

神经网络与大模型的现在与未来

摘要：神经网络已从早期的感知机与浅层连接模型，演化为以卷积神经网络、循环神经网络、图神经网络和 Transformer 为代表的多类深度架构；在大规模数据、算力和预训练范式的共同推动下，进一步发展出具有通用迁移能力的大模型与基础模型。当前，大模型正在自然语言处理、计算机视觉、多模态理解、医学辅助、科学发现、工业优化与边缘智能等场景中快速落地，但同时也暴露出幻觉、偏见、能耗高、部署昂贵、可解释性不足、数据治理复杂与安全对齐困难等问题。本文在综合经典文献、近年综述与公开报告的基础上，围绕神经网络和大模型的主要应用领域、核心问题及解决方法、不同网络架构的适用任务，以及未来可能实现突破的方向进行系统梳理。本文认为，神经网络的发展主线正在从“更深更大”转向“更高效、更可信、更具多模态与行动能力”，未来的关键突破将更多出现在高效架构、具身智能、AI for Science、端侧部署与可信治理的交叉地带。

关键词：神经网络；大模型；基础模型；多模态学习；具身智能；可信人工智能

1 引言

神经网络是人工智能中最具生命力的技术路线之一。从 McCulloch–Pitts 神经元模型提出计算性神经元概念开始，研究者逐步建立起从感知机、反向传播、多层感知机到深度神经网络的完整技术谱系^[1]。进入深度学习时代后，卷积神经网络在视觉任务上取得突破，循环神经网络及 LSTM 在语音、文本和时间序列处理中占据主导，而 Transformer 凭借自注意力机制显著提升了长距离依赖建模能力，推动语言模型、视觉模型乃至多模态基础模型进入统一的预训练时代^[2,3,4,8,20]。

与传统“为单一任务设计单一模型”的思路不同，大模型更强调在海量语料、图像、代码与多模态数据上进行预训练，再通过提示学习、参数高效微调或检索增强等方式适配到下游任务^[9,10,11,12]。这种范式极大提升了模型的通用性与迁移性，也使模型从工具性算法逐渐变成智能基础设施。2026 年 Stanford HAI 发布的 AI Index 报告显示，前沿模型在多模态推理、科学问答和产业应用上的能力仍在快速提高，人工智能的影响范围正由互联网应用扩展到科学研究、医疗服务、教育和制造等更广阔的社会系统^[19]。

然而，能力扩张并不意味着问题消失。大模型在回答中仍会出现事实幻觉，在训练与推理中依赖高昂算力，在隐私保护、价值对齐、可解释性和鲁棒性方面也远未成熟^[9,13,14]。因此，对“神经网络与大模型的现在与未来”进行综述，不能

只讨论其能力边界，更需要从应用、问题、架构匹配和未来趋势四个层面建立系统认识。本文按照课程作业要求展开分析，以期后续进一步学习神经网络、大模型及其工程实践提供一份结构化参考。

2 神经网络与大模型的主要应用领域

2.1 自然语言处理、知识服务与软件工程

大模型首先在自然语言处理领域形成最明显的范式替代。以 Transformer 和 GPT-3 为代表的自回归语言模型展示了强大的上下文建模、少样本学习与文本生成能力，使机器翻译、文本摘要、智能问答、搜索增强、法律文书辅助以及教育答疑等任务的技术底座发生变化^[4,8,10]。在软件工程中，代码大模型已能够完成函数补全、单元测试生成、接口解释和文档整理，推动“人机协同编程”成为现实工作流的一部分。与早期针对单项任务训练的模型相比，这类系统的优势在于能够通过统一语言接口处理多类任务，减少任务切换和特征工程成本。

这类应用的本质，是把语言建模能力转化为知识组织和交互能力。大模型不仅在“生成答案”，也在重构信息检索、知识整理与任务执行方式。然而，语言能力越强，用户越容易忽视其错误边界，这也使事实性与可验证性成为其落地中的首要门槛。

2.2 计算机视觉、多模态理解与生成式内容生产

在视觉领域，CNN 长期是图像分类、目标检测和语义分割的主力架构；近几年，Vision Transformer 和大规模视觉基础模型进一步扩展了视觉模型的泛化能力^[2,15,20]。以 Segment Anything 为代表的视觉基础模型表明，分割任务可以从“为单一数据集训练一个专用模型”，转向“训练一个可提示、可迁移、可零样本使用的通用视觉模型”^[15]。与此同时，扩散模型在文生图、图像编辑和视频生成中的表现不断提升，生成式 AI 已深刻影响设计、媒体和内容创作行业^[7]。

更值得关注的是，多模态模型正在打通文本、图像、语音甚至传感器数据之间的表示边界。模型不再只做“看图说话”，而是在统一表示空间中完成理解、检索、规划与交互。这意味着神经网络的应用正在从单模态感知迈向多模态协同认知，也为机器人、自动驾驶和交互式智能体提供了新的基础能力。

2.3 医学辅助、科学发现与高价值专业场景

医疗和科研是大模型最具潜力、同时也最需要谨慎验证的应用场景。医学领域的大模型已被用于病例问答、文献检索、病历摘要、医学编码和临床决策支持等任务。Med-PaLM 相关研究显示，经过领域对齐和评测框架设计后，语言模型在医学问答中已具备相当高的知识覆盖和推理潜力^[16]。与此同时，AI for Science

的快速发展使神经网络被广泛应用于蛋白质结构预测、分子生成、材料筛选、实验规划和跨学科知识发现等方向，模型的作用正从“数据分析器”演变为“科研助手”甚至“实验建议生成器”^[19]。

但在这些高风险领域，模型错误的代价远高于通用问答或娱乐生成。因此，当前应用往往采取“模型辅助专业人员”的模式，而非直接完全替代人类专家。这也说明，大模型在专业场景的价值更可能体现在提高效率、拓宽候选方案和加速信息整合，而不是简单追求全自动化。

2.4 工业互联网、推荐决策与边缘智能

除互联网内容生产外，神经网络已经深入制造、交通、通信与运维等工业场景。时间序列模型和图神经网络可用于设备状态监测、负载预测、交通流分析和供应链优化；视觉模型则承担质检、检测和安全监控任务；而大模型开始进入工业知识问答、操作规程生成和多步骤任务编排等环节^[5,10,19]。在边缘计算环境中，神经网络还需要面对算力受限、时延敏感和隐私约束等现实条件，这促使学界和产业界持续探索轻量化模型、端云协同与稀疏激活等路线。

表 2-1 神经网络与大模型的主要应用领域概览

领域	代表任务	常见模型或架构	主要特点与挑战
自然语言处理	问答、摘要、翻译、代码生成、知识助手	Transformer、GPT 类模型、RAG 系统	通用性强、迁移方便，但事实性与可控性要求高
视觉与多模态	分类、检测、分割、视觉问答、文生图	CNN、ViT、SAM、扩散模型	感知能力强、跨模态潜力大，但训练成本高、幻觉与安全问题突出
医疗与科研	医学问答、病例摘要、分子设计、实验辅助	领域大模型、图网络、生成模型	专业价值高，但评价标准严格、容错空间小
工业与边缘智能	质检、预测维护、时序预测、运维问答	CNN、LSTM、GNN、轻量化 Transformer	场景落地明确，但受时延、算力与隐私约束明显

总体来看，神经网络和大模型的应用版图正在从“单点任务自动化”转向“跨场景能力平台化”。这既体现了模型通用性的提升，也意味着未来竞争的焦点将从模型是否可用，转向模型是否可靠、经济、可治理。

3 神经网络与大模型的主要问题及当前解决方法

3.1 事实幻觉与可靠性不足

大模型最受关注的问题之一是幻觉，即模型生成的内容在语义上通顺，却与事实、上下文或输入条件不一致^[14]。这一现象在开放域问答、法律文本、医疗建

议和多模态描述中都可能出现。究其原因，一方面，语言模型本质上是在学习条件概率分布，优先生成“统计上像真的内容”；另一方面，训练语料噪声、知识时效性不足以及推理过程缺乏外部验证，也会放大错误输出。

当前解决思路主要包括三类。第一类是检索增强生成，通过从外部知识库检索证据并与参数化模型联合生成，提升事实性与可溯源性^[11]。第二类是偏好对齐和监督微调，通过人工反馈、排序数据和规则约束改善回答风格与正确率^[13]。第三类是多轮自检、思维链或工具调用，让模型在生成前后显式展开中间推理、计算或查询，从而降低“张口就来”的概率。不过，这些方法通常是“缓解”而非“根治”，模型仍可能在证据选择、推理链连接和不确定性表达上出错。

3.2 训练与推理成本过高

大模型的另一核心问题是资源消耗。模型越大，训练时需要的算力、能耗、通信和存储成本越高；而在真实应用中，推理成本往往比训练成本更直接决定可落地性。Transformer 的自注意力复杂度随上下文长度近似二次增长，这在长文本、多模态和实时应用场景中尤为突出^[4,17]。

针对这一问题，研究者提出了多条技术路线。参数高效微调方法如 LoRA 只更新少量参数即可完成任务适配，降低了训练门槛^[12]；模型量化、蒸馏和剪枝有助于压缩部署体积；稀疏 MoE 与线性复杂度架构如 Mamba，则试图从模型结构上缓解长序列和高吞吐压力^[17]。这些方法共同体现出一个趋势：未来神经网络的竞争不只是参数规模竞争，更是单位算力产出效率的竞争。

3.3 数据治理、偏见与安全对齐问题

神经网络和大模型依赖大规模数据，但数据规模的扩大也带来了版权、隐私、偏见与安全问题。训练语料中的刻板印象、地域偏差、噪声标签和敏感信息，都会在模型输出中被继承甚至放大^[9,10,14]。在开放式系统中，提示注入、越狱攻击、有害内容生成和隐私泄露也是实际部署必须面对的风险。

当前解决方法一方面包括数据侧治理，如数据清洗、脱敏、合成数据筛选和高质量对齐语料构建；另一方面包括模型侧和系统侧治理，如 RLHF、内容过滤、拒答策略、权限分层、审计日志和安全沙箱等^[13,19]。值得注意的是，安全问题已经从“模型参数是否被窃取”扩展为“模型输出是否安全、调用链是否可信、外部工具接入是否会引入新攻击面”的系统性问题。因此，可信 AI 需要算法、平台和制度三者共同作用，而不能只依赖某一种训练技巧。

3.4 可解释性、泛化边界与评测困难

神经网络尤其是大模型仍具有明显黑箱特征。即便模型在基准测试中取得高分，也不代表其真正理解了任务机制。模型可能借助训练数据污染、模板记

忆或统计捷径取得表面性能，却在分布变化、组合泛化和极端场景中失效^[9,10,19]。这解释了为什么许多模型在演示中表现惊艳，但在长期稳定使用中仍暴露脆弱性。

围绕这一问题，当前研究正在加强可解释性方法、细粒度评测基准和不确定性估计机制。例如，从激活分析、概念瓶颈、注意力可视化到对抗测试与红队评估，都是为了更清楚地回答“模型为什么这样输出、何时会失败、失败能否提前预警”这三个关键问题。可以说，评测体系本身已经成为大模型时代的重要基础设施。

表 3-1 主要问题与典型解决方法

问题类型	典型表现	当前主要解决思路
幻觉与事实错误	回答似是而非、证据不足、多模态描述失真	RAG、工具调用、链式推理、自检与引用增强
训练和推理成本	训练昂贵、部署困难、长上下文推理缓慢	PEFT、量化蒸馏、MoE、Mamba 等高效架构
偏见与安全风险	有害输出、隐私泄露、提示注入、对齐不足	数据治理、RLHF、规则过滤、审计与权限隔离
解释与评测不足	黑箱决策、泛化边界不清、基准高分但落地不稳	可解释性分析、细粒度评测、红队测试、不确定性估计

4 不同类型神经网络分别适用于解决哪些问题

从任务视角看，“神经网络”并不是单一模型，而是一组带有不同归纳偏置的架构家族。不同任务的数据形态不同，决定了适合的网络结构也不同。理解这一点，有助于避免“只要是 AI 问题就一律用大模型”的误区。

4.1 CNN、RNN 与 GNN：针对特定数据结构的经典强项

卷积神经网络通过局部感受野和权值共享机制，天然适合图像、视频帧和其他网格化数据，因此在分类、检测、分割和医学影像分析中长期保持高效与稳定^[2]。即使在 Transformer 流行之后，CNN 在资源受限、样本规模有限和低时延视觉任务中仍具有明显工程优势。

循环神经网络及其 LSTM、GRU 变体则适合时间顺序强、历史状态重要的序列数据，如语音、传感器信号、金融序列和部分文本任务^[3]。虽然它们在长文本建模上被 Transformer 广泛替代，但在短序列建模、流式处理和对参数规模敏感的场景中仍然具有价值。

图神经网络面向节点和边构成的非欧式结构数据，适用于社交网络分析、推荐系统、知识图谱、交通网络和分子性质预测^[5]。当问题本质上依赖对象之间的关系而非单个样本本身时，GNN 往往比 CNN 或普通 MLP 更自然、更有效。

4.2 Transformer、大模型与生成模型：通用能力与生成能力的核心路线

Transformer 最大的优势在于全局依赖建模与高度并行化，这使它成为自然语言理解、生成、多模态融合和基础模型训练的主干架构^[4,8,10]。几乎所有主流大语言模型和多模态基础模型都以 Transformer 或其变体为核心。当任务需要统一处理长上下文、多任务迁移和提示学习时，Transformer 通常是首选。

生成对抗网络和扩散模型则主要适用于“生成”问题，如图像合成、数据增强、风格迁移、语音生成与分子设计。GAN 生成速度快，但训练不稳定；扩散模型生成质量更高、控制性更强，近年已成为高质量图像和视频生成的主流路径^[6,7]。因此，在需要高保真内容生成时，扩散模型通常优于传统 GAN。

4.3 MLP、混合架构与强化学习结合模型

多层感知机虽然结构简单，但并未过时。它仍是分类器、特征融合器和轻量级预测器的重要组成部分。在表格数据、低维表示学习和部分高效模型中，MLP 因训练简单、部署便捷而被广泛使用。当前越来越多的系统采用混合架构，如 CNN 与 Transformer 结合、图网络与序列模型结合、多模态编码器与语言解码器结合，以兼顾局部特征、全局依赖和跨模态对齐。

此外，当任务不是静态预测，而是“连续决策与行动”时，强化学习与神经网络的结合更具优势。例如机器人控制、自动驾驶策略优化、博弈对抗与资源调度等问题，需要模型在环境反馈中不断更新策略，而不仅仅是学习输入到输出的映射。这说明网络架构的选择不能脱离任务目标、数据结构和部署约束。

表 4-1 主要神经网络架构及适用问题

架构类型	最适合的数据或问题	主要优势	典型局限
CNN	图像、视频、栅格化信号	局部特征提取强，参数共享，高效稳定	对远距离依赖建模不足
RNN、LSTM、GRU	时间序列、语音、流式文本	顺序建模自然，适合历史状态依赖任务	并行性差，长序列训练困难
GNN	图结构、知识图谱、推荐、分子图	关系建模能力强，适合非欧式数据	训练复杂，扩展到超大图较困难
Transformer	长文本、多模态、通用基础模型	全局依赖强，可扩展性好，适配多任务	计算和显存开销大
GAN/扩散模型	图像、音频、视频等生成任务	生成能力强，扩散模型质量高	训练或采样成本较高
MLP/混合模型	表格数据、特征融合、轻量部署	结构简单，部署方便，可作通用组件	对结构先验利用有限

因此，与其问“哪一种神经网络最好”，不如问“当前问题的数据形态、决策方式和资源约束最匹配哪一类神经网络”。这是神经网络工程实践中最重要的架构

意识。

5 神经网络未来可能取得突破的领域

5.1 高效架构与端侧部署

未来数年，一个明确趋势是从单纯追求参数规模，转向追求更高的能力密度和部署效率。面向长上下文与高吞吐需求，线性复杂度架构、稀疏激活模型、模型压缩和硬件协同优化将继续成为热点^[12,17,19]。这不仅关系到云端训练成本，更关系到模型能否真正进入手机、机器人、车载终端和工业边缘设备。换言之，谁能够把强能力压缩到可接受的能耗和时延范围内，谁就更可能定义下一阶段的大模型生态。

5.2 多模态智能体与具身智能

当前模型擅长“看”和“说”，但未来更重要的是“理解环境并采取行动”。具身智能强调模型在真实或仿真环境中的持续感知、规划、执行与反馈闭环，是从静态问答走向主动智能的关键一步。围绕机器人和基础模型的综述研究指出，未来的具身 AI 系统将大量依赖语言、视觉、动作与世界状态的统一表示，以及工具调用、长期记忆和任务分解能力^[18]。这意味着神经网络的研究边界正在从屏幕内的文本与图像，扩展到与物理世界交互的智能体系统。

5.3 AI for Science 与医学智能

在科学发现和医学智能中，神经网络有望扮演更核心的“研究助手”角色。一方面，模型可在海量文献、实验数据和结构信息中发现潜在规律，加速候选筛选和假设生成；另一方面，医学大模型可以继续向专科问答、病程管理、辅助诊断和个性化健康建议推进^[16,19]。不过，这一方向的突破将高度依赖高质量数据、严格评测与人机协同机制。未来真正有价值的并不一定是“全自动的 AI 科学家”，而是能够和研究者、医生形成稳定协作关系的高可信系统。

5.4 可信治理、可解释性与神经符号融合

未来突破还可能出现在“让模型更可信”而非“让模型更大”这一方向。大模型要成为社会基础设施，就必须回答隐私、偏见、责任归属、内容安全和可追责等问题^[9,14,19]。同时，纯粹依赖统计学习的模型在严格逻辑推理、组合泛化和长期规划方面仍存在短板，因此神经网络与符号推理、外部工具系统、规则库和知识图谱的融合，可能成为提升稳定推理能力的重要杠杆。

5.5 脑启发计算与新型计算范式

从更长远看，神经网络的演进还可能受到脑科学、神经形态硬件和新型存算体系的深刻影响。当前深度学习系统虽然名称上借鉴“神经”，但在能效、持续学习和鲁棒性上与生物神经系统仍有明显差距。未来若能在脉冲神经网络、神经形态芯片、记忆机制和持续学习理论上取得突破，神经网络或许会出现一次新的范式转移。这类方向短期内未必直接替代 Transformer 和现有大模型，但很可能成为下一代低能耗智能系统的重要源头。

6 结语

总体而言，神经网络的发展已经从“针对单任务的专用模型”进入“以基础模型为核心的通用智能平台”阶段。大模型显著扩展了神经网络的能力边界，使其在语言、视觉、多模态与专业领域都展现出前所未有的迁移性和生产力；但与此同时，幻觉、成本、偏见、安全与可解释性等问题也提醒我们，能力增长并不自动等于可靠可用。

面向未来，真正决定神经网络上限的，未必只是参数数量，而是模型是否能够在更低成本下实现更高质量推理，是否能够与环境、工具和人类协同工作，是否能够在高风险场景中做到可信、可控与可审计。可以预见，神经网络与大模型的下一阶段竞争，将是效率、行动力和可信度的综合竞争。

参考文献

- [1] McCulloch W S, Pitts W. A Logical Calculus of the Ideas Immanent in Nervous Activity[J]. Bulletin of Mathematical Biophysics, 1943, 5(4): 115–133.
- [2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[C]. Advances in Neural Information Processing Systems, 2012: 1097–1105.
- [3] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735–1780.
- [4] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[C]. Advances in Neural Information Processing Systems, 2017, 30: 5998–6008.
- [5] Zhou J, Cui G, Hu S, et al. Graph Neural Networks: A Review of Methods and Applications[J]. AI Open, 2020, 1: 57–81.
- [6] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets[C]. Advances in Neural Information Processing Systems, 2014, 27: 2672–2680.
- [7] Ho J, Jain A, Abbeel P. Denoising Diffusion Probabilistic Models[J/OL]. arXiv:2006.11239, 2020.
- [8] Brown T B, Mann B, Ryder N, et al. Language Models are Few-Shot Learners[J/OL]. arXiv:2005.14165, 2020.
- [9] Bommasani R, Hudson D A, Adeli E, et al. On the Opportunities and Risks of Foundation Models[J/OL]. arXiv:2108.07258, 2021.
- [10] Zhao W X, Zhou K, Li J, et al. A Survey of Large Language Models[J/OL]. arXiv:2303.18223, 2023.
- [11] Lewis P, Perez E, Piktus A, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks[J/OL]. arXiv:2005.11401, 2020.
- [12] Hu E J, Shen Y, Wallis P, et al. LoRA: Low-Rank Adaptation of Large Language Models[J/OL]. arXiv:2106.09685, 2021.
- [13] Ouyang L, Wu J, Jiang X, et al. Training Language Models to Follow Instructions with Human

- Feedback[J/OL]. arXiv:2203.02155, 2022.
- [14] Huang L, Yu W, Ma W, et al. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions[J/OL]. arXiv:2311.05232, 2023.
 - [15] Kirillov A, Mintun E, Ravi N, et al. Segment Anything[J/OL]. arXiv:2304.02643, 2023.
 - [16] Singhal K, Azizi S, Tu T, et al. Large Language Models Encode Clinical Knowledge[J]. Nature, 2023, 620: 172–180.
 - [17] Gu A, Dao T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces[J/OL]. arXiv:2312.00752, 2023.
 - [18] Xu Z, Wu K, Wen J, et al. A Survey on Robotics with Foundation Models: toward Embodied AI[J/OL]. arXiv:2402.02385, 2024.
 - [19] Stanford Institute for Human-Centered AI. AI Index Report 2026[EB/OL]. <https://hai.stanford.edu/ai-index/2026-ai-index-report>.
 - [20] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J/OL]. arXiv:2010.11929, 2020.