**附件4：**

**2025年度中国电子学会-腾讯**

**博士生科研激励计划（混元大模型专项）重点支持方向**

1. **语言大模型研究**

语言大模型主要开展语言模型的预训练、强化学习等方向的研究。其中预训练基础算法层面包括高效模型结构、训练策略、Scaling Law、数据合成等。强化学习基础算法层面包括RLHF/RLAIF等，涵盖从人类反馈学习到大模型自我对弈学习等，也包含广泛的智能问题，例如：长文、创作、代码、数学、情感、情商等，以及广泛的应用问题，例如：情感陪伴智能体、AI搜索、智能客服等。

**二、多模态理解大模型研究**

多模态能力是大模型感知真实世界和走向物理世界的基础，多模态理解大模型主要构建具备全面和可靠理解视觉信号的多模态大模型，具体研究内容包括图像视频数据搜集和清洗，视觉encoders，多模态预训练，强化学习，深度推理模型等。

**三、语音大模型研究**

语音是人与人之间，以及人与大模型之间交互和交流的重要模态，GPT-4o是其中一个里程碑，将语音原生融入语言大模型中。语音大模型主要构建全面可靠的具备语音理解、生成和对话能力的语音大模型，具体研究内容包括语音数据清洗，语音encoders和decoders，语音模态预训练，强化学习，ASR，TTS，语音通话，音乐生成等。

**四、视觉生成大模型研究**

视觉生成大模型主要开展能够生成高质量、逼真、且遵循语言指令的图像、视频或3D的模型研究。典型的生成模型包括生成对抗网络（GANs）、变分自编码器（VAEs）、自回归模型，扩散模型等。具体研究内容包括视觉数据的收集与清洗、视觉生成模型的架构设计、生成过程中的控制与调节、风格迁移等，以及艺术创作、广告设计、电影制作、虚拟现实、游戏开发等领域的应用研究。

**五、大模型训练和推理优化研究**

大模型训练和推理优化主要开展提高大模型训练和推理的效率和效果的研究。大模型训练的研究内容包括分布式训练、模型并行，低精度训练、优化算法的改进等，涉及大规模数据的高效处理与存储、训练过程中的资源调度与管理、训练速度与精度的平衡、训练过程中的稳定性与鲁棒性等。大模型推理的研究内容包括模型量化、模型剪枝、推理加速器的设计与优化、推理过程中的资源调度与管理等。具体课题涉及推理速度与精度的平衡、推理过程中的能耗优化、推理结果的稳定性与鲁棒性等，以及大模型在边缘设备、移动设备等资源受限环境中的推理优化策略研究。

**六、超级智能和超级对齐研究**

超级智能和超级对齐主要开展探索和开发超越人类智能的大模型技术，以及如何确保超级智能系统的行为与人类价值观和目标保持一致等研究。超级智能的研究内容包括通用人工智能（AGI）的理论与实现、超大规模模型的训练与优化、模型的自我学习与进化、智能体的自主决策与行为等。一些超级智能的例子也包括使大模型具备解决奥赛难度问题的能力，以及让其具备科研能力等。超级对齐的研究内容包括智能体的价值对齐、智能体的行为规范与约束、智能体的透明性与可解释性、智能体的安全性与可靠性等。近期GPT-o1和Deepseek-R1等深度推理模型，以及通用agents的探索，均是迈向超级智能和超级对齐较有潜力的方向，这些方向均还有较多开放性问题需要进一步研究。