

# 深圳大学考试答题纸

(以论文、报告等形式考核专用)  
二〇二四~二〇二五学年度第二学期

课程编号	1504550001	课程名称	计算机伦理	主讲教师	朱庆灵	评分	
学号	2023150001	姓名	李文俊	专业年级	2023 级计算机科学与技术(创新班)		

教师评语：

一、生成式人工智能与数字人的社会影响分析 (20 分)：

二、对生成式人工智能的监管与自由的平衡讨论 (15 分)：

三、从计算机伦理角度分析技术所涉及的关键伦理问题 (30 分)：

四、个人观点与案例分析 (25 分)：

五、报告写作与格式规范 (10 分)：

题目： 人工智能生成内容 (AIGC) 与数字人的伦理挑战与社会责任分析

说明：

- 不要删除或修改蓝色标记的文字，也不要删除线框。
- 请在相应的线框内答题，答题时请用 5 号、宋体、黑色文字、单倍行距。

题一（20 分）、题二（15）、题三（30 分）、题四（25 分）、报告格式和总体撰写情况（10 分）

### 一、生成式人工智能与数字人的社会影响分析（20 分）

生成式人工智能（AIGC）与数字人的融合应用正以前所未有的深度与广度重塑着当代社会的生产范式、社交生态与文化景观。自 2024 年底以来，随着预训练大模型与多模态生成技术的日益成熟，虚拟偶像、合成主播、全能 AI 助手等应用层出不穷。这场技术浪潮在驱动社会进步的同时，也催生了关于伦理、法律与治理的严峻挑战，考验着人类社会的适应能力与价值准绳。

#### 一、 正面影响：效率、普惠与文化传播的范式跃迁

首先，AIGC 带来了内容生产与专业服务的“效率革命”。生成式模型的核心在于“通过学习大规模数据分布，实现对多模态内容的快速合成与创作”（《人工智能概论》，清华大学出版社，2023 年版），这一特性从根本上解放了人类的生产力。例如，AI 写作助手能将寥寥数语的提示词迅速扩展为逻辑严谨、结构完整的文章初稿；AI 视频工具则可在数分钟内生成包含虚拟主持的专业级短片，这在过去需要一个团队数小时甚至数天的工作量。与此同时，7x24 小时全天候在线的合成主播与智能客服，不仅优化了用户体验，更有效缓解了传统服务业在高峰时段的人力瓶颈，实现了资源配置的优化。其次，技术进步正在催生“创作民主化”与产业新动能。正如经济学者李明所预见的：“AIGC 技术让创作门槛大幅降低，为中小企业和个体创业者提供了平等参与文化创意产业的机会”（《经济研究》，2024 年第 6 期）。这种技术普惠的价值尤为突出，使得预算有限的中小企业能借助数字人进行低成本、高效率的品牌营销和客户服务，从而在激烈的市场竞争中获得一席之地。在教育领域，个性化的虚拟教师 and 自适应智能题库，正逐步消弭地域鸿沟，让优质教育资源惠及更多偏远地区的学子。再者，AIGC 显著增强了跨文化传播的深度与效度。依托强大的多语言模型，合成主播不仅能实现多语种的实时、无缝切换，更能保留甚至模拟出贴近人类情感的自然语调。这使得信息与文化的全球化传播摆脱了语言障碍。网络媒体的观察一针见血：“AI 合成翻译不仅加快了信息传播的速度，还通过语气、语调的仿真，提升了与受众的情感共鸣”（TechWeb，2025 年 3 月 5 日）。这不仅是效率的提升，更是情感连接的深化，为构建全球文化互信开辟了全新路径。

#### 二、 负面挑战：信任赤字、权利困境与衍生风险

然而，技术的另一面是潜在的巨大风险。首当其冲的是，深度伪造技术正对社会信任体系构成系统性侵蚀。网络安全专家张华在其《数字信任白皮书》（2024 年 12 月）中发出严正警告：“深度伪造技术门槛持续降低，任何人仅需少量样本数据即可生成高保真度的人脸替换视频，这在政治选举、金融诈骗、名誉毁损等领域构成了迫在眉睫的威胁。”当眼见不再为实，社会共识的基石便会开始动摇。

其次，知识产权的界定在 AIGC 时代陷入了前所未有的“灰色地带”。绝大多数基础模型在训练阶段，都不可避免地“吸收”了海量受版权保护的互联网数据，这种“数据投喂”过程往往未经原创者明确授权，存在“数据原罪”的嫌疑。王淼在《AI 与版权法改革》一文中指出的问题具有普遍性：“AI 模型的训练数据集往往缺乏透明度，导致模型输出内容的来源无法追溯”（《北京大学法律评论》，2025 年 2 月）。当 AI 生成物与原作构成实质性相似时，原创作者的维权之路将因举证困难、成本高昂而异常坎坷，长此以往将严重挫伤创作生态的活力。

此外，AIGC 与数字人还带来了一系列复杂的次生风险。第一，情感异化与心理健康问题。部分用户可能在与虚拟伴侣或 AI 心理咨询师的交互中寻求情感寄托，但这种沉浸式体验若缺乏引导，极易导致个体在现实世界中的社交隔离，甚至引发对真实人际关系的疏离感。第二，劳动力市场的结构性冲击。内容创作、客户服务、语言翻译等领域的传统岗位正面临被数字人“优化”的风险，这要求政府与社会必须前瞻性地布局大规模的职业再培训体系与更具韧性的社会保障网络。第三，技术滥用下的责任归属难题。当 AI 系统决策失误或被恶意利用造成损害时，责任主体（开发者、使用者还是 AI 本身？）的界定在现有法律框架下极为模糊，亟待建立清晰的法律追责与赔偿机制。

综上所述，生成式人工智能与数字人的结合，正以前所未有的力量推动社会进入一个生产力跃升和

文化繁荣的新纪元。然而，这场变革亦如同一枚硬币的两面，深度伪造、版权争议、隐私侵犯等阴影随之而来。人类社会正站在一个关键的十字路口，唯有坚持技术创新与法律伦理的协同进化，构建敏捷而审慎的治理框架，才能确保这场由“AIGC+数字人”引领的深刻变革，最终导向一个更加公平、可信和普惠的未来，真正增进全人类的福祉。

## 二、生成式人工智能的监管与自由的平衡讨论（15分）

AIGC的发展使其成为一把双刃剑，它既是推动社会进步的强大引擎，也可能成为侵蚀社会信任的催化剂。因此，关于“是否应该对AI创作进行内容监管”的讨论，其核心并非“是否监管”，而是“如何监管”，即如何在防范风险与激励创新之间找到最佳平衡点。

### 一、加强监管的必要性与潜在风险

**放任技术自流发展必然导致风险失控，对AIGC加强监管是维护社会秩序，公众利益的必然要求。**

其一，防范系统性社会危害：AIGC最直接的威胁在于其制造虚假信息的能力。深度伪造（Deepfake）技术可被用于炮制假新闻、诽谤个人、操纵舆论乃至影响选举结果，对社会信任和国家安全构成直接冲击。此外，AI生成有害内容（如暴力、色情、极端言论）的规模化和自动化，其传播速度和影响力远超传统手段。因此，设置内容审核、来源追溯（如数字水印）和明确标识（明确标注内容由AI生成）等监管红线，是防止技术滥用的第一道防线。

其二，保护个人核心权益与知识产权：数字人技术的发展使得盗用他人肖像、声音等生物特征变得异常轻易，极易引发人格权滥用和身份欺诈。同时，AIGC的训练数据常包含海量受版权保护的作品，这种未经授权的“学习”引发了严重的知识产权争议。正如法律学者所指出的，“若不加以规制，创作者的权益将难以保障，最终会侵蚀整个内容生态的创新根基”。建立清晰的版权和人格权保护规则，确保训练数据的合法性与透明度，是维持公平创作环境的基石。

**与此同时，不当或过度的监管同样会带来负面效应，可能扼杀技术发展的潜力。**

首先会抑制技术创新与产业竞争力：过于严苛和模糊的监管标准会大幅增加企业的合规成本与法律风险，尤其是对于资源有限的初创公司和中小企业。这可能导致创新活动“降温”，技术迭代放缓。例如，若要求模型对所有生成内容负有绝对责任，可能会迫使开发者过度“保守”，限制模型的创造力和实用性。长此以往，可能导致技术壁垒固化，市场被少数能承担高昂合规成本的科技巨头垄断，损害整个国家的产业竞争力。

其次威胁言论自由与导致“寒蝉效应”：内容监管的尺度极难把握。一个潜在的风险是，监管机构或企业为规避责任，可能会采用“一刀切”的过度审核策略，过滤掉大量无害但具有争议性或讽刺性的内容。这不仅会限制艺术创作和多元文化表达，还可能被滥用为压制批评声音、进行社会议程控制的工具，从而引发“寒蝉效应”，损害言论自由这一现代社会的基本价值观。

### 二、鼓励开放自由的益处与可能的问题

**鼓励开放自由，特别是在开源社区和学术研究中，对AIGC的健康发展至关重要。**

其一可以加速技术迭代与实现“技术普惠”：开放自由的环境，特别是开源模式，允许全球的开发者和研究者共同参与模型的改进、测试和应用开发。这种“集体智慧”模式极大地加快了技术迭代速度，降低了技术使用门槛。正如开源社区所倡导的，“它打破了技术垄断，让中小企业和个人也能享受到AIGC带来的生产力红利”，从而促进了更广泛的社会创新和经济增长。

其二能够增强技术透明度与安全性：与封闭的“黑箱”模型相比，开放和开源的模型允许更多专家对其进行审查，从而更容易发现并修复其中的偏见、漏洞和潜在危害。这种“阳光下的监督”有助于提升AI系统的整体安全性和可靠性，建立公众信任。通过广泛的社区测试，可以更快地开发出对抗恶意使用（如深度伪造检测）的工具，形成一种技术生态内的自我净化和免疫能力。

**但是，绝对的自由同样会带来严峻挑战，其问题与加强监管的必要性互为表里。**

首先导致恶意应用的门槛急剧降低：当强大且不受限制的AIGC工具被完全开放获取时，它们落入恶意行为者手中的风险也随之剧增。犯罪分子可以轻易利用这些工具进行大规模、低成本的电信诈骗、网络攻击、自动化宣传战或制作非法的有害内容。技术的“民主化”在此刻可能异化为“风险的普及化”，对社会安全构成直接威胁。

其次加剧责任界定的困境与监管真空：在开放自由的环境下，一个AI应用可能由多个开源模型

和代码库拼接而成，一旦出现问题，其责任链条将变得异常复杂和模糊。用户、应用开发者、模型贡献者之间责任难以界定，导致受害者维权无门。这种“责任分散”的局面会形成事实上的监管真空，使得恶意行为难以被追究，最终可能因少数人的滥用行为，而摧毁整个社会对该技术的信任。

### 三、从计算机伦理角度分析技术所涉及的关键伦理问题（30 分）

#### 知识点 1:（10 分）

##### （1）教材中的对应章节（1 分）:

第四章第一节 道德的社会价值（第 60-64 页）

第四章第三节 职业道德与个人职业发展（第 65-66, 68-70 页）

第四章第四节 IT 职业人员的社会责任（第 70-76 页）

##### （2）对教材中对应内容的简要概括（2 分）:

**道德的社会价值**中道德并非空泛的说教，而是社会良性运转的基石。它通过构建人与人之间、个人与社会之间的信任与预期，为社会关系的健康发展（良性关系）和工作生活秩序的稳定提供了根本保障。在一个缺乏共同道德准则的社会中，协作成本将无限增高，社会效率与和谐将无从谈起。道德通过继承与彰显传统文化、调节复杂的社会关系、帮助个体认识现实社会以及规范个人行为，发挥着其不可替代的社会功能。

**职业道德**中强调一个“职业”的本质属性，在于其成员不仅拥有专门的知识 and 技能，更重要的是，他们共同遵守一套超越个人利益的职业道德规范，并致力于服务社会。这种职业精神是区分专业人士与普通从业者的关键。职业道德，作为这一精神追求的具体体现，是一系列与特定职业活动紧密相连的行为准则、道德情操与品质的总和。它通常以成文的伦理守则形式出现，这些规范为从业者在面对复杂的道德困境时提供了决策依据，旨在维护整个行业的诚信与声誉。教材明确指出，职业道德规范的核心精神与普世美德——诚实、责任、正直、宽容——是一致的，并强调了终身学习作为信息时代职业道德的特殊要求。

**IT 职业人员的社会责任**中社会责任是职业道德的最高体现，要求技术专家必须超越其直接的工作任务，预见并评估其技术创造对社会、人类及自然环境的广泛影响。教材引用了爱因斯坦的观点，强调关心人类自身及其命运必须是所有技术努力的主要兴趣点，并以“哥伦比亚”号航天飞机灾难（第 71 页）作为警示案例，深刻揭示了技术人员坚持职业责任、敢于提出不同意见的重要性。教材还分析了承担责任的心理障碍，如在大型组织中普遍存在的责任分散效应、为了私利而罔顾责任、以及对权威的不加批判地接受等。这些障碍往往使技术人员难以履行其应尽的社会责任，从而为技术滥用或技术灾难埋下伏笔。

##### （3）人工智能技术涉及相同内容（2 分）:

AIGC 的出现，将 IT 职业道德和社会责任中探讨的职业责任问题推向了一个全新的、更为复杂的维度。一方面，AI 技术的发展需要构建基于道德原则的技术规范与使用准则，正如道德通过规范个人行为来促进社会稳定一样，AI 系统的构建也必须考虑其输出内容的真实性、公正性与社会影响，以实现“良性人机关系”。另一方面，作为 AI 技术开发与应用的主体，IT 职业人员不仅需要掌握高水平的工程能力，更应肩负起“面向全社会负责”的道义责任。下面以三个实例来印证这个观点：

**一、算法偏见**，即 AI 系统可能系统性地放大社会歧视，如在招聘中对特定年龄或种族的求职者产生偏见；一个极具代表性的案例是“Mobley v. Workday, Inc.”诉讼案。在该案中，原告指控 Workday 公司提供的人力资源管理软件，其内置的 AI 筛选算法系统性地、不成比例地排除了 40 岁以上的、黑人以及残疾人求职者。原告称，他们在申请职位后极短时间内（有时甚至是凌晨）便收到拒信，这表明决策过程高度自动化，缺乏人工审核。此案揭示了一个严峻的现实：一个以提升效率为目的的技术工具，可能在无形中实施了大规模的、违背基本伦理与法律的歧视行为。

**二、虚假信息**，AIGC，特别是大型语言模型（LLM），具备生成高度逼真但完全虚假信息（即“幻觉”）的能力，同时也为恶意行为者制造“深度伪造”内容提供了便利工具。比如 AI 生成的虚假新闻以空前的速度和规模传播，其高度仿真性使得公众难辨真伪，从而加剧了社会信任危机、激化了政治极化，并带来了显著的经济与法律风险。例如，利用 AI 生成的虚假声音或文本进行诈骗和勒索的网络犯罪日益猖獗。



三、“唯快不破”创新文化，科技行业长期以来盛行的“快速行动，打破常规”（Move fast and break things）的创新理念，在 AIGC 时代面临着前所未有的伦理审视。这一理念强调速度和颠覆，鼓励在不完全理解后果的情况下快速推出产品。

**（4）关于两者相同之处的简要说明（5 分）：**

**AIGC 系统中的算法偏见**，是职业伦理失职的最突出表现之一。它直接违背了 ACM 伦理准则（ACM Code of Ethics and Professional Conduct）中“避免伤害他人”和“恪守公正”的核心要求。与传统的软件缺陷不同，算法偏见能够系统性地复制、固化乃至放大现实社会中已有的歧视与不平等，其造成的伤害更具隐蔽性、广泛性和持久性。这已非单纯的技术瑕疵，而是一个深刻的伦理失范。开发和部署存在算法歧视系统的技术人员，无论其初衷如何，都未能履行其预见并规避潜在社会危害的职业责任。

**AIGC 生成的虚假信息**直接挑战了职业道德中“做到诚实可信”的基本原则。开发者的责任边界被极大地拓宽了，他们的责任不仅在于代码本身，更在于其造物可预见的滥用方式及其社会后果。面对这一挑战，技术和政策层面的应对措施，可以被视为现代职业责任的延伸实践。技术上，学术界和工业界正积极研发深度伪造检测技术，通过分析视频、音频和图像中人眼难以察觉的不一致性来识别伪造内容。政策上，以《欧盟人工智能法案》（EU Artificial Intelligence Act, 2024 年修订草案）为代表的法规，强制要求对 AI 生成或修改的内容（如深度伪造品）进行明确标注，确保用户知情权。同时，由 Adobe、微软等公司牵头的“内容来源和真实性联盟”（Coalition for Content Provenance and Authenticity, C2PA）正在制定开放的技术标准，通过为数字媒体附加可验证的来源和历史信息（即“内容凭证”），来帮助用户追溯内容源头，辨别真伪。这些举措共同构成了抵御虚假信息泛滥的社会防线，也是 IT 专业人员履行其社会责任的具体体现。

**“唯快不破”理念的伦理审视：**完全理解后果的情况下快速推出产品的理念与 IT 职业人员的社会责任中所倡导的、深思熟虑的“工程意识”背道而驰。这种急于求成的文化，其内在逻辑是将潜在的伦理风险和技术缺陷视为一种可以日后偿还的“伦理债务”。“这一‘伦理债务’的逻辑，也正是教材第 75 页提出的‘负责行为的障碍’在技术时代的延伸表现。”然而，这种模式的危险在于，它将风险和成本不成比例地外部化了。当一个带有偏见的 AI 招聘系统投入使用时，其“伦理债务”的偿还者并非是开发公司，而是那些因此失去工作机会的弱势群体。当一个社交平台因其算法推荐机制而成为虚假信息传播的温床时，其代价由整个社会的公共信任和政治稳定来承担。这种做法从根本上违背了“避免伤害”这一最基本的职业道德准则。

**两者相同之处**在于都强调技术从业者需承担道德和社会责任，其核心伦理准则——如避免伤害、恪守公正、诚实可信——是一脉相承的。教材中的责任模型多基于个体在特定项目中的决策（例如，工程师对具体技术缺陷的警告）。AIGC 时代则将这一责任的范畴和复杂性大大扩展：问题不再是单一的技术缺陷，而是可能源于训练数据、算法设计和技术生态的系统性问题。因此，责任主体从个体工程师扩展到包括投资者、研究机构和政策制定者在内的整个生态系统。总结就是，尽管 AIGC 技术环境带来了责任边界的模糊与伦理风险的新类型，但其所需遵循的职业道德核心并未改变。正如教材中所强调的‘诚实、公正、负责任’三大原则，在 AIGC 的实践中依旧是判断道德行为的基础准则。

**知识点 2:（10 分）**

**（1）教材中的对应章节（1 分）：**

第八章第一节 隐私保护的道德和法律基础（第 145-151 页）

第八章第二节 数据挖掘与公共数据库的隐私安全（第 152-155 页）

第八章第三节 隐私保护的技术策略和伦理规范（第 155-162 页）

**（2）对教材中对应内容的简要概括（2 分）：**

**隐私的道德与法律基础**中，教材将隐私定义为一项关乎个人尊严和自主权的基本权利，其核心在于个体能够控制关于自身的信息——何时、如何以及在何种程度上与他人分享。这种控制权是个人建立自我认同、维持人际关系和独立决策的前提。教材进一步指出，隐私与公民自由紧密相连，在一个无所不在的监视环境中，言论、思想和结社自由会因“寒蝉效应”而受压制，人们会因担心言行

被记录、分析和评判而进行自我审查，从而削弱民主社会的活力。因此，保护隐私就是保护公民能够不受胁迫地参与公共生活的基本空间。

**数据挖掘与数据库的威胁**中，教材阐述了在 AIGC 出现之前，技术对隐私的主要威胁源于大规模数据库的建立和数据挖掘技术的应用。数据挖掘能够从海量数据中发现隐藏模式，推断出个人敏感信息，如健康状况、消费习惯或政治倾向。即便数据经过匿名化处理，在聚合与分析后也可能重新识别出个人身份，对隐私构成严重威胁。

**隐私保护**中，教材强调了“知情同意”这一核心伦理原则。这意味着，在收集和使用个人信息之前，组织必须以清晰、明确的方式告知数据主体其信息的具体用途、潜在风险以及数据主体的权利，并获得其自愿、明确的授权。这一原则是技术保护措施（如加密）和法律法规（如隐私法）之外，保障个人自主权的最后一道、也是最重要的一道防线。

### （3）人工智能技术涉及相同内容（2分）：

AIGC 的出现，使传统隐私威胁发生了根本性的转变。与以往技术主要通过“分析既有数据”来挖掘隐私不同，AIGC 不仅在训练阶段可能无意泄露敏感信息，还在生成阶段具备“主动推断”和“行为塑造”的能力。这种隐私侵害的模式更为隐蔽、系统且难以防范，带来了全新的伦理挑战。

首先，训练数据的“记忆化”与“推断能力”构成了 AIGC 时代最显著的隐私风险。大型语言模型在吸收庞大数据集的同时，往往无意中“记住”了其中的具体信息片段，包括姓名、住址、身份证号等敏感信息。这种“记忆”可能在用户交互过程中被触发、泄露。例如，研究人员发现，在特定提示词诱导下，模型可能重现训练语料中的原始文本内容，导致数据外泄。更值得警惕的是，AIGC 还能基于已学到的模式，对用户进行高度敏感的特征推断，如通过文本语气和历史行为预测其心理健康状况、宗教信仰或社会经济地位。这类“推断数据”虽未直接存储，但其对隐私构成的风险丝毫不亚于明文信息泄露。正如欧盟《通用数据保护条例》（GDPR）与加州《消费者隐私法案》（CCPA）均指出：“即便是通过算法生成的、间接获取的个人信息，也应受到同等保护”，但在现实操作中如何界定“可识别性”和“推断合理性”仍然是亟待解决的法律空白。

其次，AIGC 本质上是“监视资本主义”机制的加速器。在这种模式中，用户的行为痕迹不再只是副产品，而被视作构建商业模型的“原材料”。技术平台通过采集、分析、模型化用户数据，训练 AIGC 系统不断优化其对用户的理解和操控能力。这不仅实现了内容的极度个性化与商业定向，还实质性地削弱了用户对自身信息使用的控制权。正如社会学家舒欣娜在《算法与社会秩序》（2024 年）中所述：“监视资本主义的真正问题不在于我们被看见，而在于我们在不知情中被塑造。”

第三，AIGC 还具备前所未有的行为预测与操控能力。依托语言生成模型，系统可以自动创作“定制化钓鱼邮件”“诱导性广告文案”“政治宣传语句”，精准击中个体心理弱点。这类高度拟人化的“说服内容”，正逐步瓦解个体的批判性思维，使决策更依赖算法推荐而非理性判断。其深层逻辑是通过信息环境重构人的行为路径，从“被预测”到“被操纵”。这构成了对个人自由意志的直接挑战，也引发了对“认知安全”的伦理关切。

综上，AIGC 所引发的隐私问题，不仅在信息泄露程度上更加严重，更在本质上改变了个体与信息之间的关系。传统的隐私保护理念已难以应对这种系统性、生成性与操控性的风险，亟需在技术、法律与伦理三方面进行系统性重构。

### （4）关于两者相同之处的简要说明（5分）：

无论是教材中对数据挖掘技术的伦理警示，还是当今 AIGC 技术的隐私挑战，其背后的核心伦理关切始终是一致的：即如何在技术快速发展的同时，捍卫个体对于个人信息控制权、自主权与人格尊严。教材指出，“隐私权作为人们的一项基本权利，理应作为人的独立人格权的组成部分，成为法律保护的对象”（第 155 页），这一点在 AIGC 时代不仅未过时，反而显得更加迫切。

在技术路径上，教材所描述的“数据挖掘”可以看作是 AIGC 训练逻辑的早期雏形，它们本质上都基于对大规模数据的模式学习。但 AIGC 的能力已从“分析已知”演化为“生成未知”，这不仅扩大了信息处理的深度和广度，也放大了对个体隐私的威胁。教材中强调“知情同意”是保障隐私的最后一道防线，而在 AIGC 中，这道防线变得岌岌可危。用户往往在不知情的情况下，其个人数据被用于训练模型；更复杂的是，AIGC 可以基于公开信息生成关于个体的“推断数据”或“虚拟分身”，使“知情同意”难以涵盖所有衍生风险。

二者在伦理后果上也呈现出高度一致性。教材指出监控社会会带来“寒蝉效应”，人们因担心被监视而不敢表达真实思想。而 AIGC 将这种影响推向了“操控性信息生成”的新阶段：技术不再只是观察现实，而是在塑造现实。用户不仅面临“我被看见”的隐私困境，更面临“我所看到的是否真实”的认知危机。这极大挑战了个体的自主判断和社会的认知共识。

面对这一挑战，AIGC 时代所涌现出的隐私增强技术（如联邦学习、差分隐私）与用户权利保障机制（如 GDPR 赋予的解释权与数据可携带权），正是教材中伦理原则在现实语境下的制度化延展。可以说，从数据挖掘到 AIGC，尽管技术面貌不断演进，但其伦理本质和应对逻辑始终如一：即在权力与知情之间保持平衡，在创新与责任之间划定边界。这正是教材“尊重个人、保护自由、构建信任”三大隐私伦理价值，在当代人工智能语境下的延续与深化。

### 知识点 3: (10 分)

#### (1) 教材中的对应章节 (1 分):

第七章第一节 知识产权基础 (第 128-131 页)

第七章第二节 版权、专利、商标和商业秘密 (第 131-137 页)

第七章第三节 软件盗版问题与开放源代码运动 (第 137-141 页)

第七章第四节 网络知识产权 (第 141-144 页)

#### (2) 对教材中对应内容的简要概括 (2 分):

**知识产权及其形态与宗旨：**教材首先介绍了知识产权的几种主要形式：版权、专利、商标和商业秘密。其根本宗旨在于通过法律手段，赋予智力创造者对其劳动成果在一定期限内的专有权利，从而激励创新和文化繁荣。在 IT 领域，版权尤为重要，它保护软件代码、数据库、数字内容等“表达形式”，而非抽象的“思想”本身。

**数字化带来的核心挑战——盗版：**教材明确指出，数字技术的核心特征——能够以零成本进行完美、无限的复制——对传统的版权模式构成了根本性挑战。软件、音乐、影视作品的盗版问题，正是这一挑战的直接体现。这种行为被定性为对他人智力劳动成果的“盗窃”，是文明社会所不容许的，各国也为此制定了相应的法律进行保护。

**另类范式:开源运动，**与强调专有控制的传统 IP 模式相对，教材介绍了“开放源代码运动”。这一运动倡导“开放、自由、共享、合作”的伦理精神，认为知识和代码的自由流动能够催生更强大的集体智慧和更快速的技术迭代。开源模式并非否定智力劳动，而是通过不同的许可协议（如 GPL、Apache）来重新定义权利与义务，构建了一个与专有模式并存且充满活力的生态系统。

**网络知识产权的保护困境：**教材第四节专门探讨了“网络知识产权”。教材指出，网络知识产权因其依附于网络技术，具有全球性、无国界性、易复制传播和匿名性等特点。这些特点导致了一系列问题，包括更为便捷的软件盗版、网络言论自由与侵权责任的冲突、法律法规的滞后以及跨国界管辖的难题。为应对这些挑战，教材提出应采取综合性的保护措施，包括完善立法、发展技术保护模式（如加密）、加强伦理规范建设、推进行业自律与国际合作，并提高公民的信息意识与法律素养。

#### (3) 人工智能技术涉及相同内容 (2 分):

AIGC 的出现，将教材中描述的“盗版”问题，从个体或小团体的非法复制行为，升级为工业化、自动化、规模空前的数据“摄取”行为，从而引发了版权制度诞生以来最深刻的一场危机。

**一、训练数据的版权争议：**当前 AIGC 领域最核心的法律冲突，在于未经授权使用受版权保护的作品来训练 AI 模型，是否构成美国版权法下的“合理使用”或其他国家法律中类似的豁免条款。AI 开发者普遍认为，这种使用是“变革性的”，因为其目的不是复制作品本身以替代市场，而是从中提取统计模式和抽象知识来构建一个全新的工具，类似于人类的学习过程<sup>71</sup>。然而，权利人则认为，这种大规模、系统性的复制行为，特别是当生成结果与原作风格相似或直接竞争时，严重损害了其作品的潜在市场和价值，构成了赤裸裸的版权侵犯。这一争议已在多起备受瞩目的诉讼中展开，以著名的 OpenAI 和纽约时报的诉讼案为例，纽约时报起诉“OpenAI 的 LLM 不仅复制了数百万篇新闻文章用于训练，还能在输出中近乎逐字地复现受版权保护的内容，直接替代了用户访问原网站的需求”，构成市场损害，而 OpenAI 回应称“训练是为了提取事实和语言模式，而非替代新闻产品；输出的复现内容是罕见的“漏洞”而非设计意图，且可通过技术手段修复”，法院的判决无疑决定了新闻和出



版业能否保护其内容不被用于训练 AI，并维持其订阅和流量的商业模式。对“变革性使用”和“市场损害”的司法解释将产生深远影响。

**二、AI 生成物的作者身份危机：**AIGC 提出了一个根本性问题：机器能否成为“作者”？根据美国现行法律，版权只授予人类作者，这意味着，完全由 AI 自主生成的作品，由于缺乏人类作者，可能不享有版权保护，从而直接进入公共领域。美国版权局对此发布了指导意见，确立了“有意义的人类作者身份”标准。根据该标准，如果人类用户仅提供一个简单的提示词，而 AI 系统决定了作品的“表达元素”（如诗歌的韵律、画面的构图），那么该作品就不是人类创作的。然而，如果人类对 AI 的输出进行了大量的、具有创造性的修改、选择或编排，那么人类作者可以对其贡献的部分主张版权。这一标准虽然提供了一个分析框架，但也制造了巨大的法律不确定性，使得创作者陷入了一个“双重困境”：他们既希望利用 AI 作为强大的创作工具，又担心过度依赖 AI 会导致其作品失去版权保护。

### （3）关于两者相同之处的简要说明（5 分）：

**教材与现实中的 AIGC 知识产权问题，在核心理念上高度一致：即在保障创作者权益与推动社会创新之间寻找平衡。二者的相同之处主要体现在以下几个方面：**

首先，“未经授权即为盗用”的伦理判断是一以贯之的基线原则。教材将盗版界定为对他人智力成果的“偷窃”，强调其破坏公平秩序的本质。AIGC 训练数据的争议，虽披上了“技术中立”或“学习模拟”的外衣，但其实质与教材所批判的盗版行为本质相似，尤其当 AI 输出能够直接替代原始内容时，更加凸显其侵权的现实后果。

其次，AIGC 放大了教材中指出的网络知识产权保护困境。如教材所述，网络环境下的跨国管辖、技术滥用和版权执法难度，在 AIGC 时代被无限放大：算法来源匿名、内容生成难以追踪、输出风格模糊化，使传统版权机制面临“力不从心”的挑战。权利人维权成本高昂，平台又常以“技术难以控制”为由规避责任，进一步削弱了制度的公平性。

再次，教材倡导的“共享与合作”的开源精神，正在 AIGC 争议中得到新的诠释与延展。面对法律僵局，社会各界正探索共赢型替代方案，例如建立“AI 训练数据许可市场”，推动数据集“选择加入/退出”机制，甚至以“数据尊严”理论主张将内容创作者视为“数字劳动者”，应获得合理报酬。这种伦理演进正是教材中“开放源码运动”所蕴含价值观的现实体现——技术不应以牺牲个体为代价换取进步，而应实现更公平、可持续的利益分配。

总之，无论是教材中的软件盗版问题，还是现实中的 AIGC 数据争议，二者都以“如何正当使用他人智力成果”为核心，所遵循的道德逻辑与价值准则是一脉相承的。而 AIGC 所带来的问题，更像是对教材所述知识产权伦理挑战的一次全面升级，需要我们在原有理论上不断拓展法律与伦理的新边界。

## 四、个人观点与案例分析（25 分）

### 4.1 请结合前文分析，谈谈你认为面对 AIGC 及数字人格技术，未来应采取何种伦理治理策略，才能在促进创新的同时保障公众权益。（15 分）

面对生成式人工智能（AIGC）及数字人格技术的浪潮，我们正处在一个决定未来技术形态与社会结构的十字路口。为了在促进创新的同时保障公众权益，我们必须摒弃被动应对和零敲碎打的治理模式，转而构建一个多层次、前瞻性且以人为本的综合伦理治理框架。这一框架的核心，是将《计算机伦理学》中所倡导的经典伦理原则——如避免伤害、确保公平、尊重自主——从个体道德层面，提升为整个技术生态系统的结构性责任。这要求我们从创新的文化根源、个人权利的法律保障以及价值创造的经济模式三个维度，进行系统性的重塑与构建。

**首先，我们必须彻底反思并重塑创新的责任范式，**核心在于终结“唯快不破”的旧有信条，建立覆盖技术全生命周期的“生态系统问责制”。科技行业长期信奉的“快速行动，打破常规”理念，在 AIGC 时代已显现出其巨大的伦理风险。这种模式将技术缺陷和伦理风险的成本不成比例地外部化，最终由社会和最脆弱的群体承担。当一个带有偏见的 AI 招聘系统不成比例地淘汰特定人群时，当一个为追求快速迭代而疏于监管的平台成为虚假信息和深度伪造内容的温床时，其所谓的“伦理债务”便由公众信任的瓦解和个体尊严的损害来偿还。公众的忧虑是显而易见的，一项民意调查显示，



高达 72% 的民众倾向于放缓而非加速 AI 的发展。因此，未来的治理策略必须将责任前置。这不仅要求开发者在设计之初就嵌入伦理考量，秉持教材中所倡导的、深思熟虑的“工程意识”，更要求将问责范围扩大到整个生态系统。风险投资机构作为技术发展方向和速度的关键塑造者，不能再仅仅扮演资金提供者的角色，而应在投资决策中引入“伦理尽职调查”，评估初创公司技术可能带来的社会危害，并将负责任的 AI 原则作为核心标准之一。同时，高等教育机构肩负着特殊使命，它们不仅要引领负责 AI 的研究方向，更要承担起培养具备伦理素养的技术人才、向全社会普及 AI 知识的社会责任，确保下一代创新者不仅掌握技术，更懂得如何合乎道德地使用技术。

**其次，我们亟需以前所未有的力度加固个人隐私与数字自主权的防线。**AIGC 的威胁已从传统的数据挖掘演变为对个人行为的预测、操纵乃至“数字人格”的生成与复制，这直接侵蚀了作为隐私核心的个人自主权。哈佛大学教授肖珊娜·朱伯夫所提出的“监视资本主义”理论深刻地揭示了这一模式的本质：将用户的行为痕迹转化为可分析的“行为剩余”，其最终目的不仅是预测，更是为了塑造和引导行为，以实现商业利益的最大化。因此，治理策略必须超越目前在复杂 AI 系统面前显得捉襟见肘的“知情同意”模式。我们必须大力推广和部署隐私增强技术（PETs），如联邦学习、差分隐私和机密计算，从技术架构层面实现“设计即隐私”的理念，在不牺牲数据效用的前提下最大程度地保护个人信息。更重要的是，我们应在法律上明确“数字人格权”这一新型权利，确认个体对其声音、肖像及行为模式的 AI 生成复制品拥有绝对的控制权和所有权，以应对“深度伪造”等技术带来的身份盗用和人格侵犯风险。与此同时，必须强制推行算法的透明度与可解释性（XAI），将欧盟《通用数据保护条例》（GDPR）所倡导的“解释权”落到实处。尽管解释一个拥有数千亿参数的“黑箱”模型在技术上极具挑战性，但这对于保障个体的知情权、允许用户挑战不公的自动化决策、并最终建立对 AI 系统的信任至关重要。

**最后，我们必须重构知识产权与价值分配的经济模型，**以应对 AIGC 带来的创作生态危机。当前围绕 AI 训练数据的“合理使用”争议，本质上是传统版权法在面对工业化数据抓取时的失灵。未来的治理策略不应止步于旷日持久的诉讼，而应积极构建新的价值分配机制。这首先意味着要推动建立以“选择加入”为默认原则的数据授权市场，而非将举证责任不公平地推给创作者的“选择退出”模式。我们可以借鉴音乐产业的成功经验，发展高效的集体管理组织（CMOs），以处理大规模、碎片化的数据授权和版税分配，确保权利人能从其作品的 AI 应用中获得合理报酬。而更具前瞻性的方案，是采纳由计算机科学家杰伦·拉尼尔等人倡导的“数据尊严”伦理框架。该理念主张，所有为 AI 训练提供数据贡献的个人和创作者，都应被视为价值链中不可或缺的“数据劳动者”，并有权因其贡献而获得公平的经济回报。正如拉尼尔所言：“我希望未来人们能够因为向 AI 提供特别有用的数据而获得报酬，从而激励更好、更多样化的数据生产”。这种模式不仅为创作者的生计提供了保障，更能通过激励高质量、原创性数据的产生，从根本上改善 AI 生态系统的健康度与多样性。

综上所述，一个融合了前瞻性责任文化、强化型隐私权保护以及公平化价值分配的综合治理策略，是引导 AIGC 与数字人格技术走向一条既充满创新活力又维护人类尊严与福祉的可持续发展道路的唯一选择。这绝非单一行动者的任务，而需要所有利益相关方——开发者、企业、政策制定者、投资者、教育者和广大用户——的共同努力与协作，以确保这项强大的技术进步真正服务于人类的共同利益，而非侵蚀它。

#### 4.2 请举例说明和支撑你的观点和看法。（10 分）

##### 案例 1: Workday AI 招聘歧视案

###### （1）案例出处（如网址）：

CDF Labor Law LLP: “Federal Court Grants Preliminary Certification in Landmark AI Hiring Bias Case”:

<https://www.callaborlaw.com/entry/federal-court-grants-preliminary-certification-in-landmark-ai-hiring-bias-case>

###### （2）案例简述（100-200 字）：

在 Mobley v. Workday, Inc. 案中，多名求职者提起集体诉讼，指控人力资源软件公司 Workday 提供的 AI 招聘筛选工具存在系统性歧视。原告称，该 AI 算法不成比例地排除了 40 岁以上、黑人及残

疾人群体的求职者。他们提交的证据包括，在申请职位后极短时间内，有时甚至是凌晨，便收到自动发出的拒信，这表明决策过程缺乏有效的人工审核，完全由算法主导。

**(3) 案例简析 (100-200 字):**

此案是 AIGC 时代职业与社会责任问题的典型体现，深刻揭示了“算法偏见”的现实危害。它表明，技术开发者和提供商的责任已远超代码功能的实现，必须延伸至其产品的社会公平性影响。该案挑战了将 AI 供应商视为中立技术平台的传统观念，推动了将责任主体扩展至第三方技术供应商的法律认知。这警示我们，若无前瞻性的伦理审查和严格的偏见审计，旨在提升效率的工具可能反而会成为固化和放大社会不公的帮凶。

**案例 2: “AI 孙燕姿”与数字人格侵权**

**(1) 案例出处 (如网址):**

2023 年 4 月 14 日，匿名网友“幽灵创作者”(Ghostwriter) 利用歌手“Drake”和“the Weeknd”的声音训练 AI 模型并生成了歌曲《Heart on My Sleeve》:

<https://nysba.org/copyright-law-in-the-age-of-ai-navigating-authorship-infringement-and-creative-rights/>

AI 孙燕姿爆火歌曲视频: <https://www.bilibili.com/video/BV1yv4y1H7wz>

**(2) 案例简述 (100-200 字):**

近期，利用 AI 技术模仿新加坡歌手孙燕姿声音的“AI 孙燕姿”在网络上爆红，其生成的翻唱歌曲在音色、唱腔上与真人高度相似，引发了大量关注和传播。这些 AI 生成内容在未获得孙燕姿本人或其版权方授权的情况下，被广泛用于商业或非商业目的，引发了关于声音权、肖像权以及新型知识产权的激烈讨论。

**(3) 案例简析 (100-200 字):**

“AI 孙燕姿”现象集中体现了“数字人格”技术带来的伦理与法律挑战。它暴露了现有法律体系在保护个人声音、表演风格等独特人格标识方面的不足。这不仅是版权侵权问题，更是一种新型的人格权侵犯。该案例凸显了建立“数字人格权”的紧迫性，即个体应有权控制其 AI 生成的声音、形象等数字复制品的使用。这要求治理策略必须超越传统 IP 框架，为保护个人在数字时代的身份完整性与自主性提供坚实的法律基础。

**案例 3: 《纽约时报》诉 OpenAI 与微软案**

**(1) 案例出处 (如网址):**

人工智能时代: OpenAI 对 NYT 的版权侵权索赔提出合理使用辩护:

<https://entertainmentlawreview.lls.edu/openai-argues-fair-use-defense/>

OpenAI 和《纽约时报》在庭审辩论中就 AI 科技公司的版权侵权展开辩论:

<https://digiday.com/media/openai-the-new-york-times-debate-copyright-infringement-of-ai-tech-companies-in-first-trial-arguments/>

**(2) 案例简述 (100-200 字):**

2023 年 12 月，《纽约时报》对 OpenAI 和微软提起诉讼，指控其未经许可，使用数百万篇受版权保护的新闻文章来训练其大型语言模型（如 ChatGPT）。诉状指出，这些模型不仅构成了大规模版权侵犯，其生成的输出内容还能近乎逐字地复现《纽约时报》的原文，直接与原告的订阅服务构成竞争，损害了其核心商业利益。

**(3) 案例简析 (100-200 字):**

此案是全球瞩目的 AI 训练数据“合理使用”边界之战，直接关系到整个内容创作行业的未来。它揭示了传统版权法在应对工业级、系统性数据抓取时的无力，并迫使我们重新审视“合理使用”原则的适用性。该案的结果将深刻影响 AI 公司获取训练数据的成本与合法性，并可能催生新的经济模式，推动从当前的法律对抗转向建立公平的授权许可市场和价值分配机制，这与“数据尊严”的理念不谋而合。

#### 案例 4：内容来源和真实性联盟（C2PA）

##### （1）案例出处（如网址）：

C2PA 官方网站及技术说明文档：<https://c2pa.org/>

##### （2）案例简述（100-200 字）：

内容来源和真实性联盟（C2PA）是由 Adobe、微软、英特尔、Truepic 等行业巨头联合发起的项目。它旨在通过制定一个开放的、免费的技术标准来解决数字内容的信任危机。该标准通过为图像、视频等媒体文件附加一个防篡改的“内容凭证”，记录其来源、作者以及编辑历史，从而帮助用户验证内容的真实性与出处。

##### （3）案例简析（100-200 字）：

C2PA 是应对 AIGC 虚假信息挑战的典范性治理策略，体现了行业自律和多方协作的前瞻性思维。它并非依赖于强制性法规，而是通过构建一个开放的技术生态系统，为创作者提供了保护其作品完整性的工具，也为消费者提供了辨别真伪的手段。这种“技术向善”的解决方案，与单纯的法律惩戒形成互补，是实现伦理治理的重要一环，有助于在不扼杀创新的前提下，提升整个信息生态系统的透明度、问责制和可信度。