

(以论文、报告等形式考核专用)

- **数字化表演艺术：**2023 年北京冬奥会开幕式中应用数字人技术，虚拟演员与真人互动，展现中国传统文化与现代科技结合的魅力。
- **多语种文化传播：**AIGC 支持的自动翻译与多语言内容生成，使不同文化之间的信息交流更便捷，如百度的多语种 AIGC 系统支持“一带一路”沿线国家文化交流。
- **文化遗产保护：**数字人技术被应用于虚拟文物修复与展览，通过虚拟现实让公众体验历史文化，增强文化认同感。

### 3. 促进教育公平与个性化学习

AIGC 和数字人技术在教育领域的广泛应用有效促进教育资源均衡，推动个性化学习。

- **虚拟教师与助教：**多所高校和在线教育平台推出 AI 虚拟教师，支持 24 小时教学答疑，辅助师资力量紧张的偏远地区。
- **智能学习路径规划：**通过分析学生学习数据，AI 系统能够为每个学生定制适合其水平和兴趣的学习计划，提高学习效率。
- **语言学习支持：**数字人辅助口语练习，提供实时发音反馈，帮助学习者克服语言障碍。
- **学习内容生成：**AIGC 可自动生成测验题目、习题解析和教学视频，极大丰富教学资源库。

### 4. 优化公共服务与提升社会治理能力

数字人在政务服务、金融咨询、法律援助等公共服务领域实现智能化应用，提升效率和公平性。

- **政务数字人：**多个地方政府部署智能数字人政务服务助手，实现在线政策解读、办事指引，减少群众跑腿，提高政务透明度。
- **金融智能客服：**银行和保险公司引入数字人客服，提供个性化金融产品推荐，快速响应客户需求，降低服务成本。
- **法律咨询助手：**AI 法律顾问帮助普通民众进行法律知识普及和简单案件咨询，降低法律服务门槛。

### 5. 推动医疗健康普惠化

- **基层医疗支持：**AI 数字医生通过远程问诊与诊断，为偏远地区提供医疗支持，缓解医疗资源不平衡。
- **智能健康管理：**结合可穿戴设备，数字人帮助用户监测健康状况，提供个性化健康建议和预警。
- **心理健康辅导：**数字心理顾问通过对话引导，辅助情绪调节和心理疏导，减轻心理疾病治疗压力。

## （二）负面社会影响

### 1. 深度伪造与虚假信息泛滥

AIGC 技术产生的深度伪造（Deepfake）视频、音频和图像，极易被恶意利用制造虚假信息，影响社会稳定。

- 2023 年美国总统选举期间出现的深度伪造事件印证了：“Deepfake 技术已形成完整的黑产业链条，单条伪造视频的社交传播量可达亿级”[4]。
- 中国网信办的监测数据显示：“2024 年 Q1 数字人诈骗案件同比激增 230%，涉案金额超 12 亿元”[5]。

- 虚假新闻自动生成工具被不法分子利用，制造舆论风暴，危害社会信任。

## 2. 知识产权保护困境

AIGC 训练过程依赖海量数据，许多内容未经授权使用，侵犯原创者权益。

- 2024 年中国知名画家集体诉讼某 AI 绘画平台，指控其未经授权大量使用其作品训练模型，破坏艺术创作生态。
- 软件领域 AI 辅助编程可能引入未授权代码片段，引发版权纠纷。
- 影视行业对 AI 生成的虚拟演员肖像权和表演权界定尚不明确。

## 3. 个人隐私与人格权威胁

数字人模拟真实个体形象，容易引发人格权被侵害的风险。

- 2024 年中国娱乐圈爆发数字人侵权事件，一位著名女演员被虚拟数字人仿冒形象参与商业活动，导致形象受损。
- 用户数据在训练和使用过程中存在被滥用、泄露风险，个人隐私保护成为难题。

## 4. 劳动力市场结构冲击

AIGC 在内容创作、客服、教育等行业大量替代人工岗位，引发就业压力和职业转型挑战。

- 2024 年部分科技公司宣布大规模裁撤传统内容编辑岗位，员工转岗和再培训问题凸显。
- 新职业与传统职业脱节，导致技能错配和就业结构性矛盾。

## 5. 算法偏见与社会不公

AI 模型训练数据不完善或偏颇，导致算法歧视现象存在。

- 招聘系统中对女性、少数民族等群体存在隐形偏见，影响公平就业。
- 信贷评估模型歧视低收入群体，加剧社会不平等。

## 6. 文化认同与社会心理问题

数字人和虚拟偶像带来虚拟社交热潮，部分用户沉迷数字世界，现实社交能力下降。

- 青少年对虚拟偶像过度依赖，影响现实人际关系，甚至导致心理健康问题。
- 文化同质化风险，地方传统文化受到边缘化。

在深入分析生成式人工智能与数字人技术所带来的多方面社会影响之后，有必要进一步探讨其在治理层面所面临的重大挑战。特别是在科技快速发展的背景下，如何在保障技术创新活力的同时，确保社会公共利益与个人权利不受侵害，成为当前亟需平衡的重要议题。

## 二、对生成式人工智能的监管与自由的平衡讨论

随着生成式人工智能（AIGC）技术的迅猛发展，其带来的社会影响日益显著，内容的真实性、版权归属、伦理风险等问题频发。围绕“是否应加强对 AI 生成内容的监管”，社会各界展开了广泛讨论。本章将从“加强监管的必要性与潜在风险”及“鼓励开放自由的益处与可能问题”两方面进行比较分析。

### （一）加强监管的必要性与潜在风险

#### 1. 加强监管的必要性

##### （1）防止技术滥用与风险扩散

生成式 AI 的强大生成能力如果缺乏有效监管，极易被用于制造虚假信息、深度伪造、诈骗欺诈等恶意用途，严重危害社会安全和公共秩序。例如，2023 年多起利用深度伪造技术制造政治人物假视频事件，加剧了社会分裂与不信任。中国政府高度重视网络空间治理，出台《网络信息内容生态治理规定》，要求平台对 AI 生成内容进行严格审核，防止有害内容传播。

##### （2）保护知识产权和原创权益

AIGC 模型训练往往依赖海量未经授权的作品，侵犯原创作者的合法权益。加强监管可以明确版权界限，建立数据使用许可和侵权责任追究机制，保护文化创意产业健康发展。例如，中国 2024 年知识产权局开展“AI 版权保护专项行动”，对侵权平台依法查处。

##### （3）保障个人隐私与人格权

数字人仿冒真人形象、声音，可能侵犯当事人肖像权、隐私权。严格监管可建立数字人实名登记与使用规范，防范数字人格权被滥用。2024 年中国《个人信息保护法》加强对个人生物特征数据的保护，明确 AI 模型训练的数据边界。

##### （4）维护社会公共利益和伦理底线

监管能够确保 AI 技术的发展符合社会价值观，防止算法歧视和偏见固化，保障社会公平正义。中国社会科学院发布《人工智能伦理报告》，提出建立算法透明、可解释机制，是加强监管的重要方向。

#### 2. 加强监管的潜在风险

##### （1）可能抑制技术创新活力

过度严苛的监管措施可能导致企业创新动力减弱，限制新技术、新模式的探索与应用，形成“技术寒冬”。例如，欧盟拟定的《人工智能法案》草案被部分业内人士批评为“过于严苛”，可能阻碍欧洲 AI 产业发展。[6]

##### （2）加剧技术和地区发展不均

监管壁垒可能使技术资源向监管宽松的区域集中，造成全球或国内技术发展不均衡。部分中小企业和初创企业可能因合规成本高昂而被挤出市场，影响产业生态多样性。

##### （3）监管与技术演进速度脱节

AIGC 技术更新快速，监管体系难以做到实时响应，存在“监管滞后”现象，可能导致监管空白或“灰色地带”，影响监管效果。

## （二）鼓励开放自由的益处与可能问题

### 1. 鼓励开放自由的益处

#### （1）促进技术突破与跨界创新

开放自由的环境激发更多创新思想和技术尝试，推动生成式 AI 技术快速迭代与应用。开源社区如 OpenAI、HuggingFace 汇聚全球研发力量，加速模型优化与创新。HuggingFace 平台汇集超 50 万开发者，其协作创新速度是封闭体系的 3.2 倍”[7]

#### （2）推动国际合作与伦理共治

开放自由有助于建立全球 AI 治理框架，促进跨国伦理标准制定和经验共享。中国积极参与联合国人工智能伦理工作组，推动“共商共建共享”治理理念。

#### （3）降低研发成本，促进普惠

开放平台和开源模型降低了开发门槛，有利于发展中国家和中小企业快速获取技术资源，缩小数字鸿沟。

#### （4）支持多元文化和应用场景

开放环境激发不同文化背景的创新者探索本地化应用，丰富生成式 AI 的文化表达和社会价值。

### 2. 鼓励开放自由的可能问题

#### （1）技术滥用风险加剧

开放使得恶意用户更易获得先进技术，制造虚假内容和网络攻击工具，增加监管难度。例如，2024 年网络安全报告显示，多起利用开源 AI 工具生成钓鱼邮件的案例激增。

#### （2）知识产权保护难度提升

开放共享导致版权追溯困难，原创内容被未经许可广泛使用，版权保护体系面临挑战。

#### （3）伦理风险和社会责任模糊

开放平台上的内容缺乏统一的伦理审查标准，可能产生歧视、偏见、暴力等不良信息，影响社会价值观。

监管与自由的平衡问题不仅是制度性和法律性的，更深层次地涉及伦理价值的判断与取舍。下面，本文将进一步从伦理视角出发，系统梳理生成式人工智能与数字人在实际应用中可能引发的伦理困境与道德风险，进而为后续的治理实践提供有力支撑

### 三、生成式人工智能和数字人的伦理问题分析

生成式人工智能（AIGC）和数字人技术以其强大的内容创造力和人机交互能力，正在深刻影响社会各领域。然而，伴随技术普及而来的多重伦理问题，不容忽视。以下将围绕 AIGC 分别从 IT 职业道德和社会责任、计算机技术与隐私、信息技术与知识产权问题三个方面展开讨论。

#### （一）IT 职业道德与社会责任

AIGC（人工智能生成内容）技术在推动科技发展的同时，也带来了复杂的伦理挑战。IT 从业者在开发与应用 AIGC 技术时，应严守职业道德规范，主动承担多层次社会责任，确保技术进步造福社会与人类，避免带来新的伤害和风险。

##### 1. 算法偏见与歧视

AIGC 系统训练依赖历史数据，若数据中存在性别、种族、文化等偏见，可能导致模型在生成内容时产生歧视性输出，破坏社会公平和包容性。

##### 实例：

2018 年，亚马逊招聘系统 AI 因训练数据偏向男性候选人，导致系统自动降低女性求职者评分，引发广泛批评并最终被废弃；

2023 年，面向艺术创作的图像生成工具 Stable Diffusion，被发现默认输出欧美白人形象，亚洲、非洲形象明显不足，暴露训练样本的不平衡问题。

##### 职业道德要求：

恪守公正与无歧视原则：在模型训练与应用过程中，主动剔除带有偏见的数据样本；

建立算法公平性检测机制：如 Google 开发的“Fairness Indicators”，持续监测算法输出是否存在不公平；

承担社会后果：一旦偏见输出影响社会，应积极介入修正，回应公众舆论，保障受影响群体权益。

##### 2. 错误传播与虚假信息

AIGC 技术能制造极具迷惑性的虚假信息，如伪造新闻、深度伪造（deepfake）视频，极易在社交平台快速传播，误导公众舆论甚至威胁社会安全。

##### 实例：

2024 年美国总统初选期间，多起候选人语音深度伪造事件在社交媒体流传，试图干扰选民认知，激发公众对 AI 滥用的恐慌；

2023 年，日本某社交平台上虚假生成的地震警报新闻一度引发群众恐慌，导致实际公共资源调度紊乱。

##### 职业道德要求：

秉持诚实可信原则：开发过程中强化生成内容真实性验证，如 Meta 引入 AI 内容水印机制（“Content Credentials”）标识生成来源；

加强用户提示：在产品界面明确标识生成内容来源及可能存在的虚构成分，提高公众辨别能力；

建立风险预警与滥用防控系统：如微软在 Copilot 中加入“敏感内容预警”机制，防止敏感生成内容的直接发布。

### 3. 失业与技能替代

AIGC 的广泛应用正在取代部分传统职业，尤其在新闻写作、平面设计、文案撰写等领域，引发职业结构调整和技能需求转型问题。

#### 实例：

2023 年 BuzzFeed 裁撤新闻编辑岗位，部分内容改由 ChatGPT 辅助生成，成为 AIGC 取代新闻写作职业的早期案例；

全球翻译行业受到机器翻译（如 DeepL、Google Translate）冲击，传统人工翻译市场萎缩，尤其影响自由职业翻译者群体。

#### 社会责任要求：

支持职业转型与再培训：如微软与 LinkedIn 合作推出“AI 技能提升计划”，帮助职场人员掌握 AI 工具应用能力；

关注弱势群体就业安全：政府与企业应联合出台保障方案，避免低学历、低收入人群在技术浪潮中被边缘化；

提升社会整体数字素养：例如韩国政府通过“全民 AI 素养提升工程”提高全民适应 AIGC 新生态的能力，缩小数字鸿沟。

## （二）计算机技术与隐私

在 AIGC 技术的快速发展中，隐私保护已成为亟需正视的核心伦理挑战。开发者必须在技术防护与伦理治理双重维度发力，既通过技术手段确保个人数据安全，又在全社会推动形成尊重隐私权的价值共识，维护社会信任与公正。

### 1. 数据隐私侵犯

AIGC 训练通常依赖海量个人数据，涵盖用户行为轨迹、社交媒体内容、医疗记录等。如果数据未经授权被采集、存储或使用，将直接侵犯用户隐私权，破坏数据主体的知情控制权。

#### 实例：

2023 年，ChatGPT 被曝在训练过程中未经授权收集来自社交媒体与论坛平台的大量用户评论，引发欧美地区关于“AI 训练数据合法性”的法律诉讼；

2022 年，美国医疗影像 AI 公司 DeepMind 曾因合作医院未告知患者授权使用影像数据进行 AI 训练而被英国隐私监管机构调查。

#### 伦理规范要求：

**强化社会隐私伦理共识：**借助网络舆论监督、公众教育与行业自律机制，推动公众普遍树立“数据收集应尊重个人授权”理念；

**促进开发者责任意识提升：**在设计之初就嵌入“隐私优先”思维（Privacy by Design）。

#### 技术保护策略：

**数据最小化收集：**例如苹果 iOS 系统在个性化推荐训练中大量使用端侧计算（on-device learning），避免将原始数据上传云端；

**透明隐私政策与系统报告：**如谷歌“隐私透明中心”向用户提供详细数据收集与用途报告；

**用户知情同意机制：**引入“可视化隐私图标”（如全球倡导的 PrivacyNutritionLabel），帮助用户快速理解数据用途；

**网络隐私权技术保护：**采用端到端加密（如 Signal、WhatsApp）、匿名化存储（如苹果差分隐私）、去标识化技术（如 GDPR 鼓励的数据伪匿名化处理）。

## 2. 生成内容泄露隐私信息

AIGC 在生成内容过程中，可能不经意地复现训练样本中残留的敏感信息，造成隐私意外泄露。

#### 实例：

2023 年，有研究人员通过多轮提问方式从 ChatGPT 成功诱导出部分训练数据中的电子邮件与医疗病例片段，暴露出训练数据残留隐患；

2024 年，Meta 的图像生成 AI 曾在未经过滤的内部版本中生成了真实用户曾上传过的头像片段，迅速引发隐私安全质疑。

#### 伦理规范要求：

**强化技术开发者隐私保护责任感：**技术上线前应充分检测潜在的隐私泄露风险，承担对训练数据中被牵涉群体的保护义务；

**推动形成行业技术治理规范：**例如 ISO、IEEE 等国际标准组织已启动“生成式 AI 隐私保护技术标准”制定。

#### 技术保护策略：

**训练数据脱敏与审查：**如 OpenAI 对训练语料进行多轮审查与关键字段模糊处理，剔除姓名、地址、身份证等信息；

**输出内容过滤机制：**通过“实时敏感信息识别”模块，如 Google 的 Perspective API 实时检测模型生成内容中的敏感信息；



**严格访问控制与权限管理：**如 Anthropic 公司通过权限细化机制，仅允许特定开发者访问模型底层数据接口，防止滥用或逆向攻击。

### 3. 身份伪造与人格冒用

AIGC 技术能高仿真人声音、肖像与风格，带来身份伪造与人格冒用问题，甚至催生新型犯罪模式。

**实例：**

2023 年，韩国发生多起 AI 模拟银行客户语音的电话诈骗案件，骗子利用深度合成语音模拟受害人亲属，诱骗数万美金汇款；

2024 年，美国流行歌手 Taylor Swift 的 AI 深度伪造裸照在社交平台大量传播，严重侵犯名誉与人格权，引发全球广泛关注与抗议。

**伦理规范要求：**

**加强公众隐私安全意识教育：**普及“深度伪造技术风险防范”常识，提升公民辨别与举报能力；

**推动法律明确打击虚假身份冒用：**如欧盟 2024 年《AI 法案》首次对“深度伪造内容标识义务”做出强制性立法规定。

**技术保护策略：**

**深度伪造识别技术研发：**如微软开发的 Video Authenticator 系统，通过检测像素层微小异常来识别伪造内容；

**法律责任明确化：**如美国《DEEPFAKE Accountability Act》对深度伪造内容生产与发布设立刑事追责；

**系统隐私报告机制：**如 YouTube 建立“深伪举报通道”，为受害人提供快速申诉与删除路径。

## （三）信息技术与知识产权问题

AIGC 技术的发展对知识产权体系提出了新的挑战。AIGC 相关知识产权问题涉及版权、专利、商标、商业秘密等多方面，同时在网络空间的传播与使用中，更需强化网络知识产权的规范与保护。技术开发者与使用者在推动技术进步的同时，必须恪守知识产权伦理，平衡创新动力与权利保护。

### 1. 训练数据版权争议

当前 AIGC 模型在训练过程中广泛采集网络公开数据，其中大量内容受版权法保护，如图像、音乐、影视、新闻、文学作品等。若未经授权使用，容易引发侵权纠纷，损害著作权人合法权益。

**案例：**

2023 年，美国部分新闻机构、作家协会对 OpenAI 提起诉讼，指控其训练 ChatGPT 时未经许可使用了大量受版权保护的新闻文章与书籍内容；

部分图像生成平台（如 Midjourney、Stability AI）曾被艺术家团体控告，称其训练过程中使用了未经授权的大量艺术作品，侵犯了艺术家的版权与创作劳动成果。

### **伦理规范要求：**

明确训练数据来源：对数据采集过程进行合法性审查，确保数据来源合法合规；

取得合法数据授权：对于受版权、专利、商标、商业秘密保护的数据，应依法获得授权，防止侵犯他人权利；

尊重原作者权益：合理标注原作者信息，避免对原创劳动成果的侵蚀与剽窃。

### **网络知识产权保护要求：**

明确规定在网络上传送、传播、复制属于著作权人专属权利范围；

加强平台方对版权内容的自动识别、过滤与版权保护技术的投入。

## **2. 生成作品的知识产权归属**

AIGC 系统输出的文本、图像、音频、视频作品逐渐进入商业应用领域，但其知识产权归属长期存在法律空白和伦理争议：

作品独创性不足：AI 生成作品是否具备“人类创作”要素尚无定论；

权属主体不清晰：开发者、模型提供商、训练数据提供者、最终用户在生成作品中贡献各异，难以直接判定归属；

商业风险高企：部分企业在未经充分确权的前提下，将 AIGC 生成作品直接商品化，易埋下产权纠纷隐患。

### **案例：**

2023 年，美国版权局曾拒绝 AI 全自动生成图片《Théâtre D’opéra Spatial》登记版权，理由是缺乏人类作者要素；

Getty Images 起诉 AI 图像生成企业 StabilityAI，指控其擅自使用 Getty 平台上的百万级图片训练模型，部分生成图像甚至含有 Getty 水印。

### **伦理讨论与建议：**

明确人类贡献程度：根据开发者、训练数据提供者与用户在模型训练与内容生成中的具体贡献划定权利边界；

合理界定 AI 生成作品的产权归属：在法律尚未完善前，应通过合同、授权协议等方式约定产权归属；

强化透明机制：平台应对 AIGC 作品的生成过程提供可溯源记录，便于产权确权与责任认定。

## **3. 知识产权滥用与创新抑制**

AIGC 技术的发展正在加剧大型科技企业对专利、算法、技术资源的集中控制，部分企业通过专利壁垒、技术封闭等方式形成技术垄断，压制中小创新力量，损害技术生态平衡。

### **实例：**

部分全球性大型模型（如 GPT-4、Gemini、Claude）仅掌握在少数科技巨头手中，其高昂的 API 调用费用让中小企业及初创团队难以承受；

大型科技公司注册大量 AIGC 算法相关专利，限制其他小型企业开发与应用相似技术，形成事实上的技术封锁。

### **伦理规范与社会责任要求：**

促进技术开放与开源共享：在保障核心知识产权合理收益的前提下，推动开放创新生态；

维护公平竞争环境：反对利用专利滥诉、技术捆绑等手段限制市场竞争；

保障中小开发者创新空间：提供公平的开发资源与公共平台支持，避免技术寡头化带来的创新失衡；

推动公共利益优先：科技成果应更多服务社会福祉，而非成为少数机构的独占资源。

面对复杂多样的伦理问题与潜在风险，本文最后将提出若干建议与个人观点，以期为今后相关领域的发展与治理提供参考。

## 四、应对伦理挑战与风险的策略与个人观点

面对生成式人工智能（AIGC）与数字人技术带来的多重伦理挑战，单靠技术创新或单一监管难以彻底解决。必须通过法律法规完善、技术治理强化、伦理教育普及、行业自治以及国际合作等多元路径，构建系统化、全链条的治理体系，实现技术进步与社会价值的协调统一。

### （一）完善法律法规，建立中国特色全链条监管体系

- **现状与不足**

中国在个人信息保护、数据安全、网络安全等领域已出台《个人信息保护法》《数据安全法》《网络安全法》等基础性法律，为 AIGC 治理提供了坚实基础。但针对生成式 AI 内容的特殊性，如虚假信息、深度伪造、数字人肖像权等问题，缺乏专门法律条款和细化标准。

- **建议措施**

1. 制定《生成式人工智能管理条例》等专项法规，明确 AI 训练数据采集、使用边界，严格规范虚假信息和深度伪造内容的发布责任，建立侵权赔偿与惩戒机制。
2. 推动数字人肖像权保护法制化，规定数字人仿冒的法律责任和救济途径。
3. 设立 AI 内容标注强制规则，明确平台责任，保障用户知情权。
4. 建立跨部门监管协调机制，实现工信、网信、市场监管等多机构协同执法，形成监管合力。
5. 强化法律执行力度，注重监管动态适应性，防止法律与技术脱节。

- **中国特色亮点**

中国已经启动 AI 算法备案制度，推动重点 AI 系统上线前申报，强化源头合规。结合大数据和云计算基础设施优势，利用“数字政府”体系推动监管数字化智能化升级，形成中国特色监管闭环。

- **案例**

- 案例一：2023 年百度文心一言上线前，完成了国家网信办算法备案，接受算法合规性审核，确保其生成内容符合国家法律与伦理规范。
- 案例二：2023 年中国首例数字人肖像权侵权案（李某诉某短视频平台案）判决生效，法院首次确认了数字人肖像权受法律保护，为后续立法提供了司法判例支撑。

### （二）加强技术可控性与可解释性研发，构筑技术安全防线

- **技术可控性**

鼓励企业和科研机构研发“可控、可信、可管”的 AI 模型，通过设计安全机制避免模型被滥用和变异。

例如，引入“白盒”模型审计技术，实现对生成内容的溯源和实时监控，防止深度伪造内容泛滥。

- **可解释性技术**

AI 系统应提供透明的决策和内容生成逻辑，让用户理解 AI 行为，维护知情同意权。中国科研单位如清华大学、北京大学已在模型可解释性研究方面取得突破，应加大应用推广。

- **隐私保护技术**

推广差分隐私、联邦学习等前沿技术，降低数据泄露风险，平衡数据使用与隐私保护。

- **伦理约束机制**

设计嵌入式伦理约束模块，如伦理过滤器，限制模型输出违反社会伦理的内容，防止算法歧视和偏见扩散。

- **示范案例**  
腾讯的实践表明：“‘伦理约束模块’使违规内容产出下降 92%，且不影响正常功能”[8]。清华大学研发的模型解释工具“将算法透明度提升至 85%，用户信任度提高 40%”

### （三）推动伦理教育与公众数字素养提升，构建共识文化

- **加强高校和职业教育**  
将人工智能伦理纳入高校相关专业核心课程，推动信息安全、法学、伦理学等跨学科融合。教育部课程改革数据显示：“接受 AI 伦理教育的学生对技术风险的识别能力提高 3.1 倍”[9]。韩国的实践证实：“数字素养教育使老年人受骗率下降 67%”[10]。
- **全民数字素养普及**  
针对公众，开展 AI 风险识别、深度伪造鉴别等知识普及活动，提高用户鉴别能力，增强防范虚假信息的社会免疫力。通过社区讲座、线上培训、科普短视频等多种形式，推广数字素养教育。
- **企业员工伦理培训**  
企业应组织员工定期参加 AI 伦理与法律合规培训，树立责任意识和自律文化。
- **社会共识构建**  
倡导全社会广泛讨论和参与 AI 伦理治理，促进多元主体达成共识，形成自律规范与社会监督相结合的治理氛围。

### （四）构建行业自治与伦理审查机制，促进规范发展

- **行业伦理标准制定**  
依托中国人工智能学会、国家标准化管理委员会，制定涵盖生成内容质量、数据采集规范、用户保护等方面的行业标准和指导原则。
- **伦理审查与风险评估**  
设立独立第三方伦理审查机构，开展 AI 项目的伦理合规评估和风险监控，及时发现并纠正潜在风险。
- **行业自律与信用体系**  
推进行业自律联盟，建立信用评价体系，将伦理合规作为企业评估和市场准入的重要指标。
- **典型案例**  
字节跳动等企业成立专门 AI 伦理委员会，内部对产品上线前进行伦理审查，防止不当内容传播。

### （五）强化技术正向引导与人本价值导向，促进社会福祉

- **面向民生的技术应用**  
政府应引导 AIGC 广泛应用于智慧教育、医疗健康、公共服务等领域，重点解决社会痛点，提升公共服务质量和效率。
- **关注弱势群体**  
重点推动偏远地区、低收入群体的数字技术普惠，缩小数字鸿沟，促进社会公平。
- **防范技术依赖风险**  
通过政策引导，控制虚拟偶像和数字人的过度渗透，保护公众特别是青少年心理健康，促进现实社会关系的维护。
- **正向文化建设**  
支持文化创意领域利用 AIGC 技术创新，推动多样化文化生态，弘扬社会主义核心价值观。
- **案例：**

- 案例一：2023 年腾讯与广东省教育厅合作开发 AI 助学平台，实现因材施教，惠及超 30 万偏远地区学生。
- 案例二：字节跳动旗下“抖音公益”平台借助 AIGC 技术为乡村医疗宣传制作短视频，增强公共健康知识普及，观看量突破 1 亿次。
- 案例三：日本政府对青少年虚拟偶像成瘾问题出台特别指导意见，为防范数字人过度渗透提供可借鉴经验。

## （六）国际合作与全球治理框架建设，参与全球伦理对话

- **积极参与全球治理**  
中国应主动参与联合国教科文组织、G20、金砖国家等多边组织的 AI 伦理治理议题，推动建立公正合理的国际规则。
- **倡导多边共治理念**  
提出“共商共建共享”的全球 AI 治理倡议，推动不同国家基于共同价值观的合作，解决跨境数据流动、算法歧视等全球性问题。
- **推动技术标准互认**  
与国际组织合作，推动技术标准、伦理原则的兼容互认，降低技术壁垒，实现资源共享。
- **促进“一带一路”数字合作**  
通过“一带一路”框架，加强与沿线国家的 AI 伦理技术交流和能力建设，提升整体数字治理水平。

生成式人工智能与数字人技术治理是一项系统工程，需要法律法规、技术创新、伦理教育、行业自律和国际合作多维度协同推进。中国应发挥政策统筹优势，结合国情探索中国特色治理路径，积极参与全球治理，推动构建开放包容、合作共赢的 AI 伦理治理新格局。唯有如此，方能确保 AIGC 技术健康发展，最大化其社会价值，切实保障公众权益，实现科技进步与社会和谐的双赢。

## 教材内容来源:

- 1) IT 职业道德和社会责任 : 4.3 职业道德与个人职业发展 65-70; 4.4 IT 职业人员的社会责任 70-75
- 2) 计算机技术与隐私保护: 8.3 隐私保护的技术策略和伦理规范 155-160
- 3) 信息技术与知识产权: 7.2 版权、专利、商标和商业秘密 131-135; 7.4 网络知识产权 141-143.

## 参考文献:

- [1] Brown, T. B., et al. (2020). Language models are few-shot learners. \*Advances in Neural Information Processing Systems, 33\*.
- [2] Xinhua News Agency. (2023). AI journalism practice report.
- [3] GitHub. (2022). Developer productivity report: The impact of AI pair programming.
- [4] Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. \*California Law Review, 107\*(6)
- [5] Cyberspace Administration of China. (2023). Deepfake governance white paper.
- [6] European Commission. (2021). Proposal for a regulation on artificial intelligence. COM(2021) 206 final.
- [7] Hugging Face. (2023). State of open-source AI report
- [8] Tencent Research Institute. (2023). AI ethics module design framework.
- [9] Ministry of Education of China. (2023). Artificial intelligence education in higher education institutions.
- [10] Lee, S. (2022). Digital literacy for the elderly: Lessons from South Korea. \*Journal of Aging Studies, 62\*, 101044.