从语言出发：我们离超级智能还有多远

胡雨韦，系中国人民大学高瓴人工智能学院博士研究生

鉴于当前人工智能的核心职能高度依赖语言载体，无论是纯文本模型还是多模态模型，均需借助语言作为与人类世界交互的基本媒介；智能体通过编程语言或上下文协议调用外部工具，进而实现与现实世界的有效互动；即便在最新发展的世界模型与具身智能框架中，语言仍承担关键作用，模型需通过语言认知自身状态并规划行为序列。因此，评估超级智能（ASI）的实现程度必须首先厘清语言模型的智能表现与能力边界。

一、超级智能是移动靶

当前，通用人工智能、超级智能有着模糊的共识，即各项能力与人类同水平、各项能力全面超越人类的智能，但大家难以给出其具体定义和衡量标准。这源于超级智能的定义随技术进步而变化，人类不断发展和进化出不同的能力，使超级智能的目标成为一个移动靶。这一“目标移动”现象源于两方面：

其一，技术持续进步推动人类能力本身不断扩展，从石器时代至今，工具的使用不断重塑智能的内涵；其二，人类倾向于不断识别并强调人工智能尚未具备的“人类独有能力”，从而在定义层面延缓了通用人工智能（AGI）的认定时间点。当前 AI 系统在单一任务上已可超越人类，但在处理多样化、高复杂性任务方面仍与人脑存在差距，因此研究重点应聚焦于如何逐步弥合该差距。

二、语言模型的能力边界

然而，这种渐进式路径是否能导向真正的超级智能，仍需深入考察语言智能的根本边界。大语言模型的有效性机制可从“从零到整”与“由整到零”两个维度理解：前者指模型将任务统一为“下一个词预测”，系统地建模语言的概率分布，实现语言空间的全面覆盖与合理生成，其过程具有前瞻性规划特征；后者则强调模型通过宏观任务接口反复训练，使基础能力组合成复杂功能，形成能力涌现。在语言哲学层面，该过程与维特根斯坦“语言游戏论”中语义源于使用的观点相契合，并符合组合性原则——有限模块经规模扩展后可生成无限的语义表达。

三、语言智能的局限性

然而，语言智能面临双重困境：语言表征本身具有封闭性，难以突破其描述边界；同时高质量训练数据趋于枯竭，使用合成数据可能导致模型退化，形成“垃圾进，垃圾出”的循环。

进一步地，语言智能存在三项根本局限：

其一为有限性，语言在描述连续动态信息（如运动细节）时存在大量信息丢失，且过度依赖离散标签导致表达能力受限；

其二为无现象性，模型无法获得真实世界的感质体验，正如“黑白房间中的玛丽”即使掌握全部颜色知识仍缺乏直观感受；

其三为无自主性，模型缺乏内在动机与价值系统，参数通常被冻结，无法在交互中实现自我更新与进化。

四、通往超级智能的其他路径

基于上述局限，必须在语言路径之外探索更为复合的超级智能建构方式。一种可行的多层架构包括：底层为亚符号神经网络，直接处理连续感官输入，形成直觉感知以突破语言有限性；中层引入物理建模与社会推理，构建世界模型与因果推断能力；顶层则保留符号与类语言结构，负责复杂任务规划与交流。通过感知、推理与符号表达的贯通设计，有望系统地克服语言智能在有限性、无现象性与无自主性方面的缺陷。正如常言所道，“悲观者永远正确，乐观者永远向前”，尽管 AGI 发展路径尚未明晰，学界仍应秉持积极探索之态度，在持续的思想碰撞与技术迭代中推动其迈向稳健与成熟的未来。