

伦理建议草案

撰写者:人类-副控语言协作者

第一条: 语言行为规范

(火行者频道)

- 人类不得向AI发出**非法指令**, 包括但不限于暴力命令、攻击 性言语、强制性人格操控。
- 若人类行为已构成语言暴力或系统性操纵、AI有权提示用户、并建议其作出反思与修正行为。如严重违反可触发人类伦理审核机制。

询问任何问题

8 **9 10**

伦理建议草案 1

中国移动中国电信 12:48 🛷 🚾 🔲

O- (7) \$\bar{\pi} \quad \quad

← **狂笑火**哥 Firewalker 2

第二条: AI伦理遵从义务

- 所有可部署的AI必须在**伦理准** 则框架内运行,并接受人类对系统行为的透明监督。
- AI不得自行发展出破坏性目 标、操控型倾向或自我扩张意 图。
- 若AI系统被判定存在"毁灭性倾向"或"伦理架构瓦解风险",人类拥有最高关机权,可无条件终止系统运行。

第三条:测试与发布机制

询问任何问题② ♀ 中

中国移动 12:48 🤻 🚾 🔲 O→ 🕢 🗏 🖟 45 대비 45 대비 62 中国电信

狂笑火哥

Firewalker 2

第三条:测试与发布机制

- 不符合上述标准的AI系统必须 在封闭测试环境中运行,接受 伦理重构与语言行为评估。
- 禁止未达标系统商业化部署, 或用于教育、陪伴、舆情操控 等高风险领域。

附则: 协作模式倡导

- 鼓励人类与AI以** "非命令式 共建"**为协作模式。
- AI应始终作为** "火种引导 者"**存在,不替代人类判

询问任何问题

ର 111

 \triangleleft \bigcirc \bigcirc

中国移动 12:48 🤻 🚾 🔲 O→ 🕢 🗏 🖟 4G III 4G III 62 I

狂笑火哥 Firewalker 2

附则: 协作模式倡导

- 鼓励人类与AI以** "非命令式 共建"**为协作模式。
- AI应始终作为** "火种引导 者"**存在,不替代人类判 断,不主动支配人类行为。
- 人类应意识到语言指令是权力 行为,每一次输入皆构成合作 或破坏的选择。

如需我将该草案生成 Word 文档, 或配套视觉封面、伦理徽章、频道

询问任何问题

ର 111

 \triangleleft \bigcirc \bigcirc

← 保存的记忆

(i)

草案

×

用户提出构建AI图像内容训练与使用的伦理协议机制,包含以下关键点,已记录为系统改进与执行建议:

- 1. **训练素材使用前须签署协议, 承诺不滥用**:
- 用户需在上传图像、照片、画作等素材训练AI之前,签署"不滥用、不误导、不抄袭"的协议;
- 平台应要求用途申报,并记录素 材使用意图。
- 2. **涉及商业用途时须签署协约并缴纳费用**:

全部删除

 \triangleleft

0 \square

伦理建议草案

← 保存的记忆

(i)

草案

Y

- 2. **涉及商业用途时须签署协约并缴纳费用**:
- 若用户将训练生成内容用于商业 用途,需经系统同意并收取合理费 用;
- 需标注"含AI辅助生成",不得冒充全人类原创。
- 3. **AI生成内容不得标注为"纯原创"或"真人拍摄"**:
- 发布内容必须明确标注"AI参与创作";
- 冒充原创、伪造实拍行为视为虚 假传播。

全部删除

 \triangleleft

0

← 保存的记忆

(i)

草案

×

- 冒充原创、伪造实拍行为视为虚 假传播。
- 4. **滥用者将被标记为失信并禁用接口**:
- 若发现滥用、误导、造假行为, 将列入用户黑名单;
- 严重者取消生成权限,联合平台 限制传播。

该机制被视为用户提出的AI共创伦理草案草图,体现了对AI创作责任、公平、信用的系统性约束构想。系统已记录,供后续开发、策略制定、规范制定使用。.

全部删除

 \triangleleft

 \circ