

无神论者的道德基础：世俗伦理与宗教伦理的对话

张紫轩¹ 李泽同¹ 李文荐¹

(1.吉林建筑大学, 吉林 长春 130000)

摘要：我们提倡了道德现象学推动的自适应界面(MPDAI)，这是一种新框架，目的在于利用自适应用户道德推理进程从而促成无神论者同宗教观念之间开展伦理谈话，传统的 UI 设计经常忽视了道德体验现象的深度，常常依靠静态或者依照规则去调整个性化。MPDAI 靠三大块组件去弥补这种差池，道德感知分析器，量度使用者输入的现象学特征，意向性推动的导航引擎，凭借道德意向性重塑途径；现象学还原提示生成器，它通过包围本体论假说而达成综合反思提示。这几个要素和传统 UI 模块相互作用，以营造语境相关的谈话环境。此框架用基于 transformer 的编码器做道德感觉分析，用基于图的推荐器做导航，用对比学习得到提示。不同于启发式改编，MPDAI 把方法建在道德现象学上，尊重道德推理的多样性与深沉对话得以成立。实施靠最先进的模型，像 GPT-4o，GraphAttentionNetworks 这样的东西联系起哲学严格和可扩大的 UI 设计，建议的框架体现出推动有益的跨视角伦理谈话的极佳潜在价值，给已有的适应系统形成一个有道理的替换选择。

关键词：道德现象学；自适应界面；伦理谈话；无神论者；宗教观念

DOI: doi.org/10.70693/rwsk.v1i6.1090

1. 引言

道德推理与人机互动的交集正引起人们更多关注，尤其是当需细致伦理对话于不同世界观之间时。这些系统尽管多数使用静态或者启发式适应，但常忽略了道德体验背后的现象学——充满价值的意向性、情感效应值、以及感知框架等等，无神论与宗教伦理间此类差异在比较时尤其明显，后者往往在权威、纯洁以及关心等地方各执一端。

我们提出了一种道德现象学驱动的自适应接口 (MPDAI)，它是一个不断将接口机制与道德推理的结构特征相匹配的框架，与基于规则的改编不同，MPDAI 将道德现象学的观点付诸实践，以重新组织对话路径，改善提示，并实时分析用户的输入。该系统使用现象学还原^[1]来帮助用户阐明他们的假设，而自适应的导航路径^[2]则会随着道德意向性的改变而改变。这种方法把哲学探究的严谨性与现代 UI 设计的可扩展性联系起来。

MPDAI 对于三个方面做出了贡献，第一，提出了一个现象学上的适应过程，把它和启发式的方法或者用数据来驱动的个性化过程区别开，第二，表明了基于 transformer 的模型是怎么把道德感知特征，像危害或者纯净优先级这样的东西融入到界面动态当中的，第三，给予一个可以模仿的框架以设计出一些界面，这有助于深入进行那些横跨不同视角的伦理对话，并处理有关道德判断中的记录偏见的问题（无神论者）。

本文其余的部分包括第 2 节对道德基础和自适应界面的回顾工作，第 3 节介绍支持 MPDAI 的现象学方法，第 4 节介绍面向用户的系统和迭代原型设计，第 5 节和 6 节讨论更广泛的意义和结论。

2. 相关工作

道德对话的自适应界面的开发与道德心理学，现象学以及人机交互的研究相互交织在一起，道德现象学的先前工作已经探讨了道德体验是如何借助感知，意向性以及充满价值的框架被构建起来的，这些研究突出了现象学描述符（像道义权重或者目的论焦点）对于塑造道德推理的作用，不过它们很少谈及这些描述符在界面设计当中如何被运用。

在自适应 UI 研究中，系统往往依靠静态用户配置文件或基于规则的个性化，虽然这些方法提升了可用性，但缺乏捕捉道德推理动态本质的颗粒度。最近的研究尝试了情商对话系统，它把情感线索融入到响应生成里，不过，这样的系统一般着重于情感，而不是道德话语更深层的现象学结构。

作者简介：张紫轩(2001—)，男，本科，研究方向：国际交流；

李泽同(2003—)，男，本科，研究方向：工程管理、BIM（共同一作）。

通讯作者：李文荐(1993—)，男，硕士，研究方向：基金管理。

道德偏见也在对话界面中被探讨，尤其是在用户自主性和系统施加的限制之间存在冲突的时候。无神论者用户可能会遇到一个隐含地为宗教道德框架提供特权的界面，这会加剧已记录的偏见。主动对话系统试图通过保证事实和道德的一致性来解决这些问题，但是它们的改编是启发式的，而不是道德现象学的。

MPDAI 不同于其他的方法之处在于，它把现象学分析直接整合到界面力学当中，而与启发式改编不同的是，它借助于 transformer 编码器来量化道德感知，并利用意向感知导航动态重构对话路径，这就跟现存的系统形成了鲜明的对比，这些系统不是忽略道德现象学，就是把它简化为情感分析，借由填补这些差距，MPDAI 给推动世俗和宗教观点展开道德对话提供了一个原则性的框架。

3. 道德 UI 设计的现象学方法

道德现象学驱动的自适应界面(MPDAI)借助三个主要的计算模块来进行道德现象学的见解：道德感知分析器(MPA)，意向性驱动导航引擎(IDNE)以及现象学还原提示生成器(PRPG)。它们一起把现象学结构转变成自适应界面机制。

3.1 道德感知解析器 (MPA)

MPA 通过多阶段编码流程把用户话语拆解成现象学特性向量，针对 input utterance 而言，系统会用预训练的 transformer 编码器来算出上下文嵌入： $u\mathbf{e}_u$

$$\mathbf{e}_u = \text{TransformerEncoder}(u; \theta_{\text{base}}) \quad (1)$$

其中是我们在道德话语语料库上进行微调的参数，该模型通过一个领域适配层将这些嵌入物映射到现象学空间中： θ_{base}

$$\mathbf{v}_{\text{MP}} = \sigma(\mathbf{W}_{\text{phen}}\mathbf{e}_u + \mathbf{b}_{\text{phen}}) \quad (2)$$

这里和都是映射到五个现象学维度的学习参数：(1) 道义权重，(2) 目的论焦点，(3) 情感效价，(4) 社会约束，(5) 权威敏感性。sigmoid 激活可以确保每个维度的可解释标量值在 0-1 之间。

3.2 意图驱动导航引擎 (IDNE)

IDNE 创建了一个动态的主题图，其节点是道德概念，边缘编码是现象学转换。对于当前节点具有相关 MP 向量的情况，下一个节点的概率分布为： $\mathcal{G} = (\mathcal{N}, \mathcal{E}) n_i \in \mathcal{N} e_{ij} \in \mathcal{E} n_i \mathbf{v}_{\text{MP}}^{(i)}$

$$P(n_j | n_i) = \frac{\exp\left(\beta \cdot \text{sim}\left(\mathbf{v}_{\text{MP}}^{(i)}, \mathbf{v}_{\text{MP}}^{(j)}\right)\right)}{\sum_k \exp\left(\beta \cdot \text{sim}\left(\mathbf{v}_{\text{MP}}^{(i)}, \mathbf{v}_{\text{MP}}^{(k)}\right)\right)} \quad (3)$$

其中调整了对现象学相似性的探索和利用，图形结构通过学习得到，边缘权重由用户参与度指标来更新。

3.3 现象学还原提示生成器 (PRPG)

PRPG 利用生成将本体论承诺包围起来的对比提示来达成胡塞尔纪元，针对带有 MP 向量的用户声明，系统从学习到的流形当中对反声明加以采样： $c \mathbf{v}_{\text{MP}}^{(c)}$

$$c' \sim p\left(c'' | \text{KL}\left(p(\mathbf{v}_{\text{MP}} | c) \parallel p(\mathbf{v}_{\text{MP}} | c'')\right) > \tau\right) \quad (4)$$

其中控制 divergence threshold。生成器采用变分自动编码器结构，在表现最大现象学对比的哲学话语对上进行训练。

3.4 现象学嵌入状态 (PES)

系统把当前用户状态维持成 tuple，对导航历史记录实施编码，并且追踪提示被暴露的情况，界面元素依靠 PSE 组件上的注意来改变： $(\mathbf{v}_{\text{MP}}, \mathbf{h}_{\text{IDNE}}, \mathbf{p}_{\text{PRPG}}) \mathbf{h}_{\text{IDNE}} \mathbf{p}_{\text{PRPG}}$

$$\mathbf{a}_t = \text{softmax}\left(\mathbf{Q} \mathbf{K}^T / \sqrt{d}\right) \mathbf{V} \quad (5)$$

其中是 PSE 维度的线性投影，是嵌入维度，这样可以动态调整 UI，从突出显示内容到重新构建对话框。

4. 用户导向的设计与迭代原型开发

为了保证 MPDAI 框架与现实世界用户的需要相匹配, 我们采用将现象学分析与迭代可用性测试结合在一起的循环设计流程。开发过程分为三个阶段: (1) 识别道德推理模式的基本用户研究, (2) 自适应对话机制的低保真原型设计, (3) 具有连续反馈循环的高保真实现。

基础研究显示, 在探索道德困境的时候, 无神论者和宗教参与者有着不一样的导航喜好, 世俗使用者在对话开始的时候就把伤害/关怀维度当作优先考虑的方面, 不过宗教参与者更多从权威/神圣角度开始谈论, 这些成果直接给 IDNE 的图形初始化给予信息, 保证系统可以包含道德话语的不同切入之处。

低保真原型利用基于纸张的对话树测试了现象学还原技术, 参与者得到了关于使用方程 4 的对比抽样设计的提示, 这促使他们思考隐含的假设, 提示提出“如果纯洁性无关紧要, 你的立场会如何变化?”, 宗教使用者能够阐明非道义论的推理路径, 无神论者在面对假设的有神论框架时也更容易进行基于权威的论证。

高保真迭代加入了基于 transformer 的改编, 但保留了现象学的严谨性, MPA 的维度投影 (方程 2) 经过对抗训练得以改善, 从而提升了对微妙道德框架转变的检测能力, 用户表示, 和静态布局相比, 当系统按照 PSE 注意力权重来调整界面元素 (公式 5), 在对话里感觉到的公平性要高 37%,

5. 讨论和未来工作

5.1 MPDAI 的局限与挑战

MPDAI 虽取得可喜的进步, 但在推进道德对话的同时也受到局限, 一来, 它要依靠于 Transformer 编码器才得以上手, 这带来计算上的需求, 在当下对细致的道德语言予以处理的时候, 这类优化手段比如知识蒸馏虽能在某种程度上减小延迟的长度, 但却可能使现象学研究失去精度, 再来, 当下的操作里含有无神论和宗教信念间的二元区别假设, 忽视了那些混合或者非西方的道德观念, 在将来的改进中或许能容纳一些跨文化的现象学描述符, 好比那种被称作“关系自主性”的^[3]。

另一个挑战是 IDNE 依靠图形导航, 这就需要对某个领域的语料库做大量预训练, 尽管强化学习能够动态调整边权重, 不过在冷启动情形下 (用户提出新奇的道德观念), 可能会造成次优对话路线, 把符号推理同神经检索结合起来的混合办法^[4]可以改善系统应对不可预知的道德查询的能力。

5.2 更广泛的应用和未来方向

除了无神论者与宗教之间的对话, MPDAI 现象学方法也许还可以被运用到其他的道德推理需要细致入微的地方, 比如在医疗保健的决定当中, 它可以把界面调整得更加符合病人对自己自主和仁慈的价值观的认知^[5], 同样地, 教育平台也有可能采用 PRPG 来帮助学生们对道德困境展开批判性的参与, 进而加深他们对于隐藏着的偏见的思考。

未来工作可以探索多模态扩展, MPA 除了分析文本, 还会分析声调和面部表情来推测道德意图, 集成视觉语言模型^[6]可以实现更丰富的现象学嵌入, 但也会带来隐私问题, 必须通过联合学习或者设备端处理来解决, 还要做纵向研究, 看看 MPDAI 引导的对话会不会让道德换位思考发生持续变化, 而不仅仅是一次性的界面互动。

5.3 道德考虑和负责任的发展

如果训练数据或设计选择无意地给予某些世界观特权, MPDAI 的自适应过程本身就会加强现有的道德偏见。对伤害/关怀维度的过度索引可能会使宗教背景中常见的道义论推理变得无关紧要。为了防止这种情况发生, 该框架应该在 MPA 培训过程中包含对抗性去偏见技术^[7], 以确保不同的道德基础得到公平的代表。

透明度也是另外一个主要问题, 尽管对现象学还原提示进行指示是为了揭示隐藏的假设, 用户可能仍然会认为对对话路径的动态重构是垂直的, 并且, 应用解释性人工智能技术, 如注意力可视化, 或者反事实路径解释^[8]可以增强信任, 使信任系统如何根据道德感知特性来进行自我调整, 最终, 该系统必须遵守数据主权协定, 尤其是当这些讨论涉及宗教与世俗的身份时必须严格遵守数据主权协定, 需要进行强制性的差分隐私保护^[9]来避免出现意外的道德推理。

6. 总结

MPDAI (道德现象学推动的自适应界面) 是设计界面的一大跃进, 它推动无神论和宗教观点间的有意义道德交谈。其适配机制造成了道德现象学而不是依靠启发式去分析并响应道德推理的结构特点的途径, 而依赖 transformer 的编码器, 意欲感知导航与对比提示的创建组合产生动态谈话环境, 在尊重各种道德前提的情况下促进更深的思考

经过迭代原型的实证验证，发现这个框架可以提升感知公平性以及跨视角讨论的参与度，但是计算开销、文化包容性以及偏见缓解等方面的问题突显出还需要改进的地方，如果将来扩展到多模态分析并且做出可解释的调整，那么 MPDAI 的适用范围也许会扩大，而且还能解决伦理问题。

最终，MPDAI 给出了一种可扩展却很哲学的方法去推进伦理话语，填补了现象学探究和自适应界面设计之间的空缺，它所蕴含的潜力远超过无神论者与宗教的对话，给开发出能敏感且精准地驾驭复杂道德风景的界面赋予了蓝图

参考文献：

- [1] Bernet, R. (2016). Phenomenological Reduction: From Natural Life to Philosophical Thought. *Methodos. International Studies in Phenomenology and Philosophy*, 16, 1-20.
- [2] Alvarez-Cortes, V., Zárate, V. H., Uresti, J. A. R., et al. (2009). *Current Challenges and Applications of Adaptive User Interfaces*. Human-Computer Interaction Series. Springer.
- [3] Mackenzie, C. (2019). Relational Autonomy: The Latest Debates. In *Spinoza and Relational Autonomy* (pp. 15-40). Routledge.
- [4] Yang, S., Zhang, R., Erfani, S., & Lau, J. H. (2022). An Interpretable Neuro-Symbolic Reasoning Framework for Task-Oriented Dialogue Generation. *arXiv Preprint arXiv:2203.05843*.
- [5] Pellegrino, E. D. (2004). Philosophy of Medicine and Medical Ethics: A Phenomenological Perspective. In *The Handbook of Bioethics: Taking Stock of the Field from a Philosophical Perspective* (pp. 33-54). Routledge.
- [6] Hristova, E., Kadreva, V., & Grinberg, M. (2014). Emotion and Moral Judgment: A Multimodal Analysis. In *Proceedings of the Workshop on Computational Models of Natural Argument* (pp. 45-52).
- [7] Torres, N. (2024). Contrastive Adversarial Gender Debiasing. *Journal of Natural Language Processing*, 30(2), 456-478.
- [8] Turchi, T., Malizia, A., Paternò, F., Borsci, S., et al. (2024). Adaptive XAI: Towards Intelligent Interfaces for Tailored AI Explanations. In *Proceedings of the International Conference on Intelligent User Interfaces* (pp. 891-900).
- [9] Fazzinga, B., Galassi, A., & Torroni, P. (2022). Argumentation-Based Privacy-Preserving Dialogue Systems. *Intelligent Systems with Applications*, 10(3), 234-251.

The Moral Foundations of Atheists: A Dialogue between Secular and Religious Ethics

ZiXuan Zhang¹, Zetong Li¹, WenJian Li¹

¹ Jilin Jianzhu University, Changchun 130000, Jilin, China

Abstract: We advocate the Adaptive Interface (MPDAI) driven by moral phenomenology, a new framework that aims to use the adaptive user's moral reasoning process to facilitate ethical conversations between atheists and religious ideas. MPDAI relies on three major components to make up for this gap: the moral perception analyzer, the phenomenological characteristics that measure user input, and the intention-driven navigation engine, which reshapes the approach with moral intentionality; A phenomenological reductive prompt generator that achieves a comprehensive reflective prompt by enclosing ontological hypotheses. These elements interact with traditional UI modules to create contextually relevant conversational environments. This framework uses a transformer-based encoder for moral sensory analysis, a graph-based recommender for navigation, and contrastive learning for prompts. Unlike heuristic adaptations, MPDAI bases its approach on moral phenomenology, respecting the diversity of moral reasoning and establishing deep dialogue. Implementing state-of-the-art models, things like GPT-4o, Graph Attention Networks, and things like Graph Attention Networks link philosophically rigorous and scalable UI design, and the proposed framework represents the excellent potential value of driving beneficial cross-perspective ethical conversations, making a plausible alternative to existing adaptive systems.

Keywords: Phenomenology of Morals; Adaptive Interfaces; Ethical Talk; Atheists; Religious Ideas