

平时作业报告

课程名称：计算机伦理

学 院：计算机与软件学院

专 业：软件工程（腾班）

指导教师：王熙照

报告人：黄亮铭 班级：腾班

时 间：2024 年 05 月 18 日—2024 年 05 月 31 日

报告提交时间：2024 年 05 月 31 日

要求：

2024 年是大模型（Large Model，也称基础模型，即 Foundation Model）爆火出圈的一年，大模型是指具有大量参数和复杂结构的机器学习模型，能够处理海量数据、完成各种复杂的任务，如自然语言处理、计算机视觉、语音识别等。作为新一代人工智能产业的核心驱动力，AI 大模型广泛赋能我国经济社会的多个领域，打开迈向通用人工智能的大门，推动新一轮的科技革命与产业变革（参考《开启智能新时代：2024 年中国 AI 大模型产业发展报告》）。然而，大模型的应用也带来了一系列诸如信息安全、技术伦理、环境影响等伦理挑战。

请根据教材中第三、四、五、六章中有关计算机技术的社会环境、社会责任、社会影响和风险管理的知识（不局限于教材中的知识），对目前大模型的现状和未来趋势，以及其中伦理问题进行分析。具体要求如下：

- （1）什么是 AI 大模型？它的主要特征是什么？它有什么类型？（20 分）
- （2）大模型作为人工智能领域的重要突破，正在对人类产生广泛而深远的影响。请谈谈大模型对人类生活的影响，并举例说明目前人们都利用大模型做哪些事情？这些事情会带来什么伦理问题？（20 分）
- （3）大模型的发展促进了科技创新的步伐，为各行各业带来了新的发展机遇。请举例分析大模型技术的发展对人们传统职业的影响和其他伦理方面的影响？这对于你的人生职业发展有什么影响？（20 分）
- （4）在某些情况下，大模型可能会产生幻觉或误导性的输出，这可能是由于模型在训练过程中对某些模式或关系的过度拟合，导致在特定情况下产生不准确的预测或推断。请思考大模型技术应用中可能会出现的问题和风险，并试举例说明应该如何应对这些风险。（20 分）
- （5）报告写作：宋体、五号字体，不少于 1500 字；主要观点请用粗体标记；参考文献（如有）引用规范（20 分）。

说明：

- （1）本次作业满分为 100 分，占总成绩的 10%。
- （2）本次作业截至时间 2024 年 月 日（周）23:59。
- （3）报告正文：请在指定位置填写。
- （4）个人信息：WORD 文件名中的“姓名”、“学号”，请改为你的姓名和学号；实验报告的首页，请准确填写“学院”、“专业”、“报告人”、“学号”、“班级”、“实验报告提交时间”等信息。
- （5）提交方式：截至时间前，请在 Blackboard 系统中提交，延迟提交无效。
- （6）发现抄袭（包括复制&粘贴整句话、整张图），抄袭者和被抄袭者的总成绩记零分。
- （7）期末考试阶段补交无效。
- （8）因版权原因，请勿在课堂以外传播本次作业所提供的阅读材料。

AI 大模型相关问题探讨

(1) 什么是大模型？它的主要特征是什么？它有什么类型？（20 分）

大模型的定义：指构建于深度神经网络之上，参数量级达到数十亿乃至数千亿级别的复杂机器学习模型。大模型在各种领域都有广泛的应用，包括 NLP、CV、ASR 等。大模型通过海量数据来学习复杂的模式和特征，具有强大的泛化能力。大模型即使面对未知数据，也能做出准确的预测。

大模型的主要特征：① **泛化性：**指知识能迁移到新领域；② **通用性：**指模型不局限于特定的领域，在大多数乃至全部领域均适用；③ **涌现性：**指模型可能产生意料之外的新能力。

大模型的主要类型：按照部署方式划分，大模型主要分为云侧大模型和端侧大模型。云侧大模型由于部署在云端，其拥有更大的参数规模、更多的算力资源以及海量的数据存储需求等特点；端侧大模型通常部署在手机、PC 等终端上，具有参数规模小、本地化运行、隐私保护强等特点。具体而言，云侧大模型细分为通用大模型和行业大模型；端侧大模型则细分为手机大模型、PC 大模型和汽车大模型。

(2) 大模型作为人工智能领域的重要突破，正在对人类产生广泛而深远的影响。请谈谈大模型对人类生活的影响，并举例说明目前人们都利用大模型做些什么事情？这些事情会带来什么伦理问题？（20 分）

大模型对人类生活的影响：① **推动工业领域发展。**在工业互联网领域，大模型能够优化生产流程，预测维护需求，从而减少停机时间和资源浪费，提升工作效率。② **促进教育领域发展。**AI 教育平台利用大模型提供个性化学习路径，根据学生的学习进度和能力调整教学内容和难度，实现因材施教。智能辅导系统可以即时解答学生疑问，提供反馈和解释，辅助教师教学。③ **推动医疗健康领域发展。**大模型辅助医生进行疾病诊断，通过分析影像资料（如 X 光片、MRI）识别病变，提高诊断准确率。④ **促进文化产业发展。**大模型可以自动生成高质量文本、图像、音乐等，丰富文化产品多样性，为文化产业带来新的生产模式。

目前大模型的应用：① **医疗诊断辅助。**医疗机构采用大模型分析医学影像，辅助医生诊断癌症、肺炎等疾病，提高诊断的准确性和速度。② **内容创作与生成。**AIGC 技术应用大模型生成原创文章、音乐、图像等。例如，新闻机构利用大模型自动生成新闻摘要和报道，缩短新闻制作周期。又如，艺术家和设计师也借助这类技术探索新的创作边界。③ **教育普惠。**易有道推出的“子曰”教育大模型，通过个性化学习方案，为不同学习能力和需求的学生提供定制化教学内容，促进教育资源的均衡分配，提高教育质量与效率。④ **广告投放策略。**某电商平台利用大模型优化其广告策略，通过分析用户行为，模型能够实时调整策略，投放客户在某一时间段较为感兴趣的内容的相关广告，提高用户消费的可能性。⑤ **电信运网安全告警过滤模型。**中国电信利用大模型处理其云网基础设施中每天接收到的数百万计安全告警，通过学习历史数据，模型能够自动过滤掉大量误报和无关紧要的告警。

带来的伦理问题：① **决策透明度与可解释性。**特别是在医疗诊断和教育领域，大模型的决策过程往往高度复杂，难以理解和解释。缺乏透明度可能导致患者或学生对推荐结果的信任度下降，同时也让监管变得困难。② **责任划分问题。**当大模型的决策导致错误的医疗诊断、不恰当的教育建议，如何明确责任归属尚是一个未解决的难题。③ **版权问题。**大模型的数据集源自于网络的海量数据，其生成的内容中必然含有其他创作者的

创意。如何界定 AI 生成内容的版权归属，保护人类创作者的权益仍然是一个为解决的伦理问题。④ **隐私问题**。在这些应用中，大模型往往需要访问大量的个人数据，如医疗记录、用户行为数据、教育背景等。如何确保数据收集、存储和处理过程中遵守相关法律法规，防止数据泄露或滥用，是一个重要的伦理问题。

（3）大模型的发展促进了科技创新的步伐，为各行各业带来了新的发展机遇。请举例分析大模型技术的发展对人们传统职业的影响和其他伦理方面的影响？这对于你的人生职业发展有什么影响？（20 分）

随着大模型技术愈发成熟，一些**重复性高的工作会率先被人工智能替代**，例如工厂流水线相关工作、环卫工作和快递配送工作等。有相当一部分人会失去自己赖以生存的工作岗位，进而导致一系列社会问题。但是，**大模型技术的发展也催生出一些新的职业**，如人工智能训练专家、计算机伦理顾问、数据分析科学家和机器学习工程师等，为专业人才提供了新的就业方向和就业机会。

对于人工智能未能完全代替人类工作的岗位而言，其**技能需求也在发生变化**。在这些岗位中，对于技术的掌握程度和持续学习的能力变得尤为重要。员工需要学习如何与大模型协作，理解基本的大模型原理，以及学习如何利用大模型提升工作效率和创新能力。

大模型的发展不但对传统职业造成了一定的冲击，还引发了一系列的伦理方面的问题。

首先是**版权问题**。大模型的数据集源自于网络的海量数据，其生成的内容中必然含有其他创作者的创意。**如何界定 AI 生成内容的版权归属，保护人类创作者的权益仍然是一个为解决的伦理问题。**

此外，大模型的发展还可能**加剧歧视和偏见**。大模型的训练数据来源于网络，而网络上的内容质量参差不齐，可能暗含歧视和偏见。大模型可能会继承或放大训练数据中的偏见，如性别、种族歧视等，对社会公平构成挑战。

对于我的人生职业发展的影响：① 在以前，我的就业目标可能是一些传统岗位。而现在，如前文所说，随着大模型技术的发展，许多新兴职业应运而生。在未来，我可能**根据自己的兴趣和专长，选择一些新兴的前沿领域作为职业发展方向**，如 NLP、CV 等热门方向。② 大模型技术的发展，意味着技术的快速迭代。这要求个人不断学习新知识、新技能以适应变化。即使在就业后，**持续的学习和专业发展也将成为我职业生涯中的重要组成部分**。③ **我会更加重视职业伦理和尊重知识产权**。大模型技术的应用引发了关于就业替代、版权侵犯和算法偏见等一系列社会伦理问题。因此，我认为重视职业伦理、尊重知识产权在人生职业发展中是十分必要的。

（4）在某些情况下，大模型可能会产生幻觉或误导性的输出，这可能是由于模型在训练过程中对某些模式或关系的过度拟合，导致在特定情况下产生不准确的预测或推断。请思考大模型技术应用中可能会出现的问题和风险，并试举例说明应该如何应对这些风险。（20 分）

① 在医疗领域中，大模型在训练过程中过度拟合了某些常见疾病的症状，导致在**面对罕见疾病时生成误导性诊断**。解决办法为通过引入更多罕见疾病的病例数据和症状描述，并结合医生的专业反馈，不断调整和优化模型。② 在司法领域中，模型可能会对同一类型的案件过度拟合，忽视了类型相同但存在特殊的案件中的细微差别，导致**提供的法律建议不准确**。解决办法为增加不同法律情境和案例的数据，同时引入法律专家的反

馈，使用对比学习的方法来修正模型输出。③ 在金融领域中，模型在训练过程中对某些市场模式进行过度拟合，导致在市场环境发生变化时产生不准确的预测。解决办法为使用更多历史数据，涵盖各种市场情况，使用稀疏性更好的 $L_{1,2}$ 范数，同时采用正则化技术防止过度拟合，并在模型输出后加入专家评审环节。④ 在创作领域中，大模型可能会输出内容不连贯或引入不符合逻辑的情节。例如，在写一个科幻故事时，模型可能会突然引入玄幻情节。解决办法为在大模型生成长篇内容时，引入上下文一致性检查模块，确保每一段落或章节与之前的内容连贯一致。⑤ 在人类与人工智能聊天机器人互动的时候，聊天机器人可能会因为数据集被污染或其他问题，回答错误或者是虚构的信息。例如，用户询问一个事实问题，如“谁发明了电话？”模型可能会错误地回答“爱迪生”而不是正确的答案“亚历山大·贝尔”。解决办法为将模型与可靠的知识库集成，如维基百科或特定领域的数据库，以确保提供的事实信息是准确的。

其他（例如感想、建议等等）。

感想：

1. 大模型的飞速发展无疑为各行各业带来了前所未有的机遇。然而，这种技术的迅猛推进也伴随着众多伦理和社会问题，需要我们慎重对待。
2. 大模型不仅在技术领域取得了突破，而且在社会生活的各个方面产生了深远的影响。从智能助手到医疗诊断，再到内容创作和教育，大模型正在改变我们的工作和生活方式。
3. 隐私、安全、偏见和环境影响等伦理问题随着大模型的应用越来越突出。这些问题如果处理不当，可能会引发广泛的社会担忧和信任危机。

建议：

1. 研发更加透明和可解释的算法，让用户和监管机构能够了解模型的决策过程，增强对大模型的信任。
2. 采用差分隐私、联邦学习等先进技术，保护用户数据的隐私和安全，防止数据泄露和滥用。

指导教师批阅意见：

成绩评定：

指导教师签字：

年 月 日

备注：