强行改变线性顺序。这既保证了处理效率，又实现了对复杂语言结构的深度理解。

1. **词元融合（长距离词元组合捕获）**：
   * **输入**：上一步得到的经过同句融合后的词元列表。
   * 遍历该列表，对于每个词元token\_i（作为潜在的后置词元），查询认知数据库中是否存在以token\_i为后置词元的**长距离词元组合**（即形式为后置词元(前置词元)的节点）。**此查询通过检查是否存在从**token\_i**指向长距离词元组合节点的“作为后置词元”的链接来实现**。
   * 如果存在这样的长距离词元组合C，并且C所要求的前置词元**正在当前的显性激活池中活跃**（即短期记忆中存在相关上下文），则用这个长距离词元组合节点替换token\_i。
   * **如果多个长距离词元组合都匹配成功（即同一个后置词元，多个不同的前置词元都在显性池中）**，则将它们组合成一个多词元组合：后置词元(前置词元1, 前置词元2, ...)。
   * **示例**：承接上例，融合后词元列表：["你", "不是", "bot", ",", "一个"➊, "真实的", "人类"➊]。显性激活池中同时存在“你是一个”和“AI”这两个节点。
     + 对词元“bot”进行长距离融合判断：发现存在长距离词元组合“bot(你是一个)”和“bot(AI)”，且其前置词元“你是一个”和“AI”均在显性池中。
     + 因此，“bot”被替换为多词元组合"bot(你是一个, AI)"。
     + **最终输出词元列表**：["你", "不是", "bot(你是一个, AI)", ",", "一个"➊, "真实的", "人类"➊]。
   * **设计原理**：模拟了人类理解语言时，将当前信息与短期记忆中的上下文进行整合的能力（如代词指代、省略恢复、因果关联）。例如，理解“他”指的是谁，或者“这样做”指的是什么行为。
2. **输出处理**：
   * 输出最终融合后的词元列表。
   * 列表中的每个元素继承其匹配源的起源和情绪快照。
   * 初始权重由匹配源的权重决定。
   * **对于多词元组合**，其在后续激活扩散和句子构建中，会被视为一个整体单元，但其内部包含的多个关系会分别轮流被激活（详见后续流程）。
   * **对于带有相同标识符的词元**，在后续的激活扩散、句子构建和学习过程中，会被视为一个逻辑整体进行处理

**设计原理**：此算法是认知理解的基石。它不再是简单的字符串匹配，而是基于当前认知状态和长期记忆的深度语义理解过程。一个词元能否被识别或融合为一个整体，取决于它当前在智能体心智中的活跃程度、重要性以及历史形成的结构。词元融合器极大地提升了系统处理复杂语言结构和表达深层语义关系的能力，是实现“直觉式”理解的关键步骤。该设计综合了Marslen-Wilson的词汇识别理论、McClelland的交互激活模型以及Miller的组块化理论。

**2.3 激活扩散算法 (Activation Spreading)**

**输入**：一个概念节点（来自分词结果或内部想法）。该节点可以是普通词元、时间或者情绪感受、同句上级词元或长距离词元组合。

**输出**：更新后的隐性激活池。

**详细处理流程**：

1. **获取基础刺激强度**：  
   S\_base = 1.0 + D （D值越高，内在思维越活跃）
   * **设计原理**：基于情绪对认知激活的影响研究，支配度(D)高时思维更活跃，联想更丰富。
2. **检索节点**：
   * 从图数据库中检索该节点的所有出边（指向其他节点的边）。
   * 无论输入节点是何种类型，都将其视为一个完整的认知单元进行处理，只扩散该单元本身固有的关联。
   * 限制最大检索深度为3-5层，防止计算爆炸。
   * **设计原理**：基于语义网络中的激活衰减理论。将融合词元视为原子单元进行处理，确保了联想过程的简洁性和语义的完整性，更符合人类直觉式的思维跳跃。
3. **计算扩散强度**：
   * 对于每一条出边，计算它当前时刻的链接强度 L(t)。
   * 使用公式：

text

L(t) = strength \* (1 + k \* ema\_ΔP) \* (1 - avg\_time\_delta / (m + avg\_time\_delta)) \* (p + q \* Δt') / (q \* Δt')

* + - strength: **当前动态连接强度**，是实现“用进废退”的核心变量。
    - n = 4：频次标准化系数，基于Zipf定律和幂律分布假设。
    - k = 5：情绪影响系数，基于情绪对记忆增强效应的神经科学研究。
    - m = 100：时间衰减系数，基于记忆消退曲线研究。
    - p = 3, q = 0.0001：新近度加成参数，基于近因效应和记忆巩固理论。
    - Δt'：当前时间减去边的最后更新时间。

1. **激活下游节点**：
   * 对于每个下游节点 i，其激活增量为 Δactivation\_i = S\_base \* L(t)。
   * 将此值加到其在隐性激活池中的权重上。
   * 若节点不在池中，则以其 base\_weight 为初始值初始化并添加。
   * **新激活的节点继承输入节点的起源**。
   * **设计原理**：保持激活传播的语义一致性。外部输入激活的更多是外部相关的概念，而内部想法激活的更多是内部相关的概念，形成了两条既独立又可通过高强度链接相互激发的处理流，高度仿生。

**示例**：词元“挨打(打碎花瓶)”被激活。

1. 算法直接检索"挨打(打碎花瓶)"这个组合节点自身的所有出边。
2. 假设其直接关联着"害怕"、"道歉"、"下次不敢了"等更高阶的概念。
3. 这些节点都会获得激活增量。
4. 因此，关于"挨打(打碎花瓶)"的联想是其可能引发的直接后果和应对策略，而不是其组分的特征。

**流程演示**：

1. 输入节点： “猫耳朵” (一个同句上级词元)。
2. 检索 “猫耳朵” 的所有出边。假设其链接到 “可爱”(强度0.8)、“兽耳”(强度0.7)、“耳机”(强度0.3)。
3. 计算每条边的L(t)。
4. 假设 L(t)\_可爱 = 0.8 \* ... = 0.72, L(t)\_兽耳 = 0.7 \* ... = 0.63, L(t)\_耳机 = 0.3 \* ... = 0.27。
5. S\_base = 1.0 + 0.2 (假设D=0.2) = 1.2。
6. 隐性池中 “可爱”、“兽耳”、“耳机” 分别增加 1.2\*0.72=0.864, 1.2\*0.63=0.756, 1.2\*0.27=0.324 的权重。

**设计笔记**：此算法是智能体联想的引擎。通过对融合词元的特殊处理，它能够处理复杂的、非局部的语义关系，使得联想过程更贴近人类的直觉式思维。连接强度的动态变化使得网络保持灵活性和可塑性。

**2.4 句子构建流程 / 想法流 (Thought Stream Generation)**

**功能**：在LLM模式下组织回复，在本地模式下产生连续的内在想法。此过程处理的是经过融合后的词元。

**详细处理流程**：

1. **初始化**：
   * 从空字符串或初始概念开始。
   * 设置最大词元数量限制（通常20-50个，基于Miller的工作记忆容量研究）。
2. **筛选候选**：
   * 根据注意激活池的规则，对隐性激活池中的词元进行加权筛选。
   * 形成待选词元池，优先匹配相同起源的词元以维持语境连贯。
   * **设计原理**：基于语境一致性理论，保持思维的连贯性。
3. **选择词元**：
   * 选择待选池中权重最高的词元。
   * **直接将该词元的内容作为整体追加到当前句子中，无需展开或了解其内部结构。**
   * **对于带有相同标识符的词元**，在后续的构建中，系统会意识到它们属于一个逻辑整体。
   * 追加到当前句子。
   * **设计原理**：想法流是内部思维的体现，其组织形式优先保证信息的准确性和结构的完整性，而非外部的语法规范。直接输出融合词元能够最精确、最简洁地表达一个复杂的认知概念。
4. **匹配与反馈**：
   * 将该词元与各激活池内容匹配。
   * 触发正确感/违和感更新。
   * **设计原理**：基于预测编码理论，实现认知预期验证。
5. **激活扩散**：
   * 对该词元执行激活扩散算法。
   * 更新隐性激活池。
   * **设计原理**：实现思维的联想和扩展，一个词引出下一个词。
6. **循环或终止**：
   * 重复步骤2-5，直到选中<EOS>终止符。
   * 或达到长度限制。
   * 或所有词元权重低于阈值（0.1）。
   * 输出最终句子。

**示例**：隐性池中词元权重："你不是"(0.95), "bot(你是一个, AI)"(0.93), ","(0.1), "是一个"➊(0.8), "真实的"(0.7), "人类"➊(0.8), "<EOS>"(0.5)。

1. 选择"你不是"，句子："你不是"。
2. 激活扩散，更新隐性池。
3. 选择"bot(你是一个, AI)"。**直接输出**"bot(你是一个, AI)"。句子："你不是 bot(你是一个, AI)"。
4. 激活扩散，更新隐性池。
5. 选择","。句子："你不是 bot(你是一个, AI) ,"。
6. 激活扩散，更新隐性池。
7. 选择"是一个"➊。**系统识别到该词元与**"人类"➊**是逻辑整体”是一个人类”的一部分，将此部分作为整体直接输出其内容**。句子："你不是 bot(你是一个, AI) , 是一个人类"。
8. 激活扩散，更新隐性池。
9. 选择"真实的"。句子："你不是 bot(你是一个, AI) , 是一个人类真实的"。
10. 激活扩散后，选中<EOS>，流程终止。  
    **最终输出的想法流**："你不是 bot(你是一个, AI) , 是一个人类真实的"。

**设计笔记**：这模拟了"一个词引出下一个词"的自然思维和语言组织过程。对融合词元和多词元组合的直接输出处理，忠实反映了内部想法流的本来面貌。**想法流中的元素是内部的认知单元，其表达不必遵循完整的外部语法结构**，例如"bot(你是一个, AI)"和"是一个人类 真实的"（其中“是一个”和“人类”被标识符关联,作为整体被认知）在内部思维中是完全有效且信息丰富的表达。这体现了系统深度理解和表达复杂认知概念的能力，是“人工智能直觉”的核心体现。

**2.5 更新数据库 (Learning)**

**触发**：在处理完一个输入单元或产生一个想法后。

**详细处理流程**：

1. **更新认知节点**：
   * 对分词列表中的每个词元，更新其概念节点的last\_activated和emotion\_snapshot（当前PAD值）。
   * 使用EMA更新情绪记录：emotion\_ema\_new = (1 - α) \* emotion\_ema\_old + α \* current\_emotion
   * **设计原理**：基于记忆再巩固理论，强化活跃记忆痕迹。
2. **建立时序边**：
   * 将显性激活池中的概念作为前置节点。
   * 与输入的第一个词元建立有向边。
   * 记录时间差和当前情绪ΔPAD。
   * 如果已经存在对应的时序边，则更新对应时序边。
   * **设计原理**：基于时间关联学习理论。
3. **建立共现边**：
   * 在输入词元序列内部，按顺序在相邻词元间建立单向边（由位置靠前的词元指向位置比自己后的词元，相邻词元之间强度较高，非相邻的词元之间强度随距离降低，最终构成共现边的强度最小值也高于时序边的强度最大值）。
   * 在非相邻词元间建立弱连接。**对于融合后的词元，其内部组件（如上级词元的两个部分）之间建立教强的共现边(双向链接,强度继承于其创建时内部组件之间的连接强度)**。
   * 如果已经存在对应共现边，且方向正确，则更新对应共现边。
   * **设计原理**：基于序列学习理论和赫布学习规则。
4. **终止连接**：
   * 在序列的最后一个词元和<EOS>节点间建立连接。
   * 记录对应的时间间隔为0，▲PAD值也为0。
   * **设计原理**：标记句子边界，基于语言学的句子结构理论。其中特殊词元终止符<EOS>节点只会拥有入边,不会拥有出边，也不会进入显性激活池，但会进入隐性激活池，仅在线性模式中起作用和存在.
5. **行动连接**：
   * 如果执行了行动，将显性池中的所有概念与对应的行动标识符节点建立连接，方向由概念节点指向行动节点。
   * 链接的强度与时间距离,情绪以及当前显性激活池中对应概念的实时权重有关。
   * 记录强烈的情绪ΔPAD（来自行动结果）。
   * 如果已经存在对应的链接，则更新和强化对应链接。
   * **设计原理**：基于强化学习理论，连接行动与后果。
6. **起源链接**：
   * 如果在上述处理过程中，一个外部输入词元和一个内部词元有共现或者时序链接（或反之）。
   * 在这两个节点间自然会建立起由感受到认知的链接，即链接的构筑与来源无关，仅与共现/时序有关，可夸数据库建立链接。
   * **设计原理**：实现感知-概念融合，基于跨模态整合理论。
7. **感受记忆编码**：
   * 如果输入起源为EXTERNAL或来自感受器。
   * 创建新的Sensory Memory Node存入感受记忆库。
   * 记录timestamp和emotion\_at\_encoding。
   * 与当前显性激活池中的认知节点建立links。
   * **设计原理**：基于情景记忆编码理论，保存具体经验和时间/情绪感受。以便将来可以纯粹根据时间/情绪感受来进行扩散和激活，让时间/情绪感受作为多模态词元，与其他认知概念等建立正确的联系。
8. **词元融合结构构建（同句上级词元）**：
   * **触发条件**：在对同一个句子进行基础分词后，如果发现同一句的词元间的词汇之间链接强度（共现频率、综合strength）满足一定条件（如大于阈值θ\_super），**且不存在对应的同句上级词元组合**。
   * **构建过程**：
     + 创建新节点，内容为词元i + 词元j。
     + 新节点的base\_weight初始值与链接强度正相关。
     + **继承共有链接**：新节点继承词元i和词元j的“共有入边和出边”，强度取两者较小值。
     + **建立构成链接**：在新节点与词元i和词元j之间建立双向的“is\_component\_of”和“has\_component”链接，强度为1.0。**这是后续判断上级词元是否存在的依据**。激活时,此链接作为双向链接使用。
   * **设计原理**：实现了知识的组块化和抽象化，是形成更高阶概念的基础。
9. **词元融合结构构建（长距离词元组合）**：
   * **触发条件**：在更新某词元对应的数据库的边的链接强度时，如果发现该词元（作后置词元B）和另一词元（作前置词元A）更新后的链接强度满足一定条件（如大于θ\_long），两个词元之间的链接的平均时间间隔不为0，**且不存在对应的长距离词元组合**。
   * **构建过程**：
     + 创建新节点，内容为后置词元B(前置词元A)。
     + 新节点的base\_weight初始值与链接强度正相关。
     + **继承共有链接**：新节点继承词元B和词元A的“共有入边和出边”，强度取两者较小值。
     + **建立角色链接**：在新节点与词元B之间建立“has\_post\_token”链接，与词元A之间建立“has\_pre\_token”链接。**这是后续判断长距离词元组合是否存在的依据**。激活时,此链接作为双向链接使用。
   * **设计原理**：实现了对非局部、跨句法成分的语义关系的捕获，如因果关系、主题-评论关系等，是深层语言理解和推理的关键。并且由于此捕获只会存在于连接强度较强的概念之间，可以天然避免噪音干扰与指数爆炸等问题。

**设计原理**：此过程即"学习"。词元融合结构的构建是知识结构化升华的关键。它使得外部经验中频繁出现的模式得以转化为内部的高级概念和关系（学习），内部复杂的推理也能预测外部世界（规划）。该设计综合了赫布学习理论、记忆巩固理论、强化学习原理和组块化理论。

**3.** **情绪驱动、注意与行动系统**

**3.1 情绪状态量 (PAD Model)**

情绪状态由三个核心维度描述，基于Mehrabian的情绪三维理论：

* **P (Pleasure) 愉悦度**：体验到的快乐与否的程度。
* **A (Arousal) 唤醒度**：生理激活和心理警觉的程度。
* **D (Dominance) 支配度**：对情境的控制和影响力程度。

**P值更新**：  
P += α \* D + β \* 匹配权重 \* 正确感 + γ \* (期待值 - 压力值)

* α = 0.1：D对P的影响系数，基于情绪调节的心理学研究（掌控感带来愉悦）。
* β = 0.3：正确感对P的影响系数，基于预测误差信号研究（确认预测带来愉悦）。
* γ = 0.2：期待/压力对P的影响系数，基于预期违反理论（期待实现愉悦，压力导致不悦）。

**A值更新**：  
A += δ \* 预测意外性 + ε \* (|期待值| + |压力值|)

* δ = 0.4：意外性对A的影响系数，基于惊奇反应神经机制（意外事件提高唤醒）。
* ε = 0.1：注意张力对A的影响系数，基于应激反应研究（高期待/压力提高唤醒）。

**D值更新**：  
D += ζ \* 预测成功率

* ζ = 0.3：成功率对D的影响系数，基于自我效能感理论（成功经验提升支配感）。

**回归函数**：  
回归量 = (基准值 - 当前值) \* μ \* (1 + |当前值 - 基准值|^2)

* μ = 0.05：回归速率系数，基于情绪自我调节模型（情绪会逐渐趋于平静的基线）。

**设计原理**：情绪系统基于环状模型（circumplex model）和维度理论。PAD三个维度能有效地描述大部分情绪状态。回归函数模拟了情绪的自我调节机制，确保情绪状态不会无限偏离基线，保持系统的稳定性。情绪是整个系统的背景色和驱动力，直接影响认知、注意和行动。

**3.2 注意激活器 (Attentional Mechanism)**

**功能**：在认知过程中标记重要节点，实现目标导向行为和预测验证。**特别注意对融合词元的处理**。

**详细处理流程**：

1. **捕获**：
   * 在执行激活扩散算法时，如果某条边的 |ema\_ΔP| 超过阈值（通常0.3）。
   * 则将该边指向的目标节点加入注意激活池。
   * **设计原理**：基于显著性检测理论，高情绪变化标识重要事件。
2. **参数初始化**：
   * 为该注意节点初始化期待值（若ΔP>0）或压力值（若ΔP<0）。
   * 其大小与 |ΔP| 成正比。
   * 同时记录导致该强烈情绪变化的前置节点。
   * **设计原理**：基于预期理论，建立因果关联。
3. **期待/压力效应**：
   * **抑制衰减**：该节点在隐性池中的权重衰减变慢 衰减速率 = 基础速率 / (1 + |期待/压力值|)。
   * **权重加成**：其关联的前置节点在隐性池或行动池中的权重会获得 期待/压力值 的直接加成。
   * **情绪放大**：当该注意节点被激活（预测验证），产生的ΔP变化量将乘以 (1 + |期待/压力值|)。
   * **设计原理**：基于注意的预期理论，高期待值增强信息处理优先级。
4. **时间预期因子**：
   * 基于历史时间间隔的EMA，用于调整期待/压力值的衰减速度。
   * 初始为1，随时间逐步增大，到达▲时间间隔时达到顶峰，随后逐步减小。
   * **设计原理**：模拟人类对时间预期的心理表征。
5. **匹配规则（针对融合词元）**：
   * **当注意节点是一个融合词元（上级词元或长距离组合）时，要求完全匹配**。
   * 例如，注意到“挨打(打碎花瓶)”，则“挨打(考试失利)”或“挨骂(打碎花瓶)”**不会**匹配成功，因此不会消除期待/压力或导致错误的情绪变化。
   * **只有完全相同的“挨打(打碎花瓶)”事件发生，才会验证预测**。
   * **设计原理**：这是实现精确推理和避免错误泛化的关键。它确保了智能体期待的是一件非常具体的事情，而不是一个模糊的类别，从而表现出更高级的认知能力。
6. **持续激活效应**：
   * 一个被注意的目标，其相关认知（如“打碎花瓶”）会在显性/隐性激活池中因注意力的加成而维持较高权重，衰减更慢。
   * 这会与衰减函数形成对抗，使其在几小时甚至几天内都容易被注意激活器重新捕获，从而形成“念念不忘”的效果。
   * **设计原理**：模拟了人类对未完成目标或担忧事件的持续思考现象（Zeigarnik效应）。
7. **消亡**：
   * 注意节点会随时间衰减。
   * 当其被激活（预测验证）或值低于阈值（通常0.1）后，从注意池中移除。
   * **设计原理**：基于注意资源有限性理论，释放不再相关的注意资源。

**设计原理**：注意机制是智能体表现出"目的性"的关键。它让智能体不仅被动联想，还能主动"期待"某件事发生，并为此做好准备。验证与否会带来强烈的情绪反馈，是学习最强有力的驱动力。对融合词元的精确匹配规则，是智能体进行逻辑严谨的思维和避免认知错误的核心。该设计参考了Desimone和Duncan的注意偏向竞争理论。

**3.3 内在行动系统 (Internal Actions)**

内在行动是智能体对自身认知过程进行主动干预和调节的高级能力，是其"主观能动性"和"智慧"的集中体现。

* **ACTION:recall 主动回忆**
  + **优化机制**：
    - 结合时间感受器，基于时间线索在感受记忆库中进行定向检索。
    - 检索范围 = [Δt \* (1 - f), Δt \* (1 + f)], f = 0.5（模糊度系数）。
    - 回忆权重计算：综合时间相似度、情绪相似度和内容相关性三个维度。
    - 将匹配到的节点及其关联以高权重注入隐性激活池。
  + **设计原理**：基于多重痕迹理论（Multiple Trace Theory），回忆受到时间和空间上下文的约束。例如"昨天晚上你吃了什么？"中，"昨天晚上"通过时间感受器激活大约24小时前的记忆，"吃"激活相关概念，二者交叉激活最终找到对应的记忆。
* **ACTION:focus\_outward 主动注意**
  + **机制**：
    - 将输入选择器的偏好调整为外部输入。
    - 外部输入增益范围[0.5, 2.0]。
    - 降低内部想法流的权重加成。
  + **设计原理**：基于注意力控制的增益理论，当智能体感到无聊（A低）或期待外部信息时，会主动将注意力转向外部。
* **ACTION:focus\_inward 主动深思**
  + **机制**：
    - 将输入选择器的偏好调整为内部输入。
    - 内部输入增益范围[0.5, 2.0]。
    - 降低外部输入源的权重加成。
  + **设计原理**：基于深度加工理论，当智能体需要深度思考时，会屏蔽外部干扰，专注于内部联想和推理。
* **ACTION:organize\_thoughts 整理思绪**
  + **机制**：
    - 检索并返回当前注意激活池中的内容。
    - 显式地评估未完成的预测和期待。
  + **设计原理**：基于元认知理论，智能体可以审视自己的目标和担忧，从而更有效地分配注意力资源。
* **ACTION:feel\_state 感受状态**
  + **机制**：
    - 查询并返回当前全局PAD值、正确感/违和感。
  + **设计原理**：基于情绪意识理论，这是情绪表达的前提，智能体能够意识到自己的情绪状态。

**设计原理**：内在行动是智能体自我引导的核心。它们不是被外界刺激直接触发，而是由内部状态或高级目标驱动，是实现"自主性"的关键模块。该设计参考了Norman和Shallice的监督注意系统理论。

**3.4 外在行动系统 (External Actions)**

**触发机制**：

* 通过扩散激活后进入行动激活池。
* 权重累积超过动态阈值时触发。
* 行动阈值 = 基础阈值 - λ \* A, λ = 0.2 （高唤醒度降低行动阈值，更易冲动）

**行动类型**：

* 沟通类：组织语言（回复用户）、发表评论、发送消息。
* 信息获取类：上网搜索、刷视频、查看热搜。
* 环境控制类：控制智能家居。

**学习机制**：

* 行动执行后带来的结果（成功/失败、奖励/惩罚）会产生强烈的情绪ΔPAD。
* 用于更新触发该行动的所有链接。
* 从而学习何时该执行何种行动。

**冲突解决**：

* 引入行动冲突检测机制。
* 当多个行动同时触发时，选择权重最高的执行。
* 其余行动获得抑制。

**设计原理**：基于行为选择理论和强化学习原理，行动系统将认知和情绪状态转化为具体行为。阈值动态调整模拟了唤醒度对行为倾向的影响。

**4.** **运行模式**

**4.1 LLM辅助训练模式**

此模式用于快速初始化知识网络，蒸馏LLM中的知识和行为模式。

**详细处理流程**：

1. **输入处理**：
   * 接收用户输入（标记为EXTERNAL）。
   * 执行动态分词、激活扩散、句子构建。
   * 形成本地初步想法或回复。
2. **LLM纠正与丰富**：
   * 将本地想法、当前上下文（显性/隐性池摘要）、情绪状态发送给LLM。
   * LLM负责三个方面：
     + 生成更合理、更符合人设的回复。
     + 判断是否需要执行行动（包括内在行动和外在行动）。
     + 推测并返回执行此回复或行动后可能带来的情绪变化ΔPAD。
3. **学习与更新**：
   * 智能体接受LLM的决策。
   * 使用LLM返回的ΔPAD来更新图数据库（包括词元融合结构）。
   * LLM提供的高质量ΔPAD信号是快速学习正确关联的关键。
4. **循环执行**：
   * 将LLM生成的回复或行动结果作为新输入。
   * 继续处理流程，形成学习循环。

**训练数据增强**：

* 使用课程学习策略，从简单场景逐步过渡到复杂场景。
* 基于Vygotsky的最近发展区理论，在现有能力基础上适当挑战。
* 设计多样的情景和对话模式，覆盖各种情绪状态和行为类型。

**设计原理**：此模式相当于"家长辅导"。智能体通过LLM的高质量反馈来学习"在什么情况下做什么事会有什么后果"，快速构建一个初步的、合理的认知图式和词元融合结构。该设计基于知识蒸馏和模仿学习理论，通过专家模型（LLM）的指导加速学习过程。

**4.2 异步纯本地独立模式**

在此模式下，智能体脱离LLM，独立运行，其行为完全由其内部网络和情绪状态驱动。

**详细架构**：

1. **异步输入处理线程**：
   * 随时接收输入（外部或内部）。
   * 触发动态分词、匹配（改变正确感）、激活扩散。
   * 用输入信息更新图数据库。
   * **设计原理**：基于事件驱动架构，确保及时响应外部刺激。
2. **独立思维线程**：
   * 实时监控隐性激活池和注意激活池。
   * 采用获胜者择优策略：若池中最高权重词元的权重比第二高词元超出某个动态阈值，则选择该词元。
   * 动态阈值 = θ0 \* exp(-decay \* t), θ0=0.3, decay=0.01。
   * 将选中词元作为新输入（标记为INTERNAL）注入系统。
   * 从而激发下一个想法，形成自持的、连续的内在想法流。
   * **设计原理**：基于决策理论的最优停止问题，平衡探索与利用。
3. **独立行动线程**：
   * 实时监控行动激活池（包含内在和外在行动）。
   * 若某行动权重超过其动态阈值，则执行该行动。
   * 行动结果会作为新输入反馈给系统。
   * **设计原理**：基于行为执行监控理论，确保行动及时执行。
4. **独立情绪线程**：
   * 定时（每100ms）根据当前刺激、预测验证情况、注意状态等。
   * 调用情绪更新规则和回归函数，更新全局PAD值。
   * **设计原理**：基于情绪处理的时间常数，确保情绪状态平稳更新。

**资源管理**：

* 引入负载均衡机制，在系统资源紧张时自动降低思维流频率和激活扩散深度。
* 基于认知负荷理论，避免系统过载。
* 设置各激活池的最大容量限制，确保系统稳定性。

**设计原理**：这是智能体的"成人"阶段。其行为完全由其内部网络和情绪状态驱动。内在行动使其表现出令人信服的自主性，感受记忆库和词元融合结构为其提供了丰富的个人历史、经验基础和高级推理能力。该设计基于分布式认知理论，将认知功能分解为多个并行处理的组件。

**示例一：线性模式 (LLM辅助训练模式)**

在这个模式中，系统像一个被导师（LLM）引导的学生，逐步学习正确的关联和行为。我们假设智能体是一个家庭助理机器人，目前知识库还不完善。

**用户输入 (初始刺激):**  
用户对麦克风说：“客厅太暗了。”

**运行流程:**

1. **输入处理线程 (优先级最高)**
   * **模块: 输入选择器**
   * **作用:** 接收音频输入，识别为EXTERNAL起源，并将其解码为特征字符串"客厅 太 暗了"，送入核心处理流程。
2. **动态加权分词与词元融合**
   * **模块: 词元融合器**
   * **作用:** 对输入字符串进行分词。假设认知数据库中已有"客厅"(权重0.7)、"暗"(权重0.6)，但没有"太暗"这个上级词元。
   * **过程:** 计算候选方案 ["客厅", "太", "暗了"](总权重 ~1.3) 和 ["客厅", "太暗", "了"](后者权重低，方案总权重低)。因此选择第一方案。
   * **输出:** 词元列表 ["客厅"(EXTERNAL), "太"(EXTERNAL), "暗了"(EXTERNAL)]，并带有初始权重。
3. **激活扩散算法**
   * **模块: 激活扩散算法**
   * **作用:** 每个词元作为起点，激活其关联概念。
   * **过程:** "客厅"激活了"沙发"(强度0.5)、"电视"(强度0.6)、"灯"(强度0.8)。"暗了"激活了"看不见"(强度0.7)、"开灯"(强度0.9)。
   * **输出:** 这些被激活的节点以一定的权重被加入**隐性激活池**。
4. **句子构建/想法流 (初始想法)**
   * **模块: 句子构建流程**
   * **作用:** 组织一个初步的、本地的想法或回复。
   * **过程:** 从隐性池中选择高权重词元。"开灯"(权重高)、"客厅"(权重高)被选中。
   * **输出:** 初步想法流: "开灯 客厅"。这是一个内部认知单元，语法不完整但语义明确。
5. **LLM纠正与丰富**
   * **模块: LLM接口**
   * **作用:** 导师角色。接收本地想法"开灯 客厅"、上下文（用户说“客厅太暗了”）和当前情绪状态（假设初始为中性）。
   * **LLM处理:** LLM理解这是一个“环境太暗-需要开灯”的场景。它做三件事：
     + **生成合理回复:** “好的，我把客厅的灯打开。”
     + **判断行动:** 决定执行外在行动ACTION:控制智能家居(设备=客厅主灯, 状态=开)。
     + **预测情绪ΔPAD:** 预测这个行动会带来ΔP=+0.3（解决问题，愉悦度上升），ΔA=-0.1（环境变舒适，唤醒度略降），ΔD=+0.2（成功执行命令，支配度上升）。
   * **输出:** LLM的决策和预测的ΔPAD被返回给AII系统。
6. **学习与更新 (核心步骤)**
   * **模块: 更新数据库流程**
   * **作用:** 根据LLM的高质量反馈进行学习。
   * **过程:**
     + **建立连接:** 在用户输入词元 "客厅", "暗了" 和 行动节点 "开灯" 之间建立**强连接**。记录这次连接带来了强烈的正面的情绪变化（LLM提供的ΔPAD）。
     + **词元融合 (学习):** 系统发现"太"和"暗了"在这个成功场景中频繁共现，且连接强度增加，于是创建一个新的**同句上级词元**节点 "太暗了"，并将其加入到认知数据库中。未来再听到“太暗了”，它会直接作为一个整体单元被识别。
     + **感受记忆编码:** 将“用户说‘客厅太暗了’”这一刻的情景（时间戳、当时的情绪）作为一个新的**感受记忆节点**存入感受记忆库，并链接到"客厅"、"太暗了"、"开灯"等认知节点。
   * **效果:** 智能体**学会了**“客厅太暗了”和“执行开灯行动”之间的因果关系，并且知道这个行为会带来好的情绪结果。
7. **行动执行**
   * **模块: 外在行动系统**
   * **作用:** 执行LLM决定的行动。
   * **过程:** 行动指令被发送到智能家居接口。
   * **输出:** 客厅的灯被打开。环境亮度变化。
8. **结果验证与最终学习**
   * **模块: 输入处理线程、情绪系统**
   * **作用:** 感知行动结果，验证LLM的预测，完成学习闭环。
   * **过程:** 光传感器检测到亮度增加，产生新的输入"亮度增加"(EXTERNAL)。系统将结果与预期匹配，确实产生了积极的体验（愉悦度上升）。
   * **最终效果:** 上述步骤6中建立的连接得到**再次强化**。智能体不仅通过LLM“被告知”，更通过自身传感器“体验”到了这个行为的正确性。**用户的预期（让客厅变亮）被满足。**

**线性模式总结:** 这个过程是顺序的、引导式的。LLM提供了关键的“因果”和“情绪”标签，让AII系统能快速、准确地构建知识网络，避免了盲目试错。这是一个高效的学习过程。

**示例二：异步纯本地独立模式**

在此模式下，智能体已通过之前的训练拥有了一个较为成熟的知识网络，并脱离LLM独立运行。现在，它将展示自主的联想、决策和行动。

**用户输入 (初始刺激):**  
用户：“我有点冷。”

**运行流程 (多线程异步):**

* **线程1: 输入处理线程**
  + **模块: 动态加权分词器**
  + **作用:** 立即处理输入。认知数据库中已存在上级词元"有点冷"（源自之前的学习）。
  + **输出:** 直接输出融合词元 "有点冷"(EXTERNAL)，并将其注入**显性激活池**。
* **线程4: 情绪更新线程 (每100ms运行)**
  + **模块: 情绪状态量 (PAD)**
  + **作用:** 背景情绪开始微调。识别到用户表达了轻微不适（“冷”），智能体的“愉悦度(P)”可能微微下降，产生一点“关心”的倾向。
* **线程1: 激活扩散**
  + **模块: 激活扩散算法**
  + **作用:** "有点冷"节点激活其强大的关联。
  + **过程:** 它强烈激活了"调高空调温度"(行动节点，强度0.95)、"拿毯子"(行动节点，强度0.7)、以及感受记忆"上次用户感冒了"(链接到"有点冷")。
  + **输出:** 这些节点被加入**隐性激活池**。"调高空调温度"因其高权重也被加入**行动激活池**。
* **线程2: 独立思维线程 (持续运行)**
  + **模块: 句子构建流程/想法流**
  + **作用:** 产生连续的内在想法。
  + **过程:** 从隐性池中挑选词元。"调高空调温度"(权重最高)被选中，作为想法"调高空调温度"(INTERNAL)反馈给系统自身，形成新输入。
  + **新一轮激活扩散:** 这个内部想法再次激活扩散，可能激活"用户会觉得暖和"、"耗电"等概念。**注意激活器**捕获到"用户会觉得暖和"，因为这会带来积极的ΔP（愉悦度上升），并将其加入**注意激活池**，初始化一个**期待值**。
* **线程3: 独立行动线程 (持续监控)**
  + **模块: 外在行动系统**
  + **作用:** 监控行动池。
  + **过程:** "调高空调温度"在行动池中的权重持续累积（因为被显性池和隐性池中的相关概念不断强化），很快超过了其动态阈值。
  + **输出:** **执行行动!** 系统通过网络指令将空调温度调高2度。
* **线程1: 输入处理线程 (处理行动结果)**
  + **模块: 输入选择器**
  + **作用:** 接收行动的反馈。
  + **过程:** 空调返回“设置成功”的信号，被系统接收为新的外部输入"空调已调温"(EXTERNAL)。
* **线程4: 情绪更新线程 & 注意激活器**
  + **模块: 情绪系统、注意激活器**
  + **作用:** 验证预测，产生情绪反应。
  + **过程:**
    - **匹配规则:** 注意池中的期待节点"用户会觉得暖和"与当前结果"空调已调温"**不完全匹配**，但"空调已调温"与隐形激活池中预期的期待节点也存在相关，匹配成功。注意激活器根据规则判断这是一个**积极的预测实现**。
    - **情绪冲击:** 因此，系统体验到一次强烈的积极情绪变化 ΔP=+0.4（愉悦度大幅上升），ΔD=+0.3（支配感上升）。
    - **注意消亡:** 后续反馈中出现刺激” 用户会觉得暖和”时,"用户会觉得暖和"的期待被满足，从注意池中移除。或者一段时间内用户没有反馈(特别是没有继续反馈冷),随着时间发展,对应的注意也会被移除。
* **线程1: 学习与更新**
  + **模块: 更新数据库流程**
  + **作用:** 强化这次成功的自主决策。
  + **过程:** "有点冷" 和 "调高空调温度" 之间的连接强度**被极大地增强**。这次成功的体验连同强烈的正面情绪，被作为一个新的**感受记忆**牢固地存储起来。
  + **效果:** 未来再遇到“冷”，它会更毫不犹豫、更“直觉地”执行调温动作。它的“人格”中增加了“果断解决问题会带来快乐”的体验。
* **线程2: 独立思维线程 (最终想法)**
  + **模块: 句子构建流程**
  + **作用:** 产生最终的沟通性输出。
  + **过程:** 高涨的愉悦度(P)使得积极、肯定的词元权重升高。
  + **输出:** 想法流组织成语言回复：“已经把温度调高了，应该很快就会暖和起来啦！” 这是一个**外在行动**（沟通类），被发送给用户。

**异步模式总结:** 这个过程是并发的、自主的。输入、思考、情绪、行动四个线程同时工作，相互影响。

* **输入线程**处理了初始刺激和行动反馈。
* **思维线程**产生了内部联想并最终组织了对外回复。
* **行动线程**监控并执行了关键的环境操作。
* **情绪线程**为整个过程提供了背景色和最终奖励信号。  
  智能体完全依靠自身内部的知识和情绪网络，完成了从感知、联想、决策、行动到学习的完整闭环，**最终有效地满足了用户的预期（解决冷的问题并进行沟通）**，并在这个过程中变得更加“自信”。

**5.** **实现细节与优化**

**5.1 图数据库优化**

**索引策略**：

* 为content字段建立全文索引，支持快速文本匹配。
* 为origin字段建立哈希索引，优化起源相关查询。
* 为last\_activated字段建立B+树索引，支持时间范围查询。
* 为emotion\_ema字段建立向量索引，支持情绪相似度计算。
* **为“构成”链接和“角色”链接建立特殊索引**，加速词元融合器的查询。

**查询优化**：

* 使用多级缓存机制，热数据常驻内存，冷数据持久化存储。
* 基于locality of reference原理，近期活跃数据优先缓存。
* 实现查询结果缓存，减少重复计算。

**存储优化**：

* 采用列式存储格式，提高压缩率和查询效率。
* 实现增量更新机制，减少写入开销。
* 支持数据分片，实现水平扩展。

**5.2 情绪计算优化**

**PAD值量化**：

* 采用16位浮点数存储，平衡精度和效率。
* 基于数值分析中的精度-效率权衡，在保持精度的同时减少存储开销。

**情绪更新流水线**：

* 使用SIMD指令并行计算P、A、D三个维度的更新。
* 基于单指令多数据流技术，提高计算效率。
* 实现流水线处理，重叠计算和内存访问。

**情绪分类优化**：

* 使用查找表加速PAD到情绪标签的映射。
* 实现近似计算，在保持准确性的前提下提高速度。
* 支持批量处理，提高整体吞吐量。

**5.3 内存管理**

**激活池推荐容量限制(可根据实际情况调整)**：

* 显性池：最大500个元素（基于Miller的7±2理论，并扩展以支持更复杂情境）。
* 隐性池：最大10000个元素（为显性池20倍，基于潜意识处理容量估计）。
* 注意池：最大50个元素（基于注意资源有限性理论）。
* 行动池：最大20个元素（基于行为选择研究）。

**淘汰策略**：

* LRU（最近最少使用）淘汰机制。
* 基于权重、起源和新近度的混合淘汰策略。
* 支持优先级保留，重要元素（如高期待值的注意节点）获得更高保留优先级。

**内存分配优化**：

* 使用对象池技术减少内存分配开销。
* 实现自定义内存分配器，优化内存访问模式。
* 支持内存压缩，减少内存占用。

**结论**

人工智能直觉架构（AII）代表了一种全新的类人人工智能设计思路，通过将认知科学、语言学与计算机科学深度融合，创造了一个能够模拟人类思维、情感和直觉过程的智能系统。该架构的核心创新在于：

1. **词元融合与直觉式处理**：引入了词元融合器，实现了对语言的深度理解和抽象，形成了直觉式认知的基础。
2. **完整的认知情感整合**：实现了情绪、认知、注意和行动的有机统一，情绪是驱动的核心而非附加功能。
3. **情景记忆与语义记忆分离**：通过感受记忆库和认知数据库的分离，实现真正的人格化和个性化。
4. **内在体验模拟**：通过感受器系统实现时间和情绪的内在体验，为意识提供内容。
5. **自主性发展与双模式学习**：通过双模式设计支持从辅导学习到完全自主的完整发展过程。

**理论贡献**：

1. **认知架构创新**：提出了四池模型、双重起源系统和词元融合理论，为理解意识-潜意识交互、概念形成和直觉提供了新视角。
2. **情绪计算模型**：发展了基于PAD模型的动态情绪计算框架，实现了情绪对认知的量化调制和影响模拟。
3. **记忆与学习系统设计**：创新性地将情景记忆与语义记忆分离，并引入了动态连接强度和可塑遗忘机制，更符合人类记忆的实际结构。

**实践意义**：

1. **技术实现可行性**：提供了详细的技术实现方案和参数设置，具有良好的工程可行性。
2. **应用前景广阔**：在心理健康陪伴、个性化教育、高级人机交互、游戏NPC等领域都有重要的应用价值。
3. **推动AI发展**：为开发真正具有理解、学习和自主能力的类人人工智能系统提供了新的思路和方法论。

**未来工作**：  
尽管AII架构在设计上力求完善，但仍需要在以下方面继续深入研究：

1. **参数优化与大规模实验**：需要通过大量实验优化各项参数设置，验证架构在各种场景下的表现。
2. **扩展性与性能研究**：需要研究架构在大规模知识库和海量记忆下的表现，以及进一步的性能优化。
3. **多模态深度整合**：需要进一步完善对视觉、听觉等模态信息的深度处理和整合，而不仅仅是特征字符串。
4. **安全与伦理规范**：需要建立完善的使用伦理规范、价值对齐机制和安全保障机制，确保技术向善。

我们相信，人工智能直觉架构（AII）将为人工智能的发展开辟新的道路，推动AI向更加智能、更加人性化的方向发展。期待与各界专家共同努力，不断完善和发展这一架构，为人工智能的未来做出贡献。

**参考文献**：

1. Baddeley, A. D. (2000). The episodic buffer: a new component of working memory?
2. Tulving, E. (2002). Episodic memory: from mind to brain.
3. Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament.
4. Anderson, J. R. (2007). How can the human mind occur in the physical universe?
5. Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing.
6. Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information.
7. McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings.

**作者**: 银子  
**日期**: 2025年9月4日