人工智能发展的一些观点

张亚勤
中国工程院院士
清华大学智能产业研究院（AIR）院长
2024 年 6 月 6 日

人工智能（AI）是当代最大的技术变革力量，是第四次工业革命的技术引擎。AI 将与各个行业深度融合，推动产业升级。本文将阐述 AI 大模型的五大核心发展趋势，提出 AI 演进过程中的五个关键观点，并展望无人驾驶技术的未来发展前景与实际应用潜力。同时，本文还提出五个具体的发展治理建议，共同推动人工智能技术的全面落地。

一、AI 大模型的五个发展方向：

AI 大模型作为数字化 3.0 的重要基石，其发展将决定未来技术的高度。以下是未来 AI 大模型架构的关键发展方向：

1. **多模态智能：**

- **全面和深度的智能分析：**结合语言、文字、视频、激光雷达点云、3D 结构信息、4D 时空信息乃至生物信息，实现多尺度、跨模态的智能感知、决策和生成。

2. **自主智能：**

- **智能体（Agent）：**将大模型作为一种工具，开发能够自主规划任务、编写代码、调动工具、优化路径的智能体，实现高度的自我迭代、升级和优化，实现自主智能。

3. **边缘智能：**

- **高效率、低功耗、低成本、低延时：**将大模型部署到边缘设备端，如 AI PC，AI 手机，AI 电视等，实现高效率、低功耗、低成本、低时延的处理和响应，从而实现边缘智能。

4. **物理（具身）智能：**

- **无人车、机器人等：**大模型正在被用到无人车、机器人、无人机、工厂、交通、 通讯、电网、电站和其他物理基础设施，提升其自动化和智能化水平，从而实现具身智能。

5. **生物智能：**

- **人体、人脑、医疗机器人：**将大模型应用到人脑、生命体、生物体里，实现大模型与生物体连结的生物智能，并最终实现信息智能、物理智能和生物智能的融合。

二、AI 发展的五个观点：

1. **大模型/GenAI 是未来 10 年的主流技术和产业路线**

大模型（如 GPT-3、BERT 等）和生成式 AI（Generative AI）将在未来 10 年内成为主流技术和产业路线。

2. **基础大模型+垂直大模型+边缘模型（开源+商业）**

基础大模型将是人工智能时代的技术底座，将与垂直产业模型及边缘模型一起形成新的产业生态，其生态规模将比 PC 时代大 100 倍，比移动互联网时代大 10 倍以上。在这个生态中，开源模型与商业模型并存，为开发者提供灵活的选择。

3. **Token-based（统一表征）+ Scaling Law（规模定律）。**

大模型最核心的要素是 Token-based（统一表征）和 Scaling Law（规模定律）。Token-based 方法通过将文本和其他类型的数据统一编码为 Token，使得模型能够处理不同形式的输入。Scaling Law 则揭示了模型规模与性能之间的关系，表明随着模型参数和数据规模的增加，模型的表现会显著提升。

4. **需要新算法体系（世界模型，DNA 记忆，Agent，RL，概率系统+决定系统，100X 效率提升），Transformer/Diffusion/AR 在 5 年内会被颠覆**

与人脑相比，现有算法存在效率低、能耗高的问题，需要开发新的算法体系，

包括世界模型、DNA 记忆、智能体（Agent）、强化学习（RL）、概率系统和决定系统，实现 100 倍的效率提升。未来 5 年内会在 AI 技术架构上有大的突破，当前主流的 AI 技术框架 Transformer/Diffusion/AR，可能在未来五年内被新技术所颠覆。

5. 大模型 => 通用人工智能（AGI）

15-20 年内实现通用人工智能（AGI），并通过“新图灵测试”。

- **0-5 年：信息智能。**0 至 5 年内，在信息智能领域，对语言、图像、声音和视频的理解、生成等方面通过新图灵测试。
- **0-10 年：物理智能（具身智能）。**0 至 10 年内，在物理智能领域，实现大模型在物理环境中的理解与操作能力，通过新图灵测试。
- **0-20 年：生物智能。**0 至 20 年内，在生物智能领域，聚焦人体、脑机接口、生物体、制药和生命科学，实现大模型与生物体连结的生物智能，通过图灵测试。

三、无人驾驶的五个观点：

1. 无人驾驶是未来五年最大物理（具身）智能应用：第一个通过“新图灵测试”的具身智能（安全 10X + 人性化 == 好司机 + 老司机）

未来五年内，自动驾驶将成为物理（具身）智能领域中最重要的应用之一，有望成为第一个通过“新图灵测试”的具身智能系统。

- **安全性：**完全无人的无人驾驶安全要比人类驾驶至少高 10 倍，达到了人类好司机的水平。
- **人性化体验：**通过模仿学习和自主学习，自动驾驶系统将具备更自然的驾驶风格，结合乘客驾驶习惯，提供更人性化的体验，达到人类老司机水平。

2. 大模型/生成式 AI 加速 L4 泛化能力：安全和智能（数据，长尾，常识）

大模型和生成式 AI 将在提升 L4 级别自动驾驶系统的泛化能力方面发挥关键作用

- **数据智能**：无人驾驶 Corner Case 数据不足，大模型生成式 AI 可结合真实数据生成高质量 Corner Case 数据。
- **长尾问题**：生成式 AI 可填补 Corner Case 中场景仿真、模拟不足的问题，解决感知领域的长尾问题。
- **常识推理**：大模型的推理能力可以帮助自动驾驶系统理解并应对道路上的突发情况，从而提高自动驾驶的能力上限。

3. 多模态（视觉+激光雷达等）、端到端、云端大模型、车端实时精确模型

自动驾驶技术将整合多模态传感器数据（如视觉、激光雷达等），采用端到端训练，实现云端大模型与车端实时精确模型的协同工作。

- **多模态融合**：相较于人类而言，机器具备多模态感知优势，通过融合视觉、激光雷达和其他传感器数据，自动驾驶系统能够更加全面地感知周围环境。
- **端到端**：无人驾驶算法形成了大量为某一任务特质化的碎片化小模型，未来可统一为端到端大模型。
- **云端与车端协同**：云端大模型提供通用性泛化能力，车端模型提供实时精确响应和本地优化部署，云端与车端协同可确保驾驶决策兼具泛化性、及时性和准确性。

4. 单车智能为主，车-路-云协同，安全冗余+智能交通

未来的自动驾驶技术将以单车智能为主，车-路-云协同工作，确保安全冗余，辅助智能交通。

- **自动驾驶车辆需具备单车智能能力**：每辆自动驾驶汽车都必须具备独立的智能驾驶能力。
- **车-路-云一体化带来安全冗余，提升驾驶安全**：通过车-路-云一体化，为自动驾驶提供多重冗余，提升驾驶安全。
- **车-路-云一体化提升交通效率，辅助智能交通**：车-路-云一体化决策控制优化交通流量、提升交通效率。

5. 2025 年是无人驾驶的“ChatGPT 时刻”，2030 年无人驾驶成为主流（10%新车具备 L4 能力）

- 2025 年，在一个具备复杂交通环境的大城市，无人驾驶将达到人类老司机水平，实现无人驾驶的“ChatGPT 时刻”。
- 2030 年，无人驾驶汽车将成为市场主流，预计有 10%的新车具备 L4 级别的自动驾驶能力。

四、AI 发展治理的五个建议：

前沿大模型能力在飞速发展，巨大的能力也带来潜在的风险。这迫使我们思考人工智能技术对社会、文化、伦理等方面的影响和责任。我们要重新审视人类与机器的关系，以及人类自身的本质和价值。对于人工智能技术的不确定性和复杂性，我们必须做好充分准备和应对。我们应充分重视人工智能可能带来的风险，将伦理问题和价值观置于技术之上。特别是当自主智能、具身智能和生物智能进入大规模部署时，大模型须有更强的可控性、更好的可解释性、和更有效的治理。

1. 建立分级体系

为大模型做好分级分类，对前沿大模型制定场景约束和评估体系。

- **场景约束**：根据不同的应用场景，设立明确的约束条件，确保大模型在特定环境中的有效性和安全性。例如，在医疗、自动驾驶等高风险场景中，制定严格的安全标准和合规要求。
- **评估体系**：建立科学、全面的评估体系，对大模型的性能、安全性、可靠性等进行评估。包括模型的准确性、响应速度、资源消耗、隐私保护等多方面指标，确保其在实际应用中的可行性、有效性和合规性。

2. ID 实体映射

进行 ID 实体映射，追溯责任主体，确保信息真实可靠。

- **AI 内容标识**：像标识广告一样，对人工智能生产的数字人等智能体进行标识，有助于用户区分 AI 生成的内容，维护信息的真实性和可靠性

- **物理+生物从属智能体**：明确机器人作为从属体，必须映射到主体，这个主体可以是人或公司等法定实体，如果机器人出现问题，就可以追溯主体责任。

3. 10%+ 投资

加大对大模型风险研究的投入。企业、国家基金会、科研机构等需要将对大模型投资的 10%以上用于 AI 风险研究，有效应对其带来的潜在风险，发展和治理融合、技术与政策融合、产学研融合。

- **政策研究**：研究制定符合伦理和法律标准的 AI 监管政策，促进各国在 AI 治理方面的国际合作，避免因政策差异引发的技术和伦理问题。
- **技术研究**：在系统设计之初就考虑 AI 风险因素，研究具备可解释性的 AI 模型，提高系统透明度，确保 AI 系统的安全性。

4. 设立红线和边界

为了避免 AI 人工智能系统的不安全开发、部署或使用带来的灾难性风险。应设立人工智能发展的红线和边界:

- **不可自主复制或改进**:任何人工智能系统都不应能够在人类没有明确批准和协助的情况下复制或改进自身。这包括制作自身的精确副本以及创造具有相似或更高能力的新人工智能系统。
- **不可权力寻求**:任何人工智能系统都不能采取不当地增加其权力和影响力的行动。

  …

5. 国际沟通合作和协调机制

AI 技术的发展具有全球性，需要加强国际间的沟通合作和协调机制。

- **国际合作**：推动国际间的科技合作，共同应对 AI 技术带来的挑战。建立国际联合实验室和科研项目，促进跨国界的学术交流和合作。
- **标准化**：推动 AI 治理的国际标准化，制定统一的技术规范和标准，促进

不同国家和地区间的技术互通和兼容。

- **协调机制**：建立国际协调机制，特别是在数据隐私、安全等领域，加强国际间的协调和合作，共同维护全球的科技安全和稳定。

五、总结

随着第四次工业革命的浪潮席卷全球，人工智能（AI）技术正成为推动社会进步和产业升级的关键力量。AI 大模型的发展正处于一个关键时期，未来的技术进步将极大地影响各行业的智能化水平。通过明确多模态智能、自主智能、边缘智能、具身智能和生物智能等五大方向，制定科学的分级体系、ID 实体映射、增加投资、设立红线和边界以及加强国际合作，AI 技术将迎来新的飞跃。特别是在无人驾驶领域，大模型和生成式 AI，结合车-路-云一体化协同，提升驾驶安全和交通效率，助力实现无人驾驶的 ChatGPT 时刻。