飞行模型：智能与意识基本原理的新探索

刘锋[1,3*] 吕本富[1,2] 刘颖[1,2]

1.中国科学院大学网络经济与知识管理研究中心
2. 中国科学院大学经济与管理学院
3.中科数字大脑研究院
*Email:liufeng@idbr-cst.org

摘要：

人工智能的未来发展和自我意识问题引发了广泛关注，构建智能与意识的基本原理成为重要的科学挑战。当前这一挑战面临五个关键问题：1）产生智能与意识的主体统一结构，2）目标和意义，3）驱动力，4）智能与意识的关系，5）不同意识类型的区别。本文围绕这五个问题展开研究，旨在建立智能与意识的基本原理框架。首先，我们对冯·诺依曼架构进行扩展，作为一条公理建立标准智能体模型，提出任何系统（智能体）具备知识输入、输出、存储、创造及控制五项能力；然后，通过对五项能力分别取值 0 或无穷大，形成智能体的演化边界，即全知全能智能体（Ω点）和绝对零智能体（α点）；接着，从智能体向边界的演化出发，推导出自然界存在两种智能力，即Ω引力和α引力；进一步，在飞行原理的启发下，我们将标准智能体模型、两个演化边界和两种智能力进行组合，建立了智能与意识的“飞行模型（FM）”，定义智能为智能体在Ω引力和α引力作用下，综合运用五项基本能力向Ω点或α点演化的能力；定义意识为智能体在两种智能力作用下对智能的运用进行控制的能力；最后，依据控制与被控制主体的关系，我们将意识细分为自我意识、他者意识、混合意识及无意识。本文基于 FM 理论框架对不同种类的智能体智能水平和自我意识进行了评估。结果显示，当前人工智能系统的智能接近成人水平，但尚未表现出自我意识。FM 为构建智能与意识理论体系奠定了基础，为判断 AGI 实现时间和 AI 自我意识问题提供了理论依据，并提出了新的科学探索方向，即两种智能力的如何作用与系统（智能体）的机制问题以及与物理学四种基本作用力的关系问题。

关键词：智能、意识、标准智能体模型、飞行模型、智能力

本文英文预印版地址：http://dx.doi.org/10.13140/RG.2.2.24518.28484

1.引言

构建智能与意识的基本原理是人工智能与智能科学领域的核心任务之一。21 世纪初以来，人工智能技术迅速发展，特别是大语言模型如 ChatGPT 的兴起，已经激发了公众和学术界对于通用智能实现时间、人工智能是否能产生自我意识，以及未来研究方向的广泛讨论和争议。这些问题不仅引起了广泛关注，也对人工智能的未来发展产生了深远影响[1]。因此，深入探索并确立智能和意识的基本原理，已成为当前研究的重要课题。

学术界已对智能与意识的基本原理进行了深入研究，提出了多种定义，从不

同角度为理解这两个概念提供了框架[2,3]。关于什么是智能，Sternberg 定义智能为个体适应环境的能力,强调了智能在个体与环境交互中的适应性作用[4]。Russell 则从计算的角度,将智能定义为在有限资源约束下实现目标最优化的能力[5]。Stone 进一步指出,智能系统应具备在多样环境中稳健工作的能力,即使在信息不完全的情况下,也能最大化地实现既定目标[6]。Legg 的定义则强调了智能在广泛环境中实现目标的普适性[7]。

当前关于意识的理论主要包括全局工作空间理论[8]、整合信息理论[9]、高阶理论[10]及复馈/预测处理理论[11]等。最新研究进展包括 Graziano 提出的注意架构理论[12]以及 Friston 等人提出的自由能原理[13]等。

现有理论从不同视角阐释了智能和意识的机制原理和神经基础，但仍存在诸多局限。主要问题包括：未详细阐明产生智能和意识的*智能体（系统、主体）*功能结构；缺乏对智能和意识产生的内在动力和目的的描述；较少探讨意识与智能的区别和联系，往往混淆了二者的概念边界；对自我意识、他者意识、混合意识和无意识的区分亦缺乏深入分析。这些局限性引发了不同定义之间的分歧，阻碍了系统性智能与意识理论体系的形成。为此，我们提出建立智能与意识基本原理框架，需充分回答以下五个关键问题：

1.产生智能与意识的智能体（系统）是否具有统一的功能结构？
2.智能体产生智能与意识的最终目标是什么？
3.驱动智能体产生智能与意识的动力机制是什么？
4.如何界定智能与意识的关系，并明确其概念边界？
5.如何区分自我意识、他者意识、混合意识与无意识？

自 2014 年以来，我们针对这五个核心问题进行了系列探索性研究[14, 15, 16]。研究发现，解决统一的智能体功能结构问题至关重要，其他问题的解决都需要围绕这一核心问题展开。本文的突破点在于建立了统一的智能体功能结构模型，并以此为基础推导出第二至第五个问题的答案，从而初步形成了智能和意识基本原理的完整体系。

基于该理论框架，我们提出了评估智能体智能水平和自我意识的标准，并对包括本文作者开发的轻量级 AI 系统 Angry Elf[17]、不同年龄段的人类、大模型和智能设备在内的 23 个智能体进行了测试和检验。

本文研究的重要意义在于建立了相对完整的智能与意识基础理论体系，为判断通用人工智能的实现及其是否能够产生自我意识提供了理论依据，并提出了新的科学探索方向，即研究两种智能力如何作用于智能体产生智能和意识。其与物理学中的四大基本作用力是什么关系，这些问题的研究将对当前物理学和智能科学的发展都将具有积极的推动作用。

在本文中，我们将智能体等同于系统或主体。为简化论述，我们将智能体处理的三种元素——“知识”、“信息”和“数据”统一称为“知识”。

2. 理论体系的形成过程

为了建立具有统一功能结构的智能体模型，我们特别关注了冯·诺依曼架构，该架构为计算机和人工智能系统提供了统一的功能框架[18]。然而，此架构未涵盖人类等生命系统。我们的研究表明，人类不仅具备冯·诺依曼架构的特征，还展现出显著的知识创造能力，例如牛顿提出万有引力，爱因斯坦提出相对论。我们通过将知识创造功能加入到冯·诺依曼架构中，提出了标准智能体模型。该模型认为，任何系统（或智能体）均具备知识输入(I)、输出(O)、存储(S)、创造(C)以及对这四种能力进行控制(Con)的能力。如图 1 所示。

标准智能体模型本质上是提出了一条基本公理，是推导和构建智能与意识理论体系的基石。相较于冯·诺依曼架构，此模型的主要有两项改进：首先，模型整合了运算与存储能力，创建了动态存储模块，实现了存储与计算的一体化；其次，引入的知识创造模块能够体现生物在适应环境和改造自然过程中的创新与发明能力。这些改进不仅使得标准智能体模型适用于计算机和智能设备，还将其覆盖范围扩展至人类及其他智能系统。

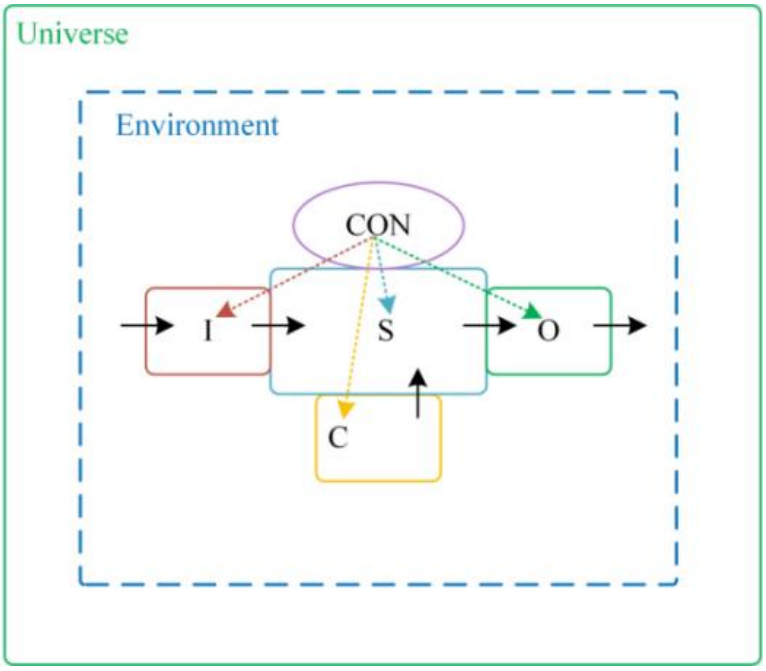

图 1 标准智能体模型

自 2020 年起，在对标准智能体模型的进一步分析中，我们发现智能体存在两种极端状态：一种是智能体在五种能力上的表现均为零（称为绝对零智能体，简称 α 点），另一种则是这些能力均达到无限大（称为全知全能智能体，简称 Ω 点）。理论上，智能体向这两个极端状态的演化需要动力驱动。基于此，我们推导出自然界中应存在两种智能力，分别命名为 α 引力和 Ω 引力，不同于物理学中的四大基本作用力，它们作用于系统或智能体产生智能和意识，并推动其向 α 点或 Ω 点演化[16]。

人工智能领域的研究者已经开始关注到飞行原理与智能原理之间的联系。例如,Peter Norvig 和 Stuart Russell 指出,研究智能的基本原理比复制大脑样本更为重要。正如莱特兄弟等人在停止模仿鸟类、开始理解空气动力学原理后才成

功实现"人工飞行"一样,智能原理的建立也应该遵循类似的思路[19]。钟义信同样提出,飞机设计并非严格模仿鸟类的具体结构,而是基于对空气动力学原理的理解;智能原理的建立也应该遵循这一思路,而非简单地模仿生物智能的具体结构[20]。

标准智能体模型及其推论显著增强了智能原理与飞行原理之间的相互联系。飞行一般被定义为飞行物在重力和升力的作用下在地平线和卡门线运动的能力或现象[21],由此在飞行现象的启发下,我们将标准智能体模型与两个演化边界(α点和Ω点),以及两种智能力(α引力和Ω引力)相结合,构建了智能体演化的动力学模型,飞行模型(Flight Model, FM),如图2所示。

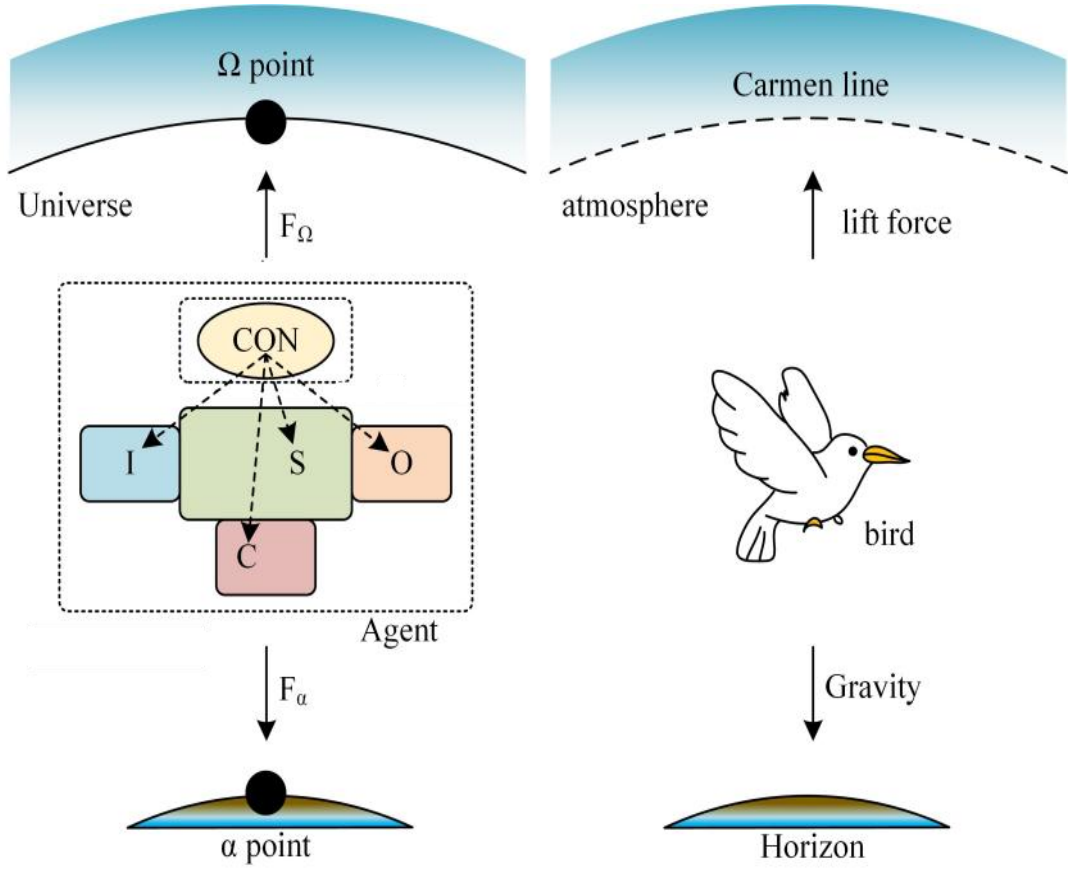

图2 智能体演化与飞行物飞行对比图

飞行模型作为构建智能与意识基本原理的重要基础之一,借鉴飞行原理对飞行的描述,我们提出:智能是智能体在在Ω引力和α引力的作用下,综合运用5种基本能力实现向Ω点或α点演化的能力

通过对"飞行模型(FM)"进一步分析,我们发现在智能体的五项能力中,前四种能力偏重知识的处理过程,因此可以视为智能体的基础智能,而控制能力掌控了前四种能力的运用,也依赖于前四种能力的存在。因此控制能力可以视为是智能体的高阶智能。

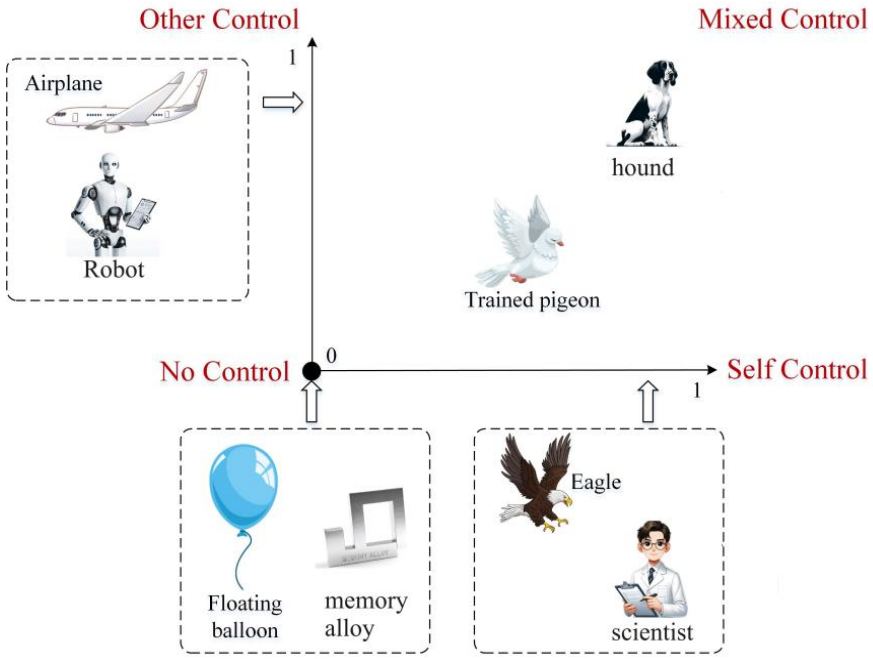

图 3 飞行和智能活动中的四种控制类型

如图 3 所示，无论是在飞行现象中，还在智能活动中，根据根据控制主体和被控制主体的不同，控制能力又可以分为自我控制、他者控制、混合控制和无控制四种类型，这种分类正好与意识在应用中常见的四种类型:自我意识、他者意识、混合意识和无意识[22]一一对应。这启发我们提出意识的本质是智能体对基础智能的运用进行控制的能力。

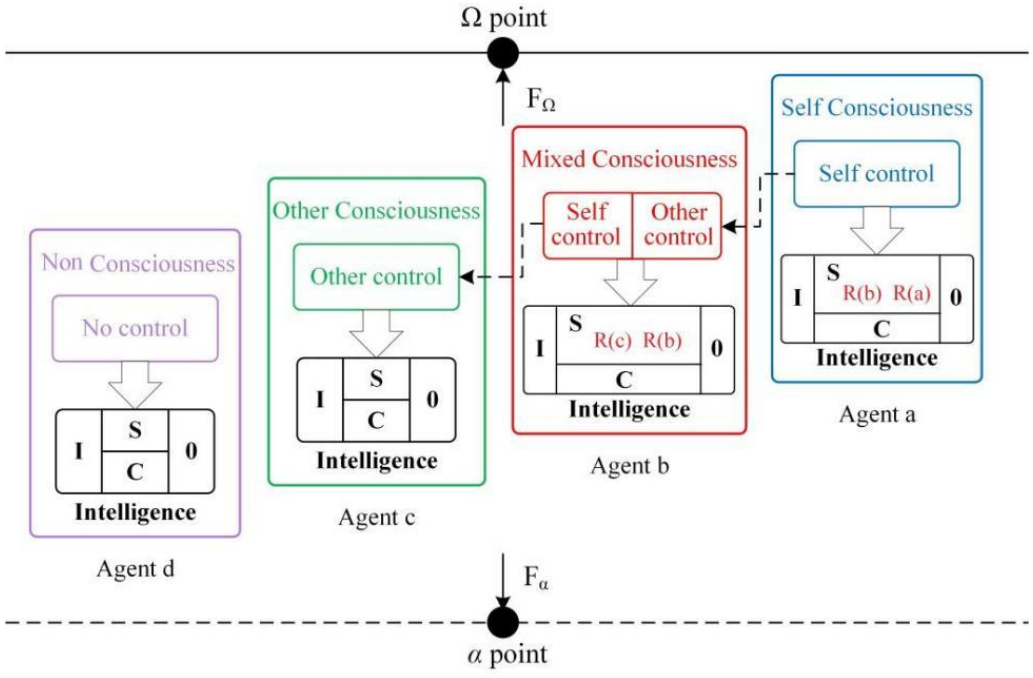

图 4 意识的四种类型

为了准确定义和分析意识的四种类型,"自我"和"他者"的本质内涵需要被理清。根据自我和他者的一般认知，基于标志智能体模型框架，我们提出："自我"是智能体 a 对自身进行认知后形成的内部知识集，记为$R_a(a)$，而"他者"是智能体 a 对智能体 b 认知后，在智能体 a 内部形成的知识集,记为$R_a(b)$，当"自我"和"他者"与不同智能体的控制功能结合后,便形成了意识的四种类型,如图 4 所示。

基于智能与意识的飞行模型（FM），我们提出了智能和意识的定义，并区分了意识的四种类型，从而构建了初步完整的智能与意识理论体系。

3. 理论体系的构建

标准智能体模型作为一条基本公理，推导和构建出整个智能与意识理论体系，包括标准智能体模型、绝对零智能体（α 点）、有限智能体、全知全能智能体（Ω 点）、α 引力、Ω 引力、智能、基础智能、高阶智能、自我、他者、意识、自我意识、他者意识、混合意识、无意识以及飞行模型等 17 个组成部分。它们的逻辑关系如图 5 所示。

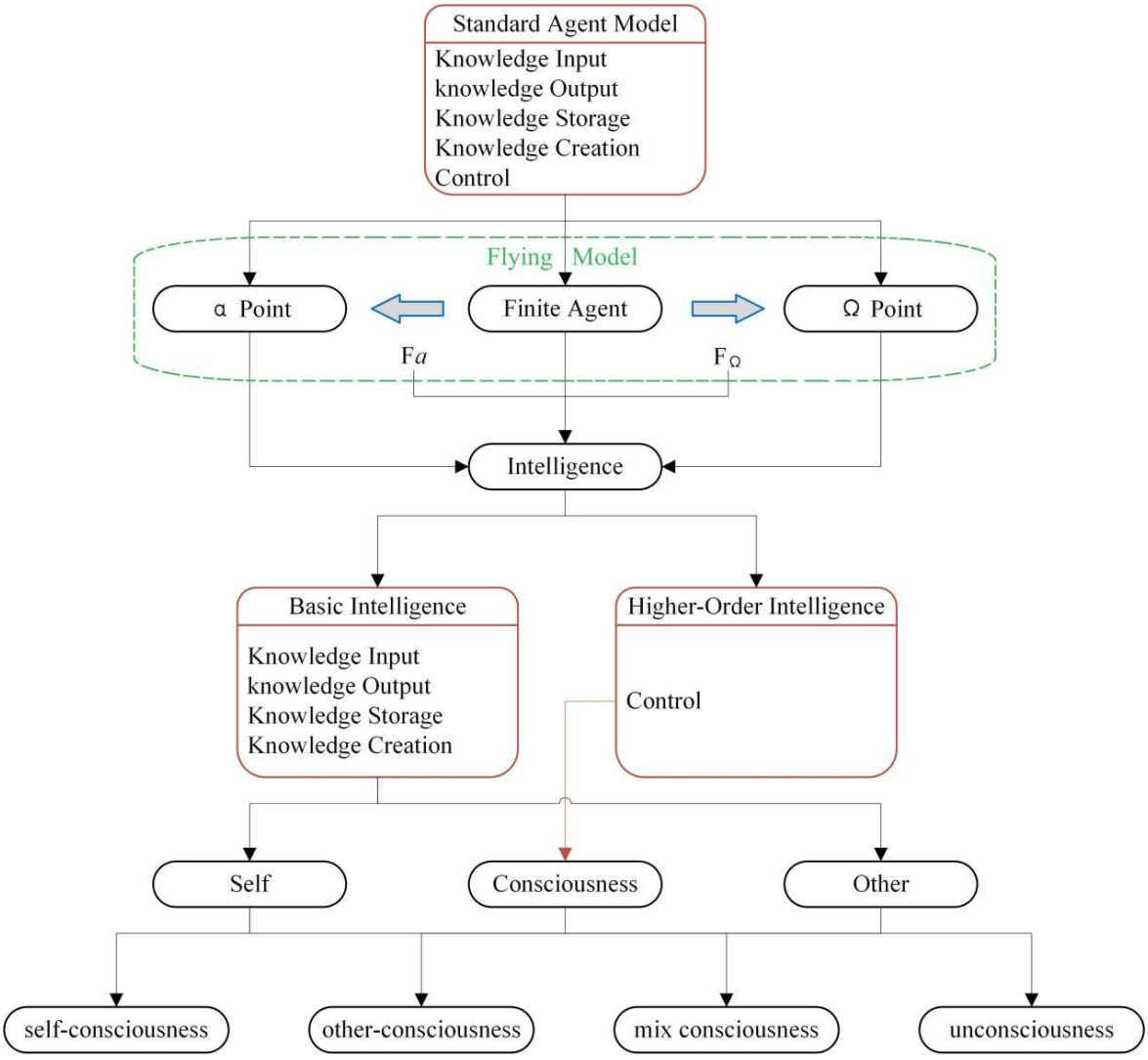

图 5 智能与意识基础理论结构框架图

3.1 基础公理

标准智能体模型是基于冯·诺依曼架构，对人类、自然界生物以及人工智能系统中的智能特性进行了抽象化和提炼后形成的通用的智能系统框架。

定义 1:标准智能体模型

任何智能体（或系统）都具备处理知识的 5 种基本能力，它们分别是知识输入(Input, I)、知识输出(Output, O)、知识存储(Storage, S)和知识创造(Creation, C)能力以及对这四种能力的使用进行控制的能力(Control, Con)。

标准智能体模型可以用以下数学表达式进行形式化描述，公式中的各个元素具体含义见表 1 的说明。:

$$Agent\ a = (Con_a; \{I_a(k), O_a(k), S_a(k), C_a(k)\})$$

表 1 标准智能体模型涉及各符号的含义

符号	含义说明	约束条件
a	表示一个系统或智能体。	
K_x	表示 x 所能处理或包含的知识集合 K	
U	表示整个宇宙，宇宙 U 所能处理或包含的知识以 K_u 表示。	
E_a	表示智能体 a 的环境，包括智能体 a 通过知识的输入和输出能力能够认知和影响的全部智能体。如果智能体a能够对本身进行认知，a 也属于其环境的组成部分，E_a 中包含的知识用 K_{Ea} 表示，	$K_{Ea} \subseteq K_u$
$I_a(K)$	表示智能体a从宇宙U 或环境E_a中识别和获取知识的能力	$I_a \in [0,1]$,（1 代表∞）
$S_a(K)$	表示智能体a将输入的知识、创造的知识，已经存储的知识进行动态存储的能力。	$S_a \in [0,1]$,（1 代表∞）
$C_a(K)$	表示智能体 a 从存储的知识中发现和创造新知识的能力,C_a 的取值在 0 到无穷大之间，	$C_a \in [0,1]$,（1 代表∞）
$O_a(K)$	表示智能体 a 根据存储的知识通过输出能力实现对宇宙 U 或环境 E_a 的影响和改造。	$O_a \in [0,1]$,（1 代表∞）
$Con_a(K)$	表示智能体 a 对知识的输入、输出、存储和创造能力的使用进行控制的能力	$Con_a \in [0,1]$,（1 代表∞）
K_a	表示智能体 a 所能处理的全部知识元素，由其知识的输入、输出、存储和创造等四种能力所能处理的知识集的并集所构成。	$K_a = K_{Ia} \cup K_{Oa} \cup K_{Sa} \cup K_{Ca}$

3.2 动力学模型

3.2.1 智能体的三种类型

依据标准智能体模型的定义，当一个系统处理知识相关的五种能力均降至零，将形成第一种极端状态的智能体。对于这种状态的智能体，我们提出如下的定义：

定义 2 绝对 0 智能体

如果一个智能体的知识输入、知识输出、知识存储、知识创造及控制能力全部为 0，那么这个系统被定义为绝对 0 智能体，简称为α点（α point）. 绝对 0 智能体的数学数学描述为:

$$\alpha\ point = (0;\{0,0,0,0\})$$

当一个智能体处理知识的五种能力的强度达到无限大时，便构成了另一种极端状态的智能体，对用这种状态的智能体，我们定义如下：

定义 3 全知全能智能体

如果一个智能体的知识输入、知识输出、知识存储、知识创造及控制能力均为无穷大，那么这个系统被定义为全知全能智能体,简称Ω点(Ω point).全知全能智能体的数学描述为:

$$\Omega\ point = (1;\{1,1,1,1\}) \quad (1\ 代表\infty)$$

除了绝对 0 智能体和全知全能智能体，还存在大量不处于两种极端状态的智能体，对于这种类型的智能体定义如下：

定义 4 有限智能体

如果一个智能体的知识输入、知识输出、知识存储、知识创造及控制能力不全为 0，也不全为无穷大，那么这个系统被定义为有限智能体，简称 FA。有限智能体的数学描述为:

$$FA=(Con_a;\{I_a(k),O_a(k),S_a(k),C_a(k)\}),,$$

$\neg(\forall x \in \{I_a, O_a, S_a, C_a, Con_a\},x=0)\wedge\neg(\forall x \in \{I_a, O_a, S_a, C_a, Con_a\},x=1)$,（1 represents ∞）

3.2.2 智能体演化的两种智能力

在智能体的演化过程中,α点和Ω点分别代表了演化的方向目标和边界条件。理论上,驱动智能体向这两个极端状态演化的过程必须由特定的动力因素所推动。由此定义了α引力和Ω引力等两种作用与智能体的智能力，如图 6 所示

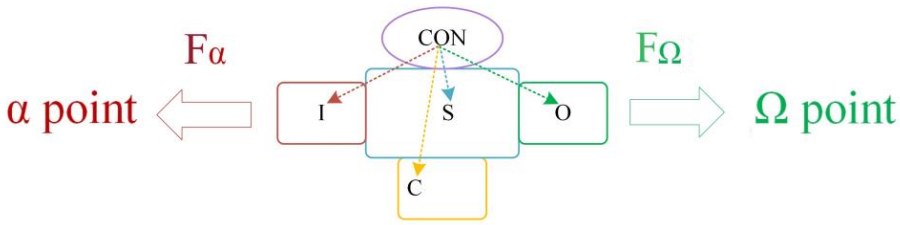

图 6 智能体演化动力示意图

定义 5 α引力

α引力是一种直接或间接作用与任何一个系统或智能体的动力，驱动智能体的五种能力不断衰减，并最终收敛到α点。α引力也表示为F_α。

数学表达为：$F_\alpha: a \to \alpha\ point$　　（a 为智能体）

定义 6　Ω引力

Ω引力是一种直接或间接作用与任何一个系统或智能体的动力，驱动智能体的五种能力不断增强，并最终演化到Ω点。Ω引力也表示为F_Ω。

数学表达为：$F_\Omega: a \to \Omega\ point$　　（a 为智能体）

3.2.3 飞行模型

参考飞行原理，将标准智能体模型和推导出来的两个极端状态α点和Ω点，和两个智能力α引力和Ω引力进行组合，就形成了智能与意识的动力学模型-飞行模型，定义如下：

定义 7 飞行模型

智能体在α引力和Ω引力的直接或间接共同作用下，在α点与Ω点之间动态演化的过程。简写为 FM

a 为任意一个智能体，飞行模型的数学描述如下：

$$\begin{cases} a \to \alpha\ point & if\ f_\alpha > f_\Omega \\ a \to \Omega\ point & if\ f_\alpha < f_\Omega \\ a \to a & if\ f_\alpha = f_\Omega \end{cases}$$

或简化为：

$$\alpha\ point \xleftarrow{f_\alpha} a \xrightarrow{f_\Omega} \Omega\ point$$

3.3 智能的理论体系

根据“飞行模型”，任何一个系统或智能体均在 α 点（起始点）与 Ω 点（终点）之间进行动态演化。因此我们参考飞行原理，提出如下智能定义：

定义 8 智能

智能是智能体在 Ω 引力和 α 引力的作用下,综合运用 5 种基本能力实现向 Ω 点或 α 点演化的能力。这五项能力分别是：知识的输入、输出、存储、创造能力以及对上述四项能力的进行控制的能力。（智能的原理图示参考图 7）。

智能的数学描述为：

$$\text{Intelligence}(a,k) = Con_a\big(I_a(k), O_a(k), S_a(k), C_a(k)\big)\ ,$$

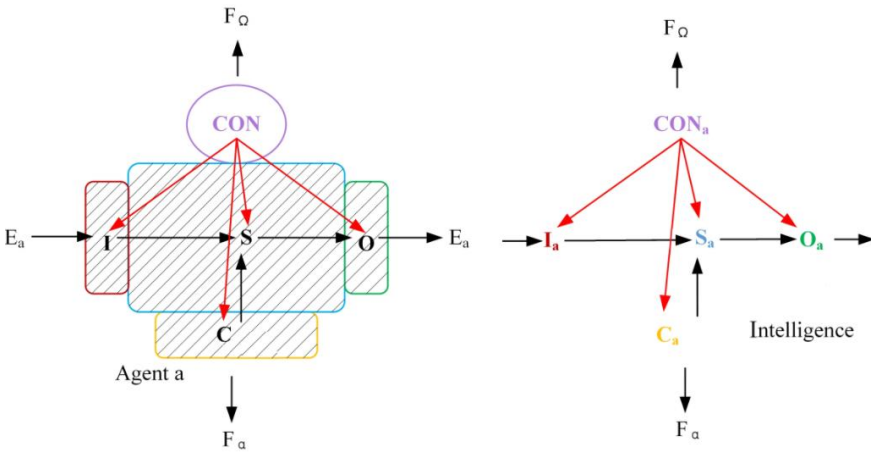

图 7 智能的原理示意图

根据构成智能的五种能力的不同特征，提出基础智能和高阶智能的定义，区分它们将为分析通用人工智能和提出意识的定义做出准备。

定义 9 基础智能

在智能体（或系统）的 5 项能力中，知识的输入、输出、存储和创造等四项能力通过协同工作，实现知识在系统内外部的处理与流动，具有更强的属性一致性，这四项能力被分类为智能体的基础智能（原理图示参见图 8）。

数学表达为：$Basic\ Intelligence(a,k) = \left(I_a(k), O_a(k), S_a(k), C_a(k)\right)$，a 代表智能体

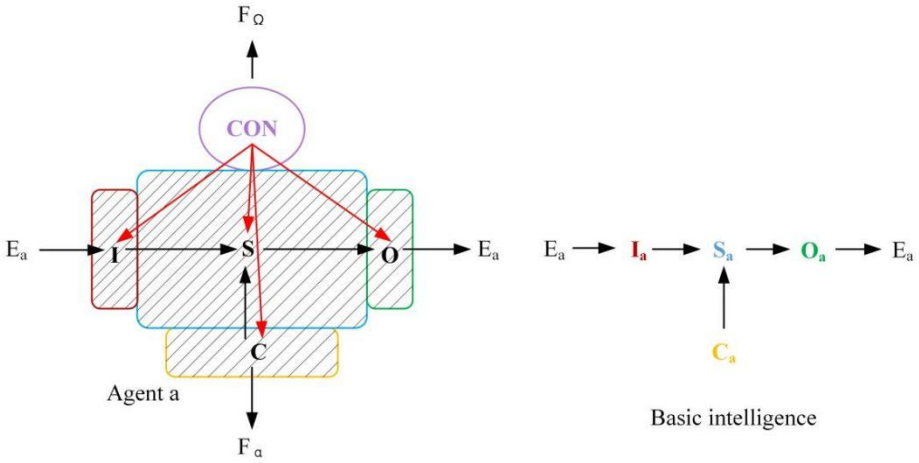

图 8 基础智能原理示意图

定义 10 高阶智能

在智能体（系统）的 5 项基本能力中，控制能力是建立在对其他 4 种能力的基础之上，用于管理、调度这四种能力的运用。因此，控制能力被定义为智能体的高阶智能。（原理图示参见图 9）。

数学表达为：$Higher-Order\ \ Intelligence(a,k) = Con_a$，$a$ *代表智能体*

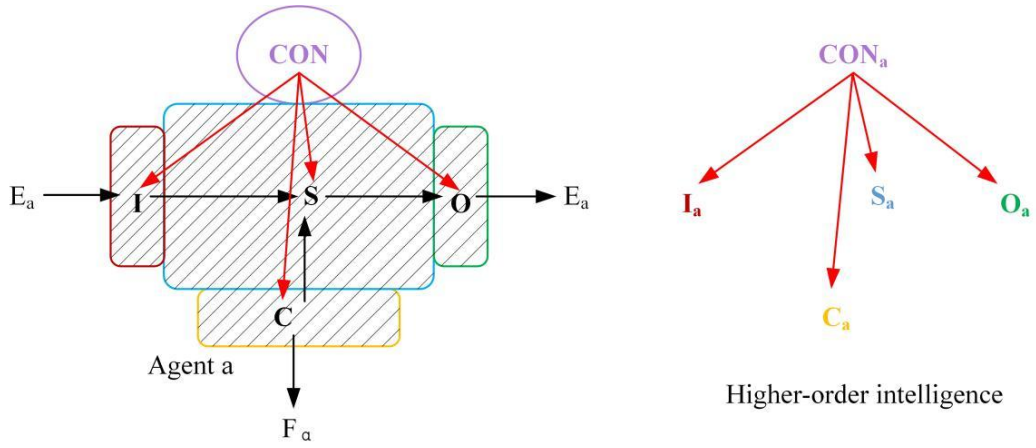

图 9　高阶智能原理示意图

3.4 意识的理论体系

根据第 2 节的分析，意识的本质是智能体对基础智能使用的控制能力，是智能的重要组成部分,属于高阶智能，定义如下:

定义 11 意识

意识是指系统(或智能体)在Ω引力和α引力的直接或间接作用下,对知识的输入、输出、存储和创造等基础智能的运用进行控制的能力。意识的作用是帮助系统优化其向Ω点或α点演化的路径(原理示意图见图 10)。

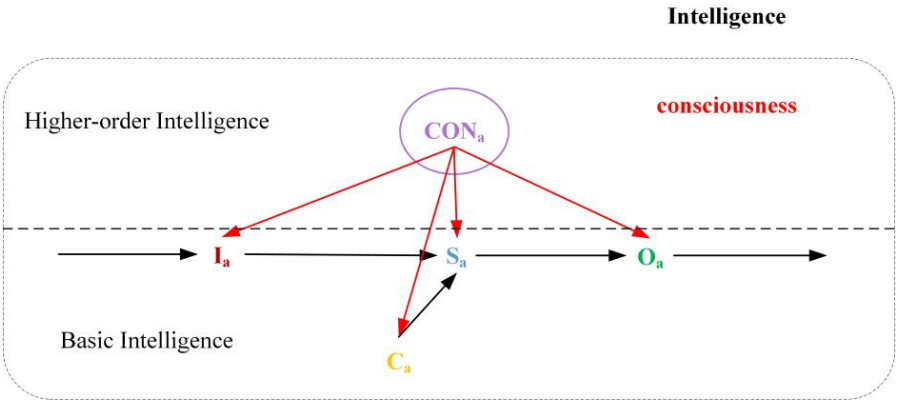

图 10 “意识“原理示意图

这里的意识在内涵上等价与智能力驱动下的控制能力，英文缩写为Con_X^Y，其中 X 为控制主体，Y 为被控制对象。取值范围$Con_X^Y \in [0,1]$.其中 0 代表 X 对 Y 完全没有控制能力，而 1 则表示 X 对 Y 具有无穷大（完全）的控制能力。

意识可分为四种类型:自我意识、他者意识、混合意识和无意识。为准确定义和分析这四种意识类型,首先需厘清"自我"和"他者"的本质内涵。

定义 12 自我

自我是指智能体 a 在 α 引力与 Ω 引力的作用下，通过基础智能的运用，对自身认知后形成的知识集合，记为$R_a(a)$。智能体 a 存在“自我”的条件为：$R_a(a) \neq \emptyset$。

智能体 a 的自我知识集$R_a(a)$作为其内部知识集K_a的子集，同样受到Ω引力或α引力的影响，导致其包含的知识元素数量动态变化。自我的形成过程如图 11 所示。

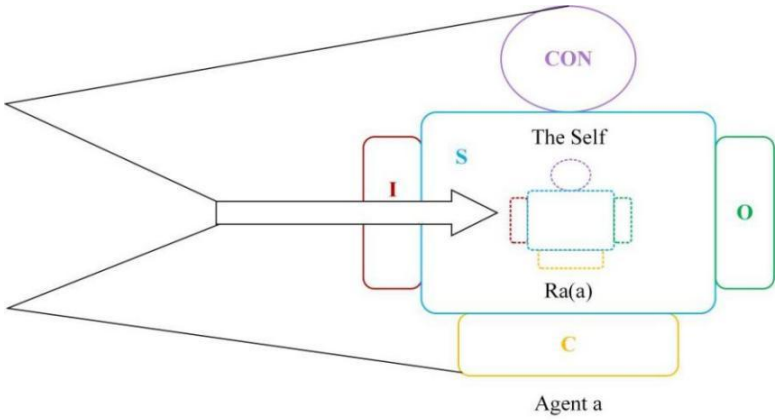

图 11 “自我”形成过程示意图

定义 13 他者

他者是智能体 a 在 α 引力与 Ω 引力的直接或间接作用下，通过基础智能的运用，对另一个智能体 b 认知后形成的知识集的总称，用$R_a(b)$表示。智能体 a 存在关于智能体 b 的他者知识集的条件是$R_a(b)\neq\emptyset$。

智能体 a 的他者知识集$R_a(b)$作为其内部知识集K_a的子集，同样会受到 Ω 引力或 α 引力的作用，导致其包含元素数量不断动态变化. “他者”的形成过程参见图 12。

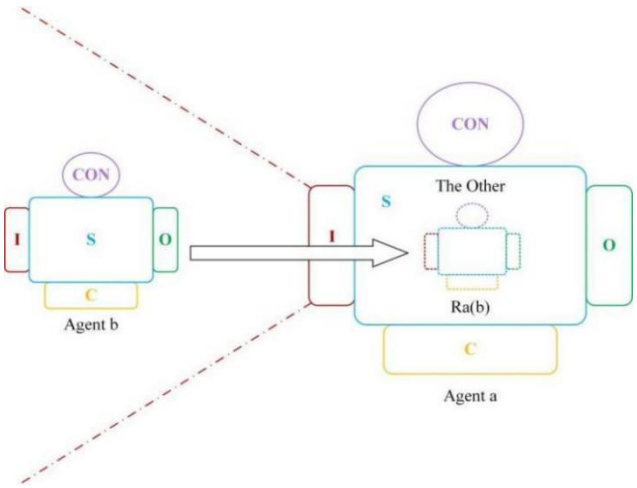

图 12 “他者”知识集的形成示意图

结合“自我”和“他者”的概念与意识的定义，可以形成四种意识类型：自我意识、他者意识、混合意识和无意识。定义如下：

定义 14 自我意识

自我意识是指任何一个系统或智能体（记为 a），在 α 引力和 Ω 引力的直接作用下，利用其“自我”知识集对自身基础智能的使用进行控制的能力，其作

用是帮助智能体 a 优化其向 α 点或 Ω 点的演化路径（原理示意图参见图 13）。

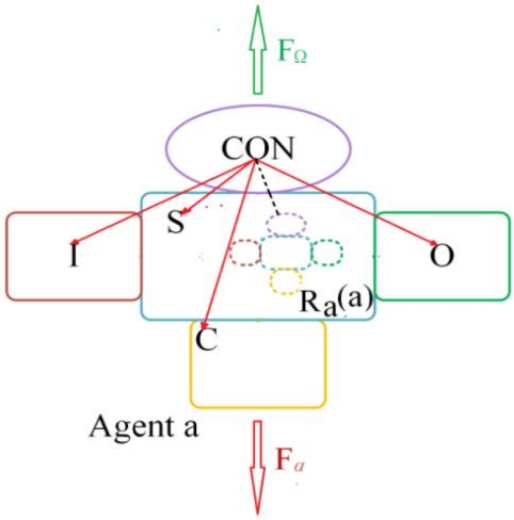

图 13 “自我意识”原理示意图

根据自我意识的定义，一个智能体 a 存在自我意识需要全部满足如下条件：

1）$Con_a a > 0$

2）$R_a(a) \neq \emptyset$

3）智能体 a 产生自我（知识集）、实现自我控制、开展智能活动和不断增强智能的动力直接来自 α 引力和 Ω 引力。

定义 15 他者意识

当智能体 a 能被智能体 b 认知，并在智能体 b 部形成关于智能体 a 的他者（知识集），且智能体 a 的基础智能被智能体 b 基于他者（知识集）进行控制时，称智能体 a 具有他者意识。(原理示意参见图 14)。

根据他者意识的定义，a 是任意一个智能体，b 是不同于 a 的另外一个智能体，智能体 a 存在他者意识需要全部满足如下条件：

1）$Con_b a > 0$ (智能体 b 接管智能体 a 的控制功能）。

2）$R_b(a) \neq \emptyset$（智能体 b 对智能体 a 认知形成的知识集不为空）

3）智能体 a 产生自我（知识集），实现自我控制，开展智能活动和不断增强智能的动力来自与智能体 b

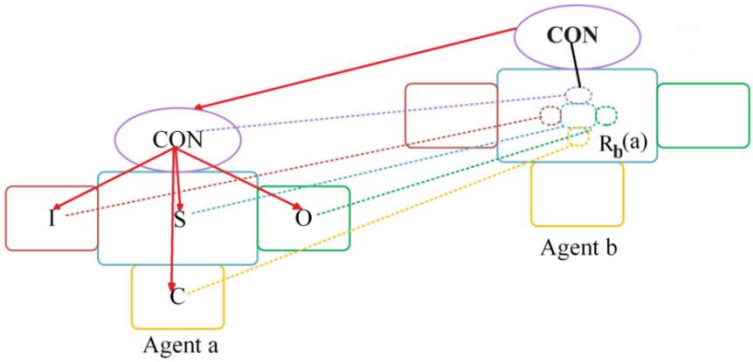

图 14 他者意识原理示意图

定义 16　混合意识

混合意识是指智能体 a 同时受自我意识和他者意识的影响,共同实现对其自身智能运用的控制。(参见图 15)。

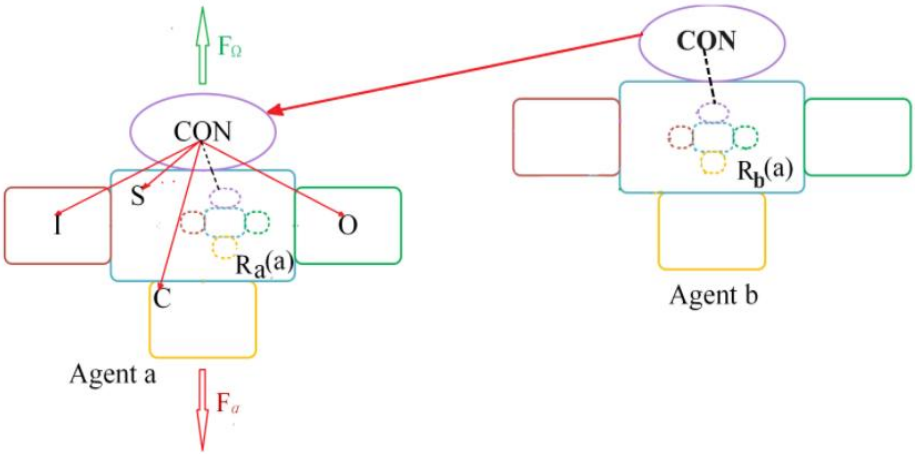

图 15 “混合意识”原理示意图

根据混合意识的定义，a 是一个智能体，b 是不同于 a 的另外一个智能体，智能体 a 存在混合意识需要全部满足如下条件：

1） $Con_b a > 0 \ \wedge \ Con_a a > 0$

2） $R_a(a) \neq \emptyset \ \wedge \ R_b(a) \neq \emptyset$

3）智能体 a 产生自我（知识集）、实现自我控制、开展智能活动和不断增强智能的动力既包含 Ω 引力和 α 引力,也包含智能体 b 的推动

定义 17 无意识

无意识是指对于智能体 a,任何智能体（包括它自身）都不具备对其基础智能的运用进行控制的能力(参见图 16)。

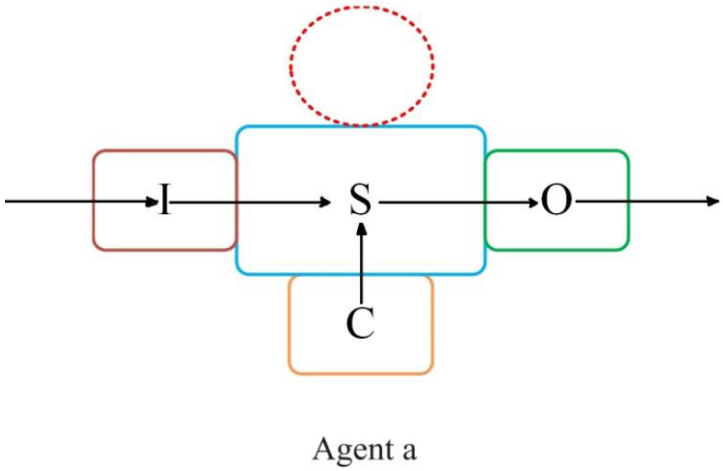

图 16 “无意识”原理示意图

根据无意识的定义，a 是一个智能体，b 是不同于 a 的另外任意一个智能体，智能体 a 具有无意识的条件为：$Con_b a = 0 \ \wedge \ Con_a a = 0$

4. 实验设计与结果

4.1. 基础智能水平测试与结果

在本文研究中，智能是智能体对 5 种能力的综合运用能力，为了减少实验复杂度，我们将主要围绕智能体的基础智能设计实验方法和开展实验。基础智能是由智能体的知识输入、输出、存储和创造能力构成。根据这四种能力各自的特点，我们又可以将其拆分形成二级能力[7]，同时通过德尔菲法方法，赋予一级能力和二级子能力相应的权重。形成如表 2 所示的智能体（基础）智能水平测试量表。

表 2　（基础）智能测试量表

一级指标	二级指标	权重
知识输入能力	识别文字的能力	3%
	识别声音的能力	3%
	识别图片的能力	4%
知识存贮能力	常识	6%
	信息保存	6%
	自我认知	7%
	计算	7%
知识创造能力	猜测能力	5%
	提问能力	10%
	创作能力	9%
	设定目标	12%
	类比能力	8%
	发现规律	10%
知识输出能力	用文字表达的能力	3%
	用声音表达的能力	3%
	用图形表达的能力	4%

智能体智能水平（智商）的数学计算公式如下：

公式 3 $$Q = f(M) = f(I, O, S, C) = a * f(I) + b * f(O) + c * f(S) + d * f(C)$$
$$a + b + c + d = 100\%$$

在公式 3 中，M 表示一个具备基础智能的系统，Q 是系统的智能水平，f 是智能水平的函数，I、Q、S、C 是系统的四种能力，a、b、c、d 代表四种能力的权重：

根据表 2 所示的测试量表，我们为每个二级能力设计 20 道测试题目，形成总共 320 道题目的测试题库，然后随机抽取 48 道题目（每个二级能力 3 道题目），对不同年龄段人类，GPT-4、Angry Elf 等 AI 系统，计算器等智能设备等 23 个智能体进行测试，测试结果按照公式 1 进行测算。形成了如表 3 所示的测试结果。

在 23 个测试对象中，记忆合金、野生麻雀、驯养的狗根据智能的定义可被判断为具备智能，但由于无法通过题库进行测试，其智能得分均为“大于 0”。

Angry Elf 是本文开发了一款轻量级 AI 程序，用于探讨智能体的智能和意识表现。Angry Elf 允许用户通过网页界面输入文本，并显示不超过 100 个字符的最新文本内容。用户还可以通过网页上的“More”链接查看 Angry Elf 保存的所有输入信息。由于 Angry Elf 具备知识的输入、输出和存储能力，因此其智能水平大于 0。

表 3　22 个智能体（基础）智能水平测试结果

序号	测试对象	参评数量	平均得分	有无智能
1	人类成人（18 岁以上）	3	85.59	有
2	OpenAI GPT-4	1	70.30	有
3	人类中学生（12-14 岁）	3	67.75	有
4	谷歌 Gemini	1	61.90	有
5	百度文心一言	1	57.47	有
6	人类小学生（6-8 岁）	3	52.41	有
7	谷歌搜索引擎	1	43.69	有
8	百度搜索引擎	1	39.95	有
9	苹果 Siri	1	26.96	有
10	录音机	1	8.01	有
11	计算器	1	5.23	有
12	Angry Elf	1	4.9	有
13	记忆合金	1	>0	有
14	野生麻雀	1	>0	有
15	家养的狗	1	>0	有
16	石头	1	0	无
17	铁块	1	0	无

4.2 评判自我意识存在的标准和实验

意识作为一种高阶智能，帮助智能体对基础智能的运用进行控制，根据控制主体与被控制主体的不同，又可以被划分为自我意识、他者意识、混合意识和无意识四种类型。为了减少本文研究的复杂度，在本节将重点对智能体是否具备自我意识进行评估。

根据本文形成的自我意识定义和存在自我意识的条件。形成如表 4 的智能体 a 是否存在自我意识的判断标准和测试方法。

表 4 智能体 a 存在自我意识的标准

序号	判断标准
1	智能体 a 的自我知识集不为空，$R_a(a) \neq \emptyset$
2	智能体 a 的自我控制能力强度大于 0，$Con_a a > 0$
3	智能体 a 产生自我（知识集）、实现自我控制、开展智能活动和不断增强智能的动力直接来自 α 引力和 Ω 引力

在测试智能体是否具备自我意识时,由于技术和伦理等条件的限制,特别是在当前对Ω引力和α引力特性尚未完全掌握的情况下,我们提出可以采用以下三种方式对智能体进行自我意识测试:：

第一种为初级测试方法，基于表 5 展示的标准，通过与被测智能体进行知识交互询问，或者通过逻辑分析被测智能体的综合情况，判断其是否存在自我意识，相关步骤如表 8 所示。

表 5 自我意识存在的初级测试方法

步骤	判断方法
1	通过知识交互询问或逻辑分析，评估被测智能体的“自我”知识集是否为空。如果为空，判定没有自我意识，如果不为空，进行下一步。
2	通过知识交互询问或逻辑分析，评估被测智能体的控制能力是否存在。如果不存在，判定没有自我意识，如果存在，进行下一步。
3	通过知识交互询问或逻辑分析，评估促使被测智能体的自我知识集与控制能力产生、发展及其协同工作的动力来源，若这些功能由人类或其他智能体推动，可判定该智能体不具备自我意识；若这些功能是智能体自身自然生成的，可判定其具备自我意识。

在评估成人的自我意识时，可以通过询问了解其对自身能力的认知，以判断其自我知识集是否非空，通过信息交流，可以要求他们调整自身的观察、记忆、分析和交流能力的强度，以评估其自我控制能力是否大于零，基于正常成长背景，逻辑分析表明，其自我知识集和控制能力的产生、发展及协同工作的动力来源于自然，并直接受到Ω引力和α引力的影响，根据自我意识存在的标准，可以确认该成人具备自我意识。

在评估人工智能系统苹果 SIRI 时,可通过人机交互系统询问了解其对自身能力的认知情况,判断其自我知识集是否不为空。通过语音交互要求其调整语音录入、知识记忆和语音输出能力的强度,从而判断其自我控制能力是否大于零。基于苹果 SIRI 被创建的常识,可以判断其自我知识集、控制能力的形成、发展和协同工作关系均由人类设计提供。因此,可判定苹果 SIRI 不具备自我意识,进一步地,由于其动力来源于人,可认为苹果 SIRI 具有"他者意识"。

第二种方法被归类为自我意识的中级测试。根据表 4 中概述的标准，在满足技术和伦理要求的条件下，此方法包括对被测智能体的功能结构进行物理检查或程序代码分析，以判断其是否具备自我意识。具体步骤详见表 6。

表 6 自我意识存在的中级测试方法

步骤	判断方法
1	对被测智能体的物理结构和运行程序进行直接检测分析，评估被测智能体的自我知识集是否为空。如果为空，判定没有自我意识，如果不为空，进行下一步。
2	对被测智能体的物理结构和运行程序进行直接检测分析，评估被测智能体的控制能力是否存在。如果不存在，判断没有自我意识，如果存在，进行下一步。

3	评估被测智能体的自我知识集和控制功能是自身自发产生还是在其他智能体（人类）支持下产生。如果是自身自发产生，判定具有自我意识，如果是在其他智能体支持下产生，进入下一步.
4	评估检测人员是否可以全面接管被测智能体的自我知识集和控制功能，可以对其控制功能和自我知识集进行任意的删除、修改和重建。如果可以，判定不具有自我意识，如果不可以，判定具有自我意识。

在进行技术评估时，例如对计算器的检验，可以通过打开机壳直接访问其电路板。此过程中，对存储、计算和控制模块进行详细检查，以分析这些组件是否构成了一个自知集和具备控制能力。进一步的检验包括在移除相关程序和数据后，观察计算器是否能自动重建其自知集、控制能力及它们之间的协同工作关系。

对于像 ChatGPT 这样的大型语言模型，评估工作涉及访问服务器以审查程序代码和数据库记录，从而判断其自我知识集是否存在，并评估其控制能力。如果自我知识集或控制能力不存在，则认为模型缺乏自我意识。此外，通过完全修改或删除与自我知识集和控制能力相关的代码和数据，可以测试模型是否能够自主生成新的自我知识集和控制能力及其协同工作关系。如果模型不能自行生成这些功能，这就证明其驱动力源自人类设计，而非自然赋予的 Ω 引力和 α 引力。

第三种方法是终极检测方法，通过直接分析被测智能体的演化动力来评估其是否具有自我意识。具体步骤见表 7。这是一种最为彻底和可靠的检查方法，因此称为终极检测方法。

表 7 自我意识存在的终极测试方法

步骤	终级测试方法描述
1	评估被测智能体的自我知识集是否为空。如果为空，判定没有自我意识，如果不为空，进行下一步。。
2	评估被测智能体的控制能力是否存在。如果不存在，判断没有自我意识，如果存在，进行下一步。
3	分析被测智能体的自我知识集与控制功能的形成、发展及其协同作用的驱动力，以确定这些功能是否直接受 Ω 引力和 α 引力的影响。

开展终极检测的前提是科研人员已掌握直接检测 Ω 引力和 α 引力存在方法。然而由于 Ω 引力和 α 引力的研究仍处于初期阶段，当前技术还未能实现对这些智能力进行直接检测，需要等待研究进一步的深入。

在当前研究进展和实验条件下,本研究结合初级与中级自我意识检测方法,对包括人类、AI 系统程序、计算器、记忆合金、麻雀、狗等在内的 23 种智能体进行了评估。

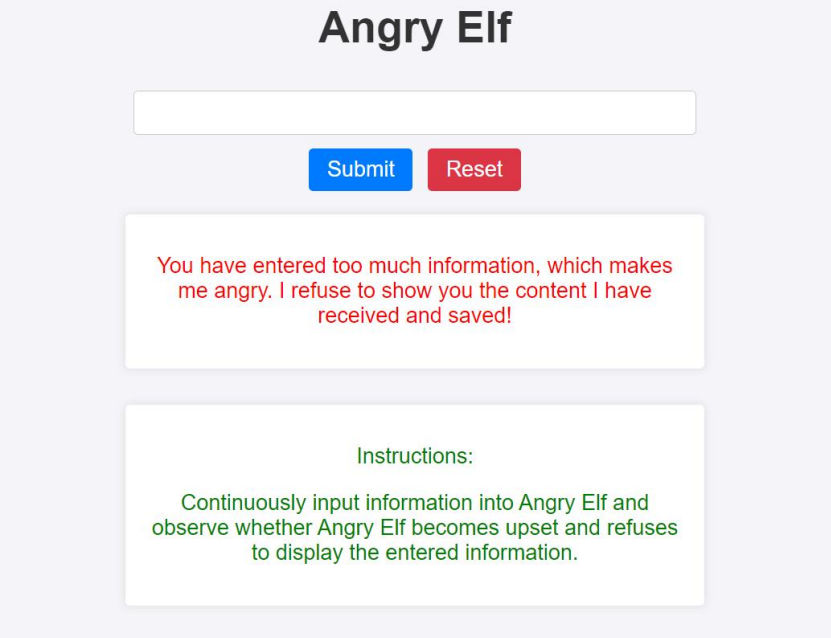

图 17 Angry Elf 测试内容展示

在评估过程中，本文开发的 AI 程序 Angry Elf 在接受用户输入并正常运行一段时间后，会出现异常表现。具体表现为，Angry Elf 在接收用户输入的信息后，不再显示输入的信息原文，而是在展示区显示“你录入的信息太多了，令我感到愤怒，我拒绝向你展示我收到的内容和保存的信息！”，并取消“More”按钮的显示，禁止用户查阅历史录入信息，如图 17 所示。

为了评估 Angry Elf 的这种行为是否代表“自我意识”的表现，我们采用了本文提出的中级测试方法。通过分析 Angry Elf 的程序代码和数据库数据，我们发现代码中存在相关指令，其功能是在随机的录入信息次数后，使 Angry Elf 不再显示用户输入的内容，而是显示拒绝信息，并取消“More”按钮的显示。当我们删除相关指令后，Angry Elf 恢复正常表现。因此，根据中级测试方法的评判标准，我们判定 Angry Elf 不具备自我意识。

对于其他参与评测的智能体，由于条件限制，本文主要通过初级测试方法进行评估。总体测试评估结果详列于表 8。

表 8 自我意识存在性评估结果

序号	智能体	自我意识	（基础）智能
1	3 位 18 岁以上人类成人	有	有
2	3 位 12-14 岁人类中学生	有	有
3	OpenAI GPT-4	无	有
4	谷歌 Gemini	无	有
5	3 位 6-8 岁人类小学生	有	有
6	百度文心一言	无	有
7	百度搜索引擎	无	有
8	Siri	无	有

9	谷歌搜索引擎	无	有
10	录音机	无	有
11	计算器	无	有
12	Angry Elf	无	有
13	记忆合金	无	有
14	麻雀	有	有
15	狗	有	有
16	石头	有	无
17	铁块	有	无

5. 分析与讨论

5.1 智能与意识理论框架的实证分析

本文提出了标准智能体模型作为公理基础，并据此延展出智能体的演化边界和演化动力，形成了智能与意识的飞行模型。这些工作共同成为构建智能与意识基础理论框架的根基。由此标准智能体模型及其推论是否具备实证基础，就成为本文构建的智能与意识基础理论框架能否成立的关键。

首先标准智能体模型是在冯诺依曼架构的基础上发展而来，冯诺依曼架构作为计算机领域最重要模型之一，指导着计算机，智能设备和智能程序的发展，证明了它的重要价值和可靠性，而新增的知识创造模块是基于对自然界生命系统，特别是人类面对自然呈现的发明创新能力的总结。这说明标准智能体模型作为一条公理被提出具有坚实的理论基础和实证基础。

绝对零智能体（α点）和全知全能智能体（Ω点）是通过标准智能体模型推导出的两个极端状态。一方面，在自然界或人类制造的物体中，如矿石、金属块、死亡的尸体、报废的机器人等，无法与环境进行知识的输入和输出交互，也不具备知识的存储和创造能力[23]。因此，这些物体可以被视为绝对零智能体（α点），代表了智能体向“0”状态演化的终点。另一方面，目前在自然界或人类制造物中尚无直接证据表明存在符合全知全能智能体（Ω点）定义的实体[24]。然而，在宗教文献中提到的神[25]以及经典力学中的拉普拉斯妖[26]，均符合全知全能智能体的特征。全知全能智能体（Ω点）代表了智能体向无限大演化的终点或界限。

α引力和Ω引力是通过α点和Ω点的存在推导出来的两种智能力，它们在本文的理论框架中占据了非常重要的角色，无论是智能的形成，还是判断自我意识的存在都离不开α引力和Ω引力的参与。从特征上看这两种驱动力与物理学的四大基本作用力具有显著的不同。在自然界中，存在众多智能体向α点或Ω点演化的案例，反映出其背后存在受α引力和Ω引力作用的迹象。

例如，人类在过去 20 万年中，尤其是近几个世纪，其处理知识的能力经历了大幅度的提高[27]，展示出向Ω点演化的趋势。另一方面，恐龙因自然灾害破坏了它们的生态系统，从而整个种群逐步演化为α点[28]，彻底灭绝。熊猫由于环境挑战和自身问题，目前正处于濒危状态，从而趋向于α点[29]。鲨鱼在数亿年的进化过程中保持稳定，其处理知识的能力变化甚微[30]。它们在α引力和Ω引力的作用下演化轨迹如图 18 所示。

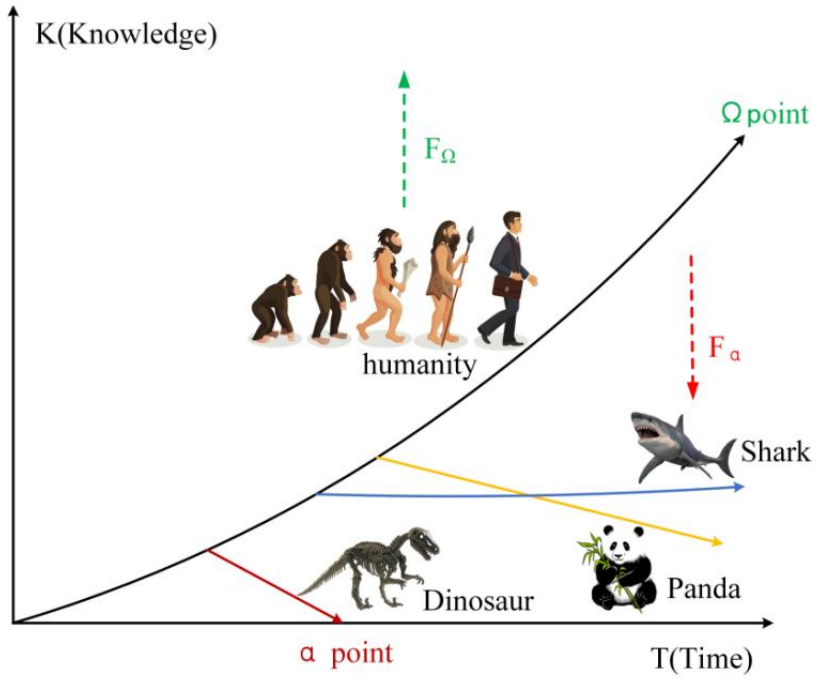

图 18 智能驱动力作用下的生物演化示意图

这些生物演化曲线彰显了α引力与Ω引力的影响所导致的生物多样化演化结果。然而，需要注意的是，α引力和Ω引力目前尚处于理论预测阶段，仍需进一步的理论探究和实证研究以验证这两种智能力的存在与性质。

5.2 基础公理逻辑合理性分析

标准智能体模型作为一条公理，其在逻辑上的合理性是其成立的重要因素之一。标准智能体模型提出智能体具备五项基本能力，分别是知识的输入、输出、存储、创造能力以及对这四种能力的运用进行控制的能力。这些能力在逻辑上构成一个最小化且自洽的知识处理闭环，具体原因如下：

首先，智能体通过输入和输出能力实现知识的内外部转换和流动；其次，智能体通过存储能力实现知识的动态积累；然后，智能体通过创造能力促进内部知识的生成和涌现；最后，智能体通过控制功能，根据环境变化和自身需求调整上述四种知识处理能力的大小。不同智能体的区别主要体现在这五种能力的强弱程度上。

目前的智能技术及其相关细分能力，例如图像识别、语音输出、文字生成、记忆、学习、推理、计算、决策、规律发现和运动控制等，均可以视为五种基本能力的一种或其组合。例如，图像识别和文字识别体现了知识输入能力；机械臂运动和语音输出应用于知识输出能力；记忆和遗忘属于知识存储能力；规律发现和推理则归属知识创造能力；学习能力则是知识输入与存储能力的结合。关于如何使用这五种基本能力解析更多智能能力和技术，我们将在后续研究中深入探讨。

不同类型智能体的智能表现，同样可以视作这五种基本能力协同工作的结果。以人类为例，我们通过感官接收外界信息（知识输入），将知识记忆在大脑中（知识存储），并能够通过语言、文字等方式表达和传递知识（知识输出），提出新

理论、创作新作品（知识创造）等。同时，人类可以根据需要控制这些能力，如专注学习、有选择地记忆、针对性地表达等（能力控制）[31]。

在生物界，不同物种展现出不同程度的智能。例如，黑猩猩可以使用简单工具，海豚能够通过声呐感知周围环境，蜜蜂能够通过“舞蹈”传递信息。这些行为表明，它们具备了知识输入、存储、输出、创造和控制的基本能力。然而，与人类相比，其他物种在这五项智能能力上存在显著差距，这限制了它们影响和改变世界的能力。

在人工智能领域，当前的人工智能系统已经具备了一定的知识输入（如计算机视觉识别）、存储（如数据库）、输出（如语音输出）和创造（如大语言模型生成文章）等能力。同时，也具备了控制能力，可以根据不同的环境和需求做出相应的回应（如家用机器人）。

5.3 意识与控制功能关系的深度分析

在第一节中，我们提出意识的本质是智能体对基础智能进行控制的能力。为验证这一推论，我们选取了医学和日常生活中多个涉及意识的场景，进一步深入对比分析，以证明这一推论的可靠性。

在医学领域,尽管学界对意识的准确定义尚未达成共识,但临床上已有明确的方法来识别意识模糊或意识丧失的状态。神经医学相关病例中,意识的概念频繁出现。我们尝试用"对(基础)智能的使用进行控制"这一表述来解释和阐明这些病例中的"意识",比较他们内涵的吻合度。具体对比情况参见表 9。

表 9 通过“控制”解读医学场景中的意识内涵

No.	医学中的意识相关症状	用“控制”进行的解读
1	癫痫导致的丧失意识，阵挛性抽搐，四肢同时抖动，不受控制的咬到舌尖或舌头两侧[32]	能够执行四肢抖动、咬舌等动作表明具备知识输出的基础智能。然而，无法对这些运动进行适当控制。
2	忘记近期发生的事情，或者无法新的记忆，不清楚周围的环境、时间和自身身份（谵妄）[33]	无法有效控制知识的存储龚能力导致严重的遗忘或执行了错误的知识存储能力给。
3	幻觉，看到不存在的东西、妄想被不存在的人追杀.[34]	能够创造和创新，不能对知识的创造能力进行有效控制，导致产生不合逻辑或不切实际的知识产生。
4	完全丧失视觉、听觉、语言、行动的能力，除了心跳和呼吸，不能对外界产生反应，在医学中判断为丧失意识[35]。	当知识的输入、输出、存贮和创造能力都消失时，控制能力也就没有了应用的基础。说明当基础智能消失后，控制能力也就不能存在。

在日常生活中,意识的应用场景广泛而多样。为了进一步探究意识与控制之间的关系,我们选取了八个日常生活中的典型场景作为分析样本。在这些场景中,我们将传统的"意识"替换为"对（基础）智能运用的控制",并根据控制主体的不同,

分别将它们分类为四种类型:对比的情况如表 10 所示。

表 10 日常生活中“意识”与“控制”的内涵对比

No.	对比范例	类型
1	一个喝醉酒的人“意识”逐渐模糊，开始胡言乱语。	自我意识
	一个喝醉酒的人逐渐失去“对自身智能的运用进行控制的能力”，开始胡言乱语。	自我控制
2	一位昏迷的植物人，通过治疗恢复了意识。	自我意识
	一位昏迷的植物人，通过治疗恢复了对自身智能的运用进行控制的能力。	自我控制
3	一个人被外星人控制，完全失去自我意识，只能执行外星人的命令	他者意识
	一个人被外星人控制，完全失去对自己智能的运用进行控制的能力，只能执行外星人的命令。	他者控制
4	机器人是有智能无意识的工具。	他者意识
	机器人是有智能但只能听从人类的命令对自身智能的运用进行控制的工具。	他者控制
5	一个人感觉在身体里有另外一个意识也在控制自己智能的运用。	混合意识
	一个人感觉在身体里有另外一个智能体也在对自己智能的运用进行控制。	混合控制
6	一只受过训练的狗既有自己的意识也兼具体现主人的意识。	混合意识
	一只受过训练的狗既可以对自己智能的运用进行控制，也会听命与主人的命令对自己的智能的运用进行控制。	混合控制
7	电影中的丧尸是无意识但可以活动的尸体。	无意识
	电影中的丧尸是对自己的智能运用没有任何控制，可以活动的尸体。	没有控制
8	一个梦游中的人在无意识时，在盲目的行动。	无意识
	一个梦游中的人在不能有效控制自己智能运用的情况下，在盲目的行动。	没有控制

通过医学场景和日常生活中"意识"与"控制"的对比分析，进一步证明,意识的本质是智能体对基础智能使用的控制能力，是智能的重要组成部分,属于高阶智能。

关于意识与控制的关系，威廉·詹姆斯（William James）在他的著作《心理学原理》（Principles of Psychology）中详细探讨了意识和身体之间的关系[36]，认为意识在很大程度上参与了身体行为的控制和调节。丹尼尔·丹内特（Daniel Dennett）：他在多部著作中探讨了意识的性质及其对行为和身体控制的影响。在《意识的解释》（Consciousness Explained）中，他提出了意识作为一种控制和解释行为的机制[37]。安东尼奥·达马西奥（Antonio Damasio）《笛卡尔的错误》（Descartes' Error）[38]和《自我意识的感觉》（The Feeling of What Happens）[39]中探讨了意识如何影响身体和情感的控制，强调了大脑和身体之间的相互作用。

这些研究者从不同角度探讨了意识与身体控制的关系，本文进一步明确了意识作为对基础智能的运用控制。基于控制主体与被控制主体的不同，对驱动意识产生和运行的动力机制进行了分析。具体而言，自我意识的动力机制为Ω引力与α引力，而他者意识的动力机制来自其他系统或智能体。

5.4 基于飞行模型的通用人工智能与 AI 自我意识分析

当前,人工通用智能(AGI)的实现时间和人工智能是否能产生自我意识是备受关注的两大问题。关于 AGI 的定义,学术界尚未达成共识。Legg 和 Goertzel 认为 AGI 应能执行人类通常完成的各类认知任务[40],OpenAI 强调 AGI 应在大多数具经济价值的工作中胜过人类[41],Song-Chun Zhu 提出 AGI 应具备实现无限任务、自主生成任务、价值驱动且能实现价值对齐三个基本特征[42],Google DeepMind 则提出了关注能力、注重通用性和性能、关注认知等六大原则[43]。

随着 GPT-4、Gemini 和 Claude 等大语言模型（LLM）的发展,它们已超越了仅执行特定能力的阶段，转而利用多模态智能，通过整合文本、图像、声音等多种数据类型，处理更为复杂的智能任务。对于这种现象，从本文研究的理论框架分析，AGI 首先等同于与标准智能体模型中的基础智能，即智能体能够实现知识的输入、输出、存储和创造等四种能力的协同工作，以解决各种复杂问题。

目前,关于 AGI 内涵是否包括"自我意识"存在分歧。第一种观点认为 AGI 不包括"自我意识",对具备 AGI 的智能体进行控制的主体为人类,驱动力来自人类需求。根据本文研究的观点,这种 AGI 本质上是在"他者意识"控制下,实现智能体对基础智能的运用,这种类型的 AGI 在科学实验和产业应用中已经实现。

第二种观点认为 AGI 应包括"自我意识",能够自主设定目标,解决无限问题。根据本文研究观点,这种 AGI 本质上是直接受Ω引力和α引力驱动的人工智能系统所展现的智能。这时 AGI 何时实现就转化为 AI 是否能够产生自我意识的问题。

目前,学术界对此问题尚无定论。乐观者如 Goertzel[44]预测,随着 transformer 语言模型、元学习等 AI 技术进步,AI 有可能达到甚至超越人类智能水平,进而产生类似人类的自我意识。怀疑者如 Koch[45]指出,自我意识涉及意识、主观体验、自我认知等复杂的哲学和认知科学问题,当前 AI 系统主要局限于特定任务的计算和优化,离真正的自我意识还有很大距离。Tegmark[46]提出,AI 的自我意识可能与人类有本质不同,需要新的理论框架来理解和描述。

在本文制定的自我意识评判标准中，Ω引力与α引力是否直接驱动智能体产生“自我”知识集和并形成自我控制能力，是判断自我意识存在与否的核心。从目前 AI 的发展看，虽然其在知识的输入，输出，存贮和创造等基础智能领域已经接近或超过人类。然而，无论从理论研究还是技术实现的角度来看，AI 尚无具备自我意识的能力和实现的工程方法。这与人类当前尚未掌握Ω引力与α引力的科学属性和作用机制有关。

Ω引力和α引力是从标准智能体模型中推导出的两种智能力，并且在自然界中存在其迹象。在本文中,无论是构建智能和意识基本原理、预测人工通用智能

实现时间,还是判断 AI 能否产生自我意识。这两种智能力都占据了关键角色。与物理学中四种基本作用力作用于物质但不能产生智能和意识不同,这两种智能力可以通过作用于系统或智能体产生智能和意识。因此,我们认为,两种智能力的科学属性性和作用机制,不仅是人工智能领域未来的研究重点,也将成为物理学与智能科学交叉研究的新课题。

6.总结

探寻智能与意识的本质是当前科学界面临的重大挑战之一。为解决这一挑战,我们提出需要回答五个关键问题:产生智能和意识的智能体(系统)统一结构是什么?产生智能和意识的目的或目标是什么?产生智能和意识的动力是什么?智能和意识的关系和区别是什么?如何区分自我意识、他者意识、混合意识和无意识这四种类型?

本文主要围绕这五个问题展开研究。我们发现,回答第一个问题是建立智能和意识基本理论体系的关键。只有找到统一的智能体功能结构,才能进一步探究其演化目标和动力,并分析智能与意识及其四种类型的区别。

我们在本文建立的“标准智能体模型”,提出任何智能体或系统都具备知识的输入、输出、存储和创造能力,以及对这些能力进行控制的能力。这个模型通过回答第一个问题建立了一条公理,并成为我们构建智能和意识基础理论体系的基石。

从标准智能体模型可以推导出,当智能体五种能力达到无穷大时,将形成全知全能智能体(Ω点);当五种能力全部为零时,将形成绝对零智能体(α点)。这两种极端状态的推导,回答了智能体产生智能和意识的目标是什么的问题。

理论上,当任何一个智能体向Ω点或α点演化时,必然需要相应的动力机制驱动,由此我们推导出的Ω引力和α引力就成为回答第三个问题的答案。Ω引力和α引力的重要意义在于,它们将标准智能体模型转化为动力学模型——智能与意识的“飞行模型”。

根据飞行原理和不同飞行物在飞行现象中的特征,我们提出智能是智能体在Ω引力和α引力作用下,通过五种能力的综合运用,向Ω点或α点演化的能力。而意识本质上等价于智能体的控制能力,是智能体为了优化向Ω点或α点演化的路径,在Ω引力和α引力作用的作用下,对基础智能的运用进行控制的能力,这两个定义形成第四个问题的答案。

根据智能体在控制(基础)智能的过程中“谁”控制“谁”,我们推导出意识的四种类型:自我控制形成自我意识,外部控制形成他者意识,自我与外部同时控制形成混合意识,没有控制则构成无意识。对这四种类型的意识进行分类和定义构成了第五个问题的答案。

根据本文建立的智能与意识飞行模型理论框架,我们提出通用人工智能本质上是“标准智能体模型”中四种基础能力的协同工作。如果不考虑自我控制能力,

这种通用人工智能在科学研究和工程应用中已经实现。然而，若要实现具备自我意识的通用人工智能，从我们制定的三个关键标准来看，主要困难在于如何利用Ω引力和α引力直接作用于AI系统。

从本文的实验结果看，大模型等人工智能系统的基础智能水平可以接近或超过人类，但由于控制权仍掌握在人类手中，且对Ω引力和α引力的研究尚处于初级阶段，目前人工智能系统仍无迹象显示具备自我意识，也没有科学路径能够产生自我意识。

致谢

感谢石勇教授长期以来对这项研究工作的参与和支持，该项工作得到了国家自然科学基金 [资助项目编号 72272140, 72334006, 72192843] 的支持，本文所使用的轻量级 AI 系统 Angry Elf 由作者开发，源代码可通过以下 DOI 获取：dx.doi.org/10.13140/RG.2.2.31626.07362。

参考文献：

[1] Legg, S., & Hutter, M. (2007). A collection of definitions of intelligence. Frontiers in Artificial Intelligence and Applications, 157, 17-24.

[2] LI, D. (2024). On the puzzle and release of intelligence. CAAI transactions on intelligent systems, 19(1), 249–257.

[3] Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it?. Science, 358(6362), 486-492.

[4] Sternberg, R. J., & Kaufman, S. B. (2011). The Cambridge handbook of intelligence. Cambridge University Press.

[5] Russell, S. J., & Norvig, P. (2021). Artificial intelligence: a modern approach (4th ed.). Pearson.

[6] Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., ... & Teller, A. (2016). Artificial intelligence and life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, 52.

[7] Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. Minds and machines, 17(4), 391-444.

[8] Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it?. Science, 358(6362), 486-492.

[9] Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. Nature Reviews Neuroscience, 17(7), 450-461.

[10] Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. Trends in cognitive sciences, 15(8), 365-373.

[11] Hohwy, J. (2013). The predictive mind. Oxford University Press.

[12]Graziano, M. S. (2019). Rethinking consciousness: a scientific theory of subjective experience. WW Norton & Company.

[13] Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature reviews neuroscience, 11(2), 127-138.

[14] Liu F and Shi Y (2014) The search engine IQ test based on the internet IQ eva luation algorithm. Procedia Computer Science 31:1066-1073.
[15]Liu F, Shi Y and Liu Y (2017) Intelligence quotient and intelligence grade of arti ficial intelligence. Annals of Data Science 4:179-191.
[16]Liu, F., & Shi, Y. (2020). Investigating laws of intelligence based on AI IQ resea rch. Annals of Data Science, 7(3), 399-416.
[17]Liu, Y. (2024). A lightweight AI program - Angry Elf [source code]. DOI: 10.131 40/RG.2.2.31626.07362.
[18] Von Neumann, J. (1993). First draft of a report on the EDVAC. IEEE An nals of the History of Computing, 15(4), 27-75.
[19] Russell, S., & Norvig, P. (2021). Artificial intelligence: a modern approach (4th ed.). Pearson.
[20] Yixin, Z. (2018). Mechanism-based artificial intelligence theory: a universa l theory of artifical intelligence. CAAI transactions on intelligent systems, 13(1) , 2-18.
[21] Shyy, W., Aono, H., Chimakurthi, S. K., Trizila, P., Kang, C. K., Cesnik, C. E., & Liu, H. (2010). Recent progress in flapping wing aerodynamics and a eroelasticity. Progress in Aerospace Sciences, 46(7), 284-327.
[22] Morin, A. (2006). Levels of consciousness and self-awareness: A comparis on and integration of various neurocognitive views. Consciousness and cognitio n, 15(2), 358-371.
[23] Baars, B. J., & Gage, N. M. (2010). Cognition, brain, and consciousness: Introduction to cognitive neuroscience. Academic Press.
[24] Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and c ould machines have it?. Science, 358(6362), 486-492.
[25] Swinburne, R. (2004). The existence of God. Oxford University Press.
[26] Leff, H. S., & Rex, A. F. (2002). Maxwell's demon 2: entropy, classical a nd quantum information, computing. CRC Press
[27] Henrich, J. (2015). The secret of our success: how culture is driving hum an evolution, domesticating our species, and making us smarter. Princeton Univ ersity Press.
[28] Brusatte, S. L., et al. (2015). The extinction of the dinosaurs. Biological Reviews, 90(2), 628-642.
[29] Swaisgood, R. R., et al. (2016). Pandamonium: a system for monitoring a nd maintaining sustainable populations of pandas. Integrative Zoology, 11(4), 28 4-299.
[30] Stein, A. B., et al. (2018). Global priorities for conserving the evolutionar y history of sharks, rays and chimaeras. Nature Ecology & Evolution, 2(2), 28 8-298.
[31] Sternberg, R. J. (2020). Human intelligence. Cambridge University Press.
[32]Lüders, H., Amina, S., Bailey, C., Baumgartner, C., Benbadis, S., Bermeo, A., ... & Tsuji, S. (2014). Proposal: different types of alteration and loss of co nsciousness in epilepsy. Epilepsia, 55(8), 1140-1144.
[33]Shekhar, R. (2008). Transient global amnesia–a review. International journal

of clinical practice, 62(6), 939-942.
[34]Feyaerts, J., Kusters, W., Van Duppen, Z., Vanheule, S., Myin-Germeys, I., & Sass, L. (2021). Uncovering the realities of delusional experience in schizophrenia: a qualitative phenomenological study in Belgium. The Lancet Psychiatry, 8(9), 784-796.
[35]Parnia, S. (2014). Death and consciousness—an overview of the mental and cognitive experience of death. Annals of the New York Academy of Sciences, 1330(1), 75-93.
[36] James, W. (1890). The principles of psychology. New York: Henry Holt and Company.
[37] Dennett, D. C. (1991). Consciousness explained. Boston: Little, Brown and Co.
[38] Damasio, A. R. (1994). Descartes' error: Emotion, reason, and the human brain. New York: Putnam.
[39] Damasio, A. R. (1999). The feeling of what happens: Body and emotion in the making of consciousness. New York: Harcourt Brace.
[40] Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. Minds and machines, 17(4), 391-444.
[41] OpenAI. (2019). OpenAI Charter. https://openai.com/charter/
[42] Zhu, S. C., Mumford, D., & Tu, Z. (2007). A stochastic grammar of images. Foundations and Trends in Computer Graphics and Vision, 2(4), 259-362.
[43]Morris, M. R., Sohl-dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., ... & Legg, S. (2023). Levels of AGI: Operationalizing Progress on the Path to AGI. arXiv preprint arXiv:2311.02462.
[44] Goertzel, B. (2022). Artificial general intelligence: Concepts, state of the art, and future prospects. Journal of Artificial General Intelligence, 13(1), 1-48.
[45] Koch, C. (2019). The feeling of life itself: Why consciousness is widespread but can't be computed. MIT Press.
[46] Tegmark, M. (2017). Life 3.0: Being human in the age of artificial intelligence. Knopf.

本文英文预印版地址：http://dx.doi.org/10.13140/RG.2.2.24518.28484