**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： 计算机科学与技术（创新班）**

**指导教师： 王熙照**

**报告人： 何泽锋 班级： 高性能特色班**

**时 间： 2024年5月17日—2024年5月20日**

**报告提交时间： 2024年5月20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点（不局限于教材中的知识点），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）**  **大模型是指具有大规模参数和复杂计算结构的机器学习模型**。这些模型通常由深度神经网络构建而成，拥有数十亿甚至数千亿个参数。这类模型能够处理复杂的任务，如自然语言处理、图像识别、语音识别等，并且在很多应用场景中展现出超越传统算法的性能。  大模型的主要特征之一是**参数量巨大**，可以达到数亿、数十亿甚至更多。这些参数在模型训练过程中通过学习数据进行调整，以实现对数据特征的捕捉和任务目标的达成。其次，大模型的计算资源需求高，由于参数量大，大模型在训练和推理时需要大量的计算资源，包括高性能的CPU、GPU以及专业的AI处理器等。此外，大模型的**数据需求大**，为了训练如此庞大的模型，往往需要大量的标注数据进行监督学习。  **结合《2024年中国AI大模型产业发展报告》**还可以总结出以下观点大模型**具有泛化性、通用性以及涌现性特征**。以 ChatGPT 为代表的 AI 大模型因其具有巨量参数和深度网络结构,能学习并理解更多的特征和模式,从而在处理复杂任务时展现强大的自然语言理解、意图识别、推理、内容生成等能力,同时**具有通用问题求解能力**,被视作通往通用人工智能的重要路径。  按照部署方式划分，AI 大模型主要分为云侧大模型和端侧大模型两类。  **云侧大模型分为通用大模型和行业大模型。**从云侧大模型来看，通用大模型具有适用性广泛的特征，其训练数据涵盖多个领域，能够处理各种类型的任务，普适性较强。行业大模型 具有专业性强的特点，针对特定行业的需求进行模型训练，因而对特定领域具有更深的业务理解和场景应用能力。  **端侧大模型主要有手机大模型、PC大模型。**从端侧大模型来看，手机和PC大模型由于直接部署在设备终端，让用户体验到更加个性化和便捷的智能体验。  **（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）**  大模型产生的影响可以从多个方面体现。**在医疗领域，**大模型的应用可以帮助医生更准确地诊断疾病，例如通过分析医疗影像来辅助诊断。**在教育领域**，大模型可以为学生提供个性化的学习建议，例如智能教育系统可以根据学生的学习进度和能力，推荐适合的学习内容和资源。**在交通出行方面，**大模型在自动驾驶技术中发挥着重要作用，可以提高驾驶安全性并减少交通事故。**在金融服务方面**，大模型可以用于信用评估、风险管理等方面，帮助金融机构更好地服务客户。可以看到，大模型已经融入了人们生活的方方面面。  大模型的大量应用难免会带来了一些伦理问题。首先，**数据隐私问题**是大模型面临的一个重要挑战，由于大模型需要大量的数据进行训练，可能会导致个人隐私信息的泄露。其次，**模型偏见问题**也需要引起关注。大模型的训练数据可能存在偏见，这会导致模型在做出决策时也表现出偏见，可能导致不公平的决策结果。最后，**模型滥用问题**也是主要问题，大模型可能被用于不当目的，例如制造虚假信息、进行网络攻击等，这可能对社会安全造成威胁。  **（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）**  大模型技术的发展对人们传统职业产生了深远的影响。一方面，大模型的出现使得许多**传统职业的工作方式发生了改变，提高了工作效率。**例如，在医疗领域，大模型可以帮助医生更准确地诊断疾病，提高治疗效果。在新闻编辑领域，大模型可以自动生成新闻报道，减轻编辑的工作负担。在金融服务领域，大模型可以用于信用评估和风险管理，提高金融机构的服务效率和风险控制能力。  然而，大模型技术的发展也带来了一些伦理方面的挑战。首先，大模型可能导致部分**传统职业的岗位需求减少，从而引发就业结构的变化**。例如，一些重复性较强的工作可能会被大模型取代，这要求从业者提高自身技能，适应新的职业环境。其次，大模型技术在处理数据时可能涉及**个人隐私问题**，如何在保护隐私的前提下发挥大模型的效益，是一个亟待解决的问题。正如计算机技术所产生的影响，大模型的应用同样**面临着法律的空白**，没有完善的规则约束大模型的使用。同时道德难以延申到网络虚拟世界，这些也都是伦理方面的问题。  对于我个人而言，这显然为我的个人职业发展产生多方面的影响。首先大模型的大规模应用让我知道，我需要提高个人的职业水平和技术能力，防止被技术发展所淘汰。再者我要提高个人的职业道德，需要规范个人行为，遵守“**一般道德守则**”。同时还要担负起责任，包括**对自然界、对社会发展和对人的责任**，要履行相应的义务，为推动社会发展做贡献。  **（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）**  大模型**预测结果不准确**，虽然在训练数据上表现良好，但在实际应用场景中遇到与训练数据分布不一致的情况时，模型的性能可能会显著下降。这可能是因为**数据类型不全面，信息可信度不高**等问题。而对于国内的大模型来说也是面临这一问题，可用于大模型训练的中文数据库体量严重不足，这就导致了应用时不能很好的处理各种语境下的问题。更为严重者，**模型偏差和公平性**也是一个重要的问题。大模型可能会从训练数据中学习并放大偏见，导致对特定群体的不公平对待。大模型可能会从训练数据学习到某些用户的隐私信息，进而导致用户信息的泄露和安全问题，这也是受到人们日益关注的风险问题。  为了解决和处理大模型的各项问题和风险，我认为首先需要**明确边界**。大模型归根揭底是通过数据进行训练，我们需要严格控制训练数据的使用，使用什么类型的数据，是否符合道德和法律这需要得到规范化处理。举例来说，例如训练医疗模型，需要保证使用的训练数据得到患者的同意，并且需要确保患者的隐私安全。诸如此类，风险防控需要多方面的共同支持，政府需要有相关法律限制，同时机构也需要严格遵守规则，从源头减少风险。  **其他（例如感想、建议等等）。**  通过本次讨论，我对当前大模型的利弊以及风险、伦理问题有了更深入的思考，相交与传统的计算机技术问题，AI的使用更有可能脱离人的控制，因此也需要有更为严格和完善的规定。计算机技术发展迅速，作为一名计算机学者，我也更加有危机意思，要提升个人能力，及时接触前沿技术。 |

|  |
| --- |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  年 月 日 |
| 备注： |