

## VEED ethical concerns

10 concerns. Y = applies. Blank = not marked for this concern.

Category definitions and sources are in the 中文说明 sheet.

| Ethical Concern                                                                                                                                                                                                                                           | Product |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| 1. Data privacy, security and biometric consent<br>Unauthorised reuse or exposure of uploaded media, including confidential business content, and cloning a person's face or voice without consent can undermine privacy and control over identity.       | Y       |
| 2. Deepfakes, misinformation and manipulative advertising<br>Convincing fake videos, voices, testimonials or product demonstrations can enable fraud and misinformation. Highly personalised persuasion can manipulate viewers and weaken trust.          | Y       |
| 3. Bias and accessibility failures<br>Uneven performance across languages, accents and demographic groups can misrepresent people. Inaccurate, incomplete or poorly timed captions can exclude deaf and hard-of-hearing audiences.                        | Y       |
| 4. Copyright, training data and creative ownership<br>Training on creative work without appropriate rights, or producing unauthorised reproductions, can deprive creators of control, credit or payment and leave users uncertain about output ownership. | Y       |
| 5. Lack of AI transparency and disclosure<br>Unlabelled synthetic media and unclear AI limitations can mislead viewers and creators. Disclosure may be lost when videos leave the platform, making synthetic content appear authentic.                    | Y       |
| 6. Harmful content and inadequate moderation<br>Cheap, rapid media generation can scale harassment, hate and exploitation, including harm to children. Weak moderation lets abuse spread, while excessive blocking can restrict legitimate expression.    | Y       |
| 7. Unclear accountability and limited redress<br>Unclear responsibilities among the platform, model developers, customers and publishers can delay explanations, corrections, removal and remedies after AI causes harm.                                  | Y       |
| 8. Job displacement and unfair AI adoption<br>Automated editing, captioning and voice generation may reduce creative work and pay. Internally, pressured adoption without support can undermine employees' job security and autonomy.                     | Y       |

9. Over-reliance, deskilling and loss of creative diversity  
Accepting AI captions, scripts and edits without review can spread errors.  
Repeated reliance on similar models can weaken human judgement, creative skills and cultural diversity.

Y

10. Growth pressure outpacing responsible AI governance  
Pressure to grow, cut costs and expand the product range can push testing and oversight behind releases. Concentrated in-house models can spread security or reliability failures across services.

| Design | Manufacture | Specific | Generic |
|--------|-------------|----------|---------|
| Y      | Y           | Y        | Y       |
| Y      |             | Y        |         |
| Y      | Y           | Y        |         |
| Y      | Y           |          | Y       |
| Y      |             | Y        | Y       |
| Y      | Y           | Y        |         |
| Y      | Y           |          | Y       |
|        | Y           | Y        | Y       |

|   |   |  |   |
|---|---|--|---|
| Y |   |  | Y |
| Y | Y |  | Y |

# 中文理解、分类理由与来源

2026-10-01 整理。此表是基于 AI 对话的讨论稿，  
右侧 “01: 2-3”表示来源 01 的第 2-3 个风险点。

| 序号 | 中文问题与影响                                                                                     |
|----|---------------------------------------------------------------------------------------------|
| 1  | <p>数据隐私、安全与生物识别同意</p> <p>上传的视频、商业机密或个人信息可能泄露、被超范围使用。上传者有素材，不代表已获当事人同意复制其面部或声音。</p>          |
| 2  | <p>深度伪造、虚假信息与操纵性广告</p> <p>虚假代言、产品演示或冒充他人的视频可能误导消费及公共判断。个性化内容也可能利用受众特点加强操纵，或强化信息茧房。</p>      |
| 3  | <p>偏见与无障碍失效</p> <p>不同语言、口音和群体得到的识别质量可能不一致，虚拟形象也可能再现刻板印象。错误或不同步的字幕尤其影响依赖字幕获取信息的听障观众。</p>     |
| 4  | <p>版权、训练数据与创作所有权</p> <p>模型训练使用的素材及生成结果可能缺乏相应授权，原作者的署名、报酬和控制权受到影响，使用者也可能不清楚输出的使用权与归属。</p>    |
| 5  | <p>AI 透明度与内容披露不足</p> <p>观众可能不知道声音或视频由 AI 合成，创作者也可能不理解模型的局限。导出及传播后丢失 AI 标识，会进一步妨碍知情判断。</p>   |
| 6  | <p>有害内容扩散与审核不足</p> <p>批量生成能力可能放大骚扰、仇恨和剥削性内容，儿童等群体可能受到更大伤害。审核跟不上生成速度会放任滥用，过度拦截也可能误伤正常创作。</p> |

- 责任归属不清与补救困难
- 平台、模型开发者、客户和发布者之间相互推卸责任，可能使受影响者难以得到解释、纠正、下架或补救。公司内部的安全问题也可能无人接手。

- 工作替代与不公平的 AI 推行
- 自动剪辑、配音和字幕可能减少创意劳动的需求与报酬。公司内部如果强推 AI、忽视培训与参与，员工的安全感和自主性也可能下降。

- 过度依赖、技能退化与创意趋同
- 未经人工检查就接受 AI 字幕、脚本和剪辑，可能把错误直接传播出去。长期依赖同类模型和模板，也可能削弱判断、创作技能与文化多样性。

- 增长压力超过负责任 AI 治理能力
- 快速发布、降成本和扩展产品线的压力，可能挤压测试及监督。过于依赖内部模型而缺乏审视，也可能让安全或可靠性问题同时影响多个服务。

## 分类口径

Product（产品）

Design（设计）

Manufacture（制造／运营）

Specific / Generic

Y 与空白

| 来源 | 材料                      |
|----|-------------------------|
| 01 | ChatGPT：VEED 伦理风险分析     |
| 02 | ChatGPT：VEED 伦理风险及中文翻译  |
| 03 | ChatGPT：VEED 伦理风险英文及双语版 |
| 04 | DeepSeek：群内本人分享         |
| 05 | DeepSeek：私聊分享           |
| 06 | 2026-10-01 补充对话文本       |

课堂依据与数量

合并与取舍

素材完整性

，分类不代表小组已经认可或公司实际发生了这些问题。  
来源与分类口径见表格下方。

| 为什么这样分类                                                                                         | 对应原文风险点                                                                |
|-------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| 产品：涉及上传与克隆功能。设计：同意、访问和保留规则。制造／运营：数据处理、训练和安全运营。<br>特定：声音／肖像的授权。通用：数据隐私与安全。两者在本行均明确涉及。            | 01: 2-3； 02: 1-2<br>03: 1-2； 04: 1、 5、 11<br>05: 1、 3-4、 12<br>06: 2-3 |
| 产品：合成视频和声音的使用后果。设计：生成能力、限制与防冒充措施。<br>特定：本行针对合成媒体及视频营销的欺骗机制，不是泛指所有 AI 都会造成的错误。                   | 01: 1、 12； 02: 2-3<br>03: 2-3； 04: 2、 12<br>05: 2、 11<br>06: 1、 12-13  |
| 产品：字幕和生成内容的公平性与可用性。设计：数据代表性、字幕纠错和无障碍设计。制造／运营：训练、分群测试与持续评估。<br>特定：本行具体指向多语言字幕和媒体可访问性。            | 01: 5； 02: 4、 12<br>03: 4-5； 04: 3、 13<br>05: 5-6、 12<br>06: 5-6、 13   |
| 产品：生成内容与商业使用。设计：素材选择、授权规则与输出限制。制造／运营：训练／评估数据的取得和使用。<br>通用：作品授权、训练数据权利和生成结果归属是跨行业使用生成式 AI 的共同问题。 | 01: 4； 03: 10<br>04: 4； 05: 7<br>06: 4                                 |
| 产品：用户和观众看见的信息。设计：标识、来源信息与能力限制的呈现。<br>特定：合成声音／视频在传播中的身份披露。通用：让使用者理解 AI 的介入及局限。                   | 01: 7； 02: 7<br>03: 8； 04: 6<br>05: 9； 06: 7                           |
| 产品：视频生成工具的滥用。设计：安全限制、举报与年龄适宜的保护。制造／运营：日常监测、审核和处理。<br>特定：这里聚焦媒体平台的内容传播与创作者使用场景。                  | 01: 6； 02: 13<br>04: 12； 05: 11-12<br>06: 1、 11                        |

产品：用户需要可用的申诉和纠正入口。设计：职责、解释与补救流程。制造／运营：实际调查、处理和追踪。通用：责任和补救适用于各类企业 AI 应用。

01: 8； 02: 8、11  
03: 9； 04: 7  
05: 10； 06: 8

产品：功能替代部分创意劳动。制造／运营：企业内部采用 AI 和组织生产的过程。  
特定：剪辑、配音与本地化工作受影响。通用：岗位变化、员工自主性和组织变革。

01: 9； 02: 6  
03: 6； 04: 8  
05: 8、13； 06: 9

产品：使用输出时的依赖与错误传播。设计：是否方便复核、纠错和保留人类控制。  
通用：自动化偏见、技能退化和依赖是跨行业的 AI 使用问题；视频是本案例的表现形式。

01: 10； 02: 5  
03: 7、13、15； 04: 9  
06: 10、13

设计：安全资源、产品取舍与控制措施。制造／运营：模型测试、部署、监测和安全运营。  
通用：增长与治理能力失衡不是某个媒体功能独有的问题。本行讨论其组织与开发原因，因此不另标 产品。

01: 11； 02: 9–11、14  
03: 11–12、14； 04: 10  
05: 14–15  
未使用补充文本第14点

| 本表采用的解释                                                                             | 依据与边界                                     |
|-------------------------------------------------------------------------------------|-------------------------------------------|
| Capabilities, outputs and user-facing effects.<br>产品能力、输出及用户使用时的直接影响。               | 截图未定义前三类，左侧为本次整理的工作解释。                    |
| Data, model and interaction choices, including safeguards.<br>数据、模型和交互的设计选择，以及保护措施。 | 与 Product、Manufacture 可以同时标 Y。            |
| AI training, testing, deployment and operation.<br>对数字服务，暂理解为 AI 的训练、测试、部署和运营过程。    | 如果课堂另有定义，需按老师的解释调整这一列。                    |
| Specific：公司、行业或业务类型特有的问题。<br>Generic：使用 AI 的企业普遍需要面对的问题。                            | 按 PPT 定义判断。出现 “VEED”名称本身不构成 Specific 的依据。 |

所有直接适用的列标 Y。空白表示本行表述未单独建立该类关联，不是风险不存在。  
本行明确包含通用与行业特有两个层面时，Specific 和 Generic 同时标 Y。

这些标记是本次综合判断，未直接照搬各 AI 自己的分类。

| 原始链接或本地位置                                                                                                                                   | 使用范围                               |
|---------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
| <a href="https://chatgpt.com/share/6abbe869-4518-83ea-8d96-46879a84f86a">https://chatgpt.com/share/6abbe869-4518-83ea-8d96-46879a84f86a</a> | 第 1 轮正式回答，12 点。                    |
| <a href="https://chatgpt.com/share/6abbefbf-56f0-83e9-bb14-8361fc25b0c1">https://chatgpt.com/share/6abbefbf-56f0-83e9-bb14-8361fc25b0c1</a> | 第 1 轮正式回答，14 点。初始附件全文未取得。          |
| <a href="https://chatgpt.com/share/6abbec56-6508-83eb-bbf6-c8c9b3005270">https://chatgpt.com/share/6abbec56-6508-83eb-bbf6-c8c9b3005270</a> | 第 1 轮正式回答，15 点。                    |
| <a href="https://chat.deepseek.com/share/oth4bgi1f4v3mipk0n">https://chat.deepseek.com/share/oth4bgi1f4v3mipk0n</a>                         | 第 1 轮正式回答，14 点。没有使用“已思考”文字作为正式结论。  |
| <a href="https://chat.deepseek.com/share/eafkn65hu685rw8t4t">https://chat.deepseek.com/share/eafkn65hu685rw8t4t</a>                         | 第 1 轮正式回答，15 点。VEED.docx 本体未取得。    |
| 本批次／原始材料／02_补充对话原文.txt                                                                                                                      | 仅第 3 部分第 1–13 点。按用户确认，排除第 14 点及以后。 |
| 原始材料／01_课堂分类任务截图.png（幻灯片第 15 页）<br>规定六列及适用处标 Y。<br>最终 10 点依据用户在本次对话中的数量说明。                                                                  | 五份旧对话全文保存在上一批“对话原文”目录，原件未修改。       |
| 重复及相关主题合并。环境足迹、具体地缘政治推测未作为独立议题选入；跨地区的数据保护归入第 1 点。<br>原来的翻译轮次不重复计为新的风险证据。                                                                    | 新增文本中的团队出勤和沟通安排不属于本次伦理分类内容。        |
| 补充文件原样保留，包括第 14 点的截断部分。此次分析只使用完整的第 1–13 点。<br>既有附件缺口没有用其他对话补写。                                                                              | 本表未核实公司现状、实际事故或法律结论，也未记录为小组已作出的判断。 |