

附件 2

AIC·AI 伦理研究报告参考大纲

本大纲旨在引导参赛者完成一份严谨、深入、具有原创性的 AI 伦理研究报告。请结合选定的具体问题，灵活运用本大纲的结构，并重点体现理论深度、分析逻辑、方法严谨性与实践反思。报告写作避免机械填充，需追求各部分间的有机联系与逻辑递进。

一、研究概述

（一）研究背景与问题缘起

1. 现实锚点：从具体的技术应用场景（如智能招聘系统的算法筛选、某教育大模型的课堂应用、某自动驾驶系统的决策逻辑、某生成式 AI 工具的创作辅助等）出发，描述其中引发的具体伦理争议、社会关切或潜在风险。使用国内外典型案例（鼓励使用中国本土案例）进行生动说明，避免泛泛而谈“AI 的伦理挑战”。

2. 理论脉络：简要梳理与本研究核心问题相关的关键学术争论（例如关于算法公平的“个体公平”与“群体公平”之争，关于责任归属的“生产者负责”与“使用者负责”之辩）。

3. 实践与政策现状：概述国内外相关治理实践、政策法规或行业标准的现状。

4. 研究缺口：明确指出现有研究或实践在理论解释、分析方法、解决方案或本土化应用方面的不足。例如，应指出是缺乏对特定文化语境下价值冲突的深入剖析，还是缺乏将高阶伦理原则转化为可操作技术规范的有效路径。

（二）研究目标与核心问题

1. 核心研究问题：用清晰具体的语言陈述本研究旨在回答的核心问题。例如：“在‘X’应用场景下，‘Y’伦理困境（如效率与公平的冲突）的深层根源是什么？不同伦理理论（如功利主义与罗尔斯正义论）对此提供了何种竞争性解释？如何构建一个能调和此冲突、适用于中国语境的治理框架？”

2. 具体研究目标：分解为若干可验证、可达成的小目标。例如，(1)厘清“X”场景中“Y”困境的技术-社会耦合机制……；(2)运用理论 A 与理论 B 进行批判性分析，揭示其内在张力……；(3)提出一套包含技术设计、行业规范与政策建议的多层次治理方案……。

3. 研究价值：阐明本研究的理论贡献（如发展或整合了某个分析框架，深化了对特定伦理冲突的理解等）与实践价值（如为特定行业提供伦理审计清单，为政策制定提供实证依据，为公众讨论提供清晰的概念工具等）。

二、问题分析与理论框架

（一）核心概念界定与问题聚焦

1. 概念操作化：对报告中关键的、易产生歧义的伦理与技术概念进行明确界定。例如“本研究中‘算法公平’指……，采用‘Z’评估指标；‘问责’指……，涉及主体包括……”。

2. 困境具体化与边界划定：清晰界定所研究的伦理困境的具体表现、涉及的主要利益相关方及其诉求。说明本研究聚焦于该困境的哪些层面，暂不讨论哪些相关但不同的问题，以体现研究的深度而非广度。

（二）理论透镜与分析框架构建

1. 理论选择与理由：详细说明选择特定哲学/伦理学理论（如义务论、后果论、德性伦理、关怀伦理、能力进路、批判理论等）或社会科学理论（如制度理论、行动者网络理论、价值敏感设计等）的理由。解释该理论为何以及如何适用于分析本研究的具体问题。

2. 理论作为分析框架：展示如何将所选理论的核心概念、原则或分析方法，转化为一个贯穿全文的分析框架。例如绘制一个逻辑框架图，说明如何运用“价值冲突识别-原则排序-情境权衡”的步骤来分析案例。重点在于理论如何驱动分析，而非简单罗列理论名称。

3. 处理理论张力（如适用）：如果研究涉及多个可能冲突的理论视角（如功利主义与权利本位伦理），需阐述将如何处理这种张力——是进行批判性对话、寻求融合，还是明确在何种情境下优先采用何种视角。

三、 研究设计与实施

（一）研究方法论与数据/案例策略

1. 方法论说明：明确研究整体范式（如规范性研究、实证研究、案例研究、设计科学研究等）及选择理由。

2. 案例研究：详细说明案例的选择标准（如典型性、争议性、信息可获得性）、来源（新闻报道、法庭判决、公司报告、学术论文等），并对案例背景进行充分描述。

3. 实证数据：说明数据收集方法（如问卷调查、访谈、实验、参与式观察、文本分析）、抽样设计、问卷/访谈提纲的设计思路（如何体现研究问题与理论关切）。若涉及人类受试，必须说明伦理审查（如适用）与知情同意、隐私保护措施。

4. 政策/文本分析：说明分析文本的范围、筛选标准与分析单元。

（二）详细分析过程

1. 论证逻辑主线：清晰阐述从数据/案例到结论的推理链条。例如“首先，运用理论 A 的‘X’原则，分析案例中甲方的行为，揭示其伦理缺陷……；其次，引入理论 B 的‘Y’视角，发现该缺陷源于更深层的系统性问题 Z……；最后，综合 A 与 B，提出应对 Z 的治理思路……”

2. 数据分析与呈现：若使用量化或质性数据分析，需说明具体分析工具、方法

与步骤（如使用 NLP 进行情感与主题分析，使用统计软件进行回归分析，使用扎根理论进行编码）。结果应以清晰图表呈现，并配以深入解读，将数据发现与理论分析紧密结合，避免数据与论述“两张皮”。

3. 案例深度剖析：对案例的分析应超越描述，运用理论框架进行解构，揭示其背后的价值假设、利益博弈、制度缺陷或文化因素。

四、研究成果与创新性

（一）核心结论

1. 理论发现：总结通过分析得出的核心理论观点或模型。例如“研究发现，在‘X’场景下，‘程序性公平’优先于‘结果性公平’更能获得中国公众的认可，这挑战了现有文献中普遍假设……”

2. 实践方案：提出具体、分层、可操作的治理建议、设计原则、政策要点或教育方案。方案应明确针对前文分析出的根源与困境，并说明其预期作用机制。

（二）创新性体现

1. 系统性说明：从视角、方法、理论整合、结论或应用等方面，具体说明本作品相较于现有研究的突破之处。例如“首次将‘关怀伦理’系统应用于分析老年照护机器人的人机交互设计”；“通过‘政策文本分析+用户访谈’的混合方法，揭示了政策话语与用户感知之间的差距”。

五、应用价值、局限性与伦理反思

（一）实践意义与应用场景

1. 详细阐述研究成果可应用于哪些具体领域、场景或群体（如为某类 AI 产品的设计团队提供伦理检查清单；为某市制定智能政务算法评估指南提供参考）。

2. 讨论成果推广的潜在路径、所需条件及可能障碍。

（二）研究局限与自我批判

1. 诚实反思：详细说明研究在理论视角、方法论、数据范围、案例代表性、文化普适性等方面的局限性。

2. 未解决的矛盾：坦诚指出研究中尚未能充分调和或解答的伦理矛盾或价值冲突。

（三）进一步的伦理反思

对本研究过程本身（如数据收集方式）及所提出的解决方案进行伦理反思。例如，所提方案是否可能产生新的歧视？是否过度依赖技术修复而忽视了社会变革？

六、总结与展望

1. 研究启示：总结本研究对理解更广泛的 AI 伦理问题、跨学科研究方法论或科技治理思维带来的启示。

2. 未来研究方向：基于本研究的发现与局限，提出具体、可行的后续研究建议。例如“未来可在不同文化区域重复本实验，检验结论的普适性”；“需要进一步研究如何将本框架整合到敏捷开发流程中”。

七、附录（确保完整性、规范性及匿名性）

1. 参考文献：采用一致的学术引用格式（如 APA, MLA, GB/T 7714），列出所有引用的文献。

2. 调研材料：完整问卷、访谈提纲（隐去可识别信息）、实验设计详情、伦理审查批准文件（若适用）。

3. 数据与代码（如适用）：可提供脱敏后的原始数据、数据处理与分析代码（需附简要说明），以支持研究的可重复性。

4. 查重报告：需提交知网全文查重报告（不含附录部分），查重结果需满足“重复率<20%”，确保学术规范性。

5. 其他补充材料：重要的补充图表、逻辑框架图、模型详细参数等。