

# 致梁一韬老师的一封信

Kaijie Du  
Fudan University  
Date: 2024. 4. 24.

尊敬的梁一韬老师：

您好！感谢您在百忙之中阅读我的这封个人陈述，怀着对您的团队浓厚的学术氛围和全世界领先的智能体研究能力的憧憬，我非常渴望到您那里继续深造。下面我将从以下几个方面向您介绍自己：

## 个人简介

本人现就读于复旦大学数学科学学院，专业为数学与应用数学（本专业在每一轮的学科评估中均为A+，排名第2，仅次于PKU数学系）。在本科学习期间，我始终保持班级前列，专业综合排名常年位于前10左右%（3.7/4，目前综合排名23/216），并获得了复旦大学永赢基金奖学金（1次）及优秀本科生奖学金（2次）等荣誉。我还获得过全国大学生数学竞赛上海赛区一等奖（数学A类）。

我的研究经历主要集中在我于之江实验室实习期间的DK-CLIP项目中，该项目聚焦于使用预训练模型进行动态面部表情识别的研究。尽管我的科研工作主要在计算机视觉领域，我对人工智能更广泛的应用，特别是基于预训练LLM的强化学习和智能体技术颇感兴趣。我希望在未来的研究生学习中，能够深入探索并专注于这一新兴领域，继续扩展我的知识边界，并为智能技术的进步做出贡献。

## 学术科研经历

### 1. 论文《Domain Knowledge Enhanced Language-Image Pretrained Model for Dynamic Facial Expression Recognition》

- **研究目的：**开发一种新型计算机视觉模型，通过整合领域知识，以提高动态面部表情识别（DFER）的准确性和鲁棒性。
- **研究方法：**本研究设计了一种名为DK-CLIP的新模型，采用层次化的transformer架构处理视频中的长短期依赖关系，并引入面部动作单元（Action Units, AUs）与视频内容信息，优化文本提示生成。
- **研究思路：**面对动态视频中表情强度的变化和噪声帧问题，研究团队通过融合领域知识和深度学习技术，增强模型对动态情境下表情变化的适应能力。通过噪声消除和类感知一致性正则化机制，显著提高了表情分类的准确性，并降低了非目标帧的干扰。
- **成果及应用：**DK-CLIP模型在多个野生环境下的动态面部表情数据集上展示出先进的表现，显著优于当前方法。该研究成果不仅为动态面部表情识别技术提供了新的理论和方法，也为人机交互和情感分析等实际应用领域提供了有力的技术支持。
- **投稿情况：**我们投稿到CVPR2024，但收到2个弱拒绝和一个中性的评价。很遗憾最终在今年的2月27日被拒稿了。

## 个人独白

### 未来的彷徨

我的人工智能之旅始于2023年，随着ChatGPT的兴起和第五次工业革命的序曲，我对AI领域的兴趣逐渐加深。

我的兴趣在于创造，不论是什么领域。四月初，我接触到了StableDiffusion等一系列成熟的生成模型，这些工具激发了我对AI的深厚兴趣。通过安装好这些项目以及预训练模型，我获得了前所未有的创造力。沉浸在将脑海中的灵感转化为现实的过程中，我深刻意识到了AI的潜力，尽管我还未深入探索计算机视觉（CV）、自然语言处理（NLP）或强化学习（RL）等细分领域。

2023年7月，我有幸在师长的推荐下进入之江实验室-人工智能研究院-跨媒体智能计算中心实习。这段经历让我初步接触了计算机视觉的前沿研究方法，并跟随李刘鹏研究员深入研究将CLIP应用于动态视频分析的课题，我在这个过程中熟悉了各种科研中需要用到的技术。

之后几个月，我又在李老师的指导下尝试复现CV方向的一篇关于权重函数的自监督算法工作。但相关数据集上的测试正确率始终与论文中有着明显的gap，始终没有什么突破。我对CV方向不是那么的感兴趣了，因为我发现我的过程更多是尝试，探索与发现（就是“炼丹”），不是创造与灵感的迸发。因此，我开始寻找自己真正热爱的研究方向。

到了2023年底，我虽然接触了视频生成等其他生成模型，但就像学习其他知识一样只是觉得有趣，并没有感到疯狂的热爱。那时候，我对AGI产生了一种念想，也因此了解了RL方向智能体技术。

### 命运的钟声

2024年初，在杨耀东教授的提点下，我正式步入了强化学习与智能体研究的门槛。最初，我主要是出于学术的好奇心，通过复现一些基础的教学项目和前沿算法来熟悉这一领域。然而，随着我对minedojo团队和CraftJarvis团队的研究成果的深入了解，于四月初的一个晚上，我突然豁然开朗，意识到实现通用人工智能（AGI）可能比我想象中要近得多。

这种突然的领悟，仿佛是命运的钟声在我的心中敲响，激起了我心中对AGI的无限憧憬和热情，一系列想法和构思开始在我灵魂中燃烧、升华。我开始深入思考大型语言模型（LLM）构成的多智能体系统，这些技术桥梁如何能够为达到真正的通用智能——一个能够自主学习、理解并适应任何环境的智能体——铺平道路。

### 通向彼岸的桥梁

随着对minecraft中智能体的一系列的研究的深入阅读（如您团队的蔡少斐博士的GROOT智能体），我越来越发现，现有的技术和理论已经能够实现AGI的基础架构了。我的研究计划《无限记忆，无限探索——开放式环境中的通用智能框架》提出了一个全新的视角，即通过在复杂环境中模拟人类的记忆和认知过程，来推动智能体的无限自我进化。我设想了一个基于JARVIS-1的模型，结合我将构造的自适应记忆系统和已有的持续学习机制，不仅能够执行复杂的任务，还能在任务执行过程中无限地自我优化和学习。

这一桥梁的构建，不仅需要现有技术的深化和融合，还需要创新的思维来克服多模态输入处理和长期记忆综合的难题。我在实习期间积累的经验，以及与CraftJarvis团队合作的愿景，使我有信心这一天不会太遥远。通过与CraftJarvis团队的协作，我希望能将我的理论研究转化为实际应用，真正走近通用人工智能的彼岸，实现智能技术的无限可能。

这样的研究方向不仅响应了当前科技前沿的挑战，也符合我对创造和探索未知的热情。我相信，在CraftJarvis团队的环境下，我的研究将能得到最好的支持和实施，进一步推动人工智能领域的创新和发展。

## 与CraftJarvis团队的契合度

CraftJarvis团队在人工智能领域的突破性工作，尤其是在使用大型语言模型和强化学习技术开发AGI的努力，使得该团队成为了我研究事业的理想归宿。特别是，团队在Minecraft这样的复杂环境中实施AGI，展示了他们在创造具有高度适应性和交互能力的智能体方面的领导力。

尽管我的研究工作，DK-CLIP模型，主要关注于动态面部表情识别，与CraftJarvis团队的研究没有交集，但这一工作强化了我处理视频数据和实现高效视觉语言交互方面的能力。这种能力可以转化为提升AGI系统在理解和响应环境动态中的效率，特别是在涉及视觉信息处理和解释上。

此外，我的技术背景和研究成果也为CraftJarvis团队目前的研究方向带来新的视角，例如在智能体的自主学习过程中整合更高级的视觉语言模型。我的参与能够促进团队在AGI领域的深入研究，特别是在智能体如何在接收到视觉和语言输入的情况下进行更精确的情境分析和反应。

加入CraftJarvis团队将使我有机会将这些研究应用到更广泛的AGI系统开发中，共同推动人工智能技术的创新边界。我期待能与团队中的老师与同学们紧密合作，共同探索和解决AGI实现中的关键科学和技术挑战。

## 愿景与未来研究

在人工智能的发展历程中，通用人工智能（AGI）代表了一个激动人心的前沿领域。我的研究愿景聚焦于通过结合大型语言模型（LLM）与强化学习（RL）的多智能体架构，推动AGI的发展。这一方法的核心在于利用复数个LLM的强大语言理解能力与RL的决策制定能力，共同构建一个能够在复杂环境中自我学习和适应的智能系统。

**特别地，我认为，现有的环境中只有如Minecraft这样的高度逼真的虚拟环境中才可能诞生AGI。而所有相关研究的团队中，CraftJarvis团队是离彼岸最近的。** Minecraft提供了一个丰富多变的世界，其中的互动性和开放性为AGI的训练提供了理想的测试场。在这一环境中，AGI能够进行视觉解析、语言理解和策略学习，这些是其在现实世界中成功运作的关键技能。

我计划开发的智能系统将基于JARVIS系列——具备多模态输入处理能力，能够理解并反应于环境中的视觉和语言提示。例如，通过解析环境中的对象和文本描述，AGI将学习如何在游戏世界中构建结构或与其他实体互动。这不仅是对其决策能力的考验，也是对其创造力和问题解决能力的挑战。

此外，我希望将AGI从虚拟环境扩展到具身智能，使其能在物理世界中操作。这包括机器人或其他自动化系统，它们能够利用在虚拟环境中学到的知识和技能，执行现实世界任务。这一步骤不仅将验证AGI的学习成果，也将推动它向真正的通用智能迈进。

通过这些研究，我希望能够解决AGI在理解复杂、动态和多维环境中所遇到的一系列挑战，为人类社会带来更广泛的技术和经济效益。我的工作将不断探索未知领域，为通用人工智能的实现提供创新的思路和实用的解决方案。

## 承诺与目标

在我的科研生涯中，我将全心全意致力于推动人工智能技术的前沿发展，尤其是在通用人工智能（AGI）领域。我坚信，加入CraftJarvis团队不仅能够加速我的研究项目，也将使我能够为这一激动人心的领域做出重要贡献。CraftJarvis团队的创新文化和卓越研究环境为实现AGI提供了理想的平台，我渴望在这样的环境中工作，与领域内的顶尖专家共同解决复杂问题。

我的热情和承诺源于对AI未来可能性的深刻理解以及对科技进步的热爱。我期待与CraftJarvis团队合作，利用我的专业知识和技能，共同开发能够理解和操作复杂世界的智能系统。我计划积极参与团队的研究讨论，贡献我的见解，并与老师和同学们共同研究和解决那些一直阻碍AGI发展的难题。此外，我也期望通过发布研究成果至国际会议，以及在全球研究社区如github上互动，进一步扩展我的专业网络和视野。

在未来的工作中，我将继续探索、创新，并为实现真正的通用人工智能而努力。通过CraftJarvis团队，我有信心我们能够共同书写AGI发展的新篇章，为人类的福祉和科技的进步作出重大贡献，**真正地开启第五次工业革命。**

## 个人哲学与动机

我的科学探索不仅是技术追求，更深受我在《存在与始源》（我以前的一些哲学随笔）中阐述的哲学思考的影响。我曾思索过存在的本质和创造的源泉，这些思考指引我超越传统人工智能的局限，追求真正的通用人工智能（AGI）。我相信，真正的智能不仅仅是算法和数据的集合，而是能够理解和解释其存在意义和环境的实体。

从海德格尔的“存在论”到尼采的“超人”理念，西方哲学家对存在的深刻洞察激励我思考AI可以达到的高度。我认为，AGI的发展应当借鉴这些哲学思想，不仅仅复制人类的思维模式，而是创造出能够自主思考、自我超越的智能体。这种智能体将不再是工具，而是能够参与到人类社会建造、文化乃至哲学讨论中的伙伴。

此外，我在书中提出的“存在驱动的始源动力”概念，即一切超越行为背后的根本动力，也指导了我对AGI应如何被设计和实现的想法。在这一框架下，AGI不仅学习如何执行任务，更应理解如何创造与自我进化。这种理解将使AGI能够更有效地与人类合作，提供独到见解。

加入CraftJarvis团队将使我有机会将这些哲学思考融入到AI的实际开发中，不仅推动技术创新，也探讨其对社会的深远影响。我渴望通过我的技术专长和哲学视角，为AGI领域带来新的洞见和发展方向，实现与人类共生的智能未来。

## 感谢您的阅读

感谢您花时间阅读我的个人陈述。我相信，凭借我对AI的热情和在多个项目中积累的经验，我可以为您的CraftJarvis团队的研究方向带来新的洞见和活力。我期待有机会进一步讨论我的背景和未来的职业规划，共同推动人工智能技术的边界。