AI 未来发展趋势

刘伟，系北京邮电大学人工智能学院特聘研究员

当前关于超级智能的讨论中，一种值得关注的立场是：超级智能的实现并非仅依赖于人工智能技术的独立发展，而是有赖于人、机器与环境三者的深度协同。相应地，对智能本质的理解也需置于“人-机-环”这一整体框架之下进行。下文将从三个方面系统阐述这一观点。

一、人工智能的理解

人工智能的独立发展存在内在局限性，唯有通过人、机、环三要素的协同整合，才能推动其向更高层次演进。人工智能不仅是一套基于数学规则与统计概率的技术体系，其应用场域已广泛延伸至人文艺术、哲学宗教等非技术领域。有研究指出（如《AI 战争》一书所述），智能问题不能简单化约为“AI+”或“+AI”的模式；前者以技术本身为驱动力，后者则以应用需求为导向。真正完善的智能发展路径在于人、机、环三者的有机融合。以军事智能为例，其典型特征可归纳为“诡”（态势的千变万化）、“诈”（借助 AI 实施的谋略与欺骗）、“算”（不仅包括计算，更涵盖“算计”即策略推理）、“胆”（基于信息的果断决策）以及“善”（既指伦理向善，亦指能力层面的擅长）。

当前主流的大语言模型，正如杨立昆所批评，存在根本性理论缺陷，难以实现人类级别的智能水平。其问题主要表现为：生成内容的指数级幻觉、对算力资源的无止境依赖，以及面对莫拉维克悖论所揭示的高层认知困难。《代数大脑》一书在二十余年前即已指出，基于多层神经网络的人工系统，因其本质上由线性函数与激活函数所构建，难以避免机器幻觉与机器欺骗现象。因此，这类系统难以满足国防、精密工业等高可靠性领域的需要，目前主要适用于对话生成与娱乐类任务。大模型的技术瓶颈在某种意义上也标示出当前人工智能的整体局限：在数据层面，真正的智能应体现为以小样本解决复杂问题；在推理机制上，将智能纯粹还原为计算或形式逻辑是一种认识误区；在表征方式上，人为割裂理性表征与感性表征并不可行；在意识维度上，智能的实现不能仅依赖语言，还必须整合思维等更深层次的认知能力。

在后大模型时代，智能的发展方向将走向人、机、环三者的深度融合。人类擅长谋划与直觉判断，而机器精于高速计算与精确执行，二者在特定环境背景下相互协调、产生共振，可共同实现安全、高效与舒适的系统目标。三者各自存在能力边界，唯有通过协同作用才能实现整体发展。因此，“人-机-环”系统智能代表着未来智能演进的基本趋势。

二、人工智能的安全问题

人工智能系统面临的安全风险，主要源于三个层面的错误：人为错误、机器错误以及环境错误。当前，AI 安全治理的关注焦点可归纳为六大核心议题：一是数据完整性问题，包括数据造假与数据投毒；二是地理限制问题，即明确 AI 系统的可用与禁用区域；三是人机关系问题，确保人在系统中始终处于核心地位；

四是自主性问题，需认识到真正的超级智能应兼容自主与他主两种模式；五是相关法律与标准体系的建立；六是系统测试与评估指标的设计。

三、人工智能的治理

相应的治理框架也应覆盖人、机、环三个维度。人所承担的职责是“正确地做事”，即保障操作与决策过程的合规与合理；机器的主要任务是“做正确的事”，即确保其行为与预设目标及伦理规范一致；环境则需“提供正确的平台”，即为系统运行提供稳定、可靠的基础设施与规制条件。

构建超越现有水平的超级智能，关键在于实现人类智慧与机器智能的深度融合，形成“人-机-环”一体的系统性能力。人机交互的本质是“共在”，即人的生理特性与机器的物理属性相结合；而人机混合智能的未来在于“共生”，即人类智慧与机器智能的互补与协同。在此进程中，计算智能可类比为“刻舟求剑”，感知智能近似于“盲人摸象”，而认知智能则更接近于“曹冲称象”——亦即人类智慧与机器能力共同应对复杂问题的过程。未来的“人-机-环”系统所追求的是深度态势感知，亦即一种洞察智能，其特质类似于“塞翁失马”，不仅能够感知当前状态，更能洞悉事物间的深层关联、发展趋势及潜在可能。为实现上述目标，“人-机-环”系统智能的发展应致力于五个关键方向：主动推荐、交互学习、高效容错、混合决策与按需组网。