arXiv:2602.03155v1 [cs.HC] 3 Feb 2026

能否让聊天机器人具备美德？探究应用于大语言模型的基于美德的设计方法论

Matthew P. Lad、Louisa Conwill 与 Megan Levis Scheirer

随着大型语言模型（大语言模型）的快速发展，对其社会影响的批评也日益增多。负责任人工智能（RAI）领域的工作聚焦于开发旨在减少伤害的AI系统。为回应负责任人工智能所受到的批评，以及将智慧传统引入人机交互的需要，我们将 Conwill 等人的美德引导的技术设计方法应用于大语言模型。我们为大语言模型编目了新的伦理设计模式，并通过与技术专家的访谈对其进行了评估。参与者认为，这些模式带来了更高的准确性与鲁棒性、更好的安全性、新的研究机会、更多的访问与控制，并减少了浪费。他们的担忧在于，这些模式可能易受越狱攻击、对模型的泛化过于宽泛，以及存在潜在的实现问题。总体而言，参与者反应积极，同时也承认伦理大语言模型设计中涉及的权衡。

CCS 概念：• **以人为中心的计算 → 人机交互实证研究**。

补充关键词　词语与短语：大型语言模型、天主教社会训导、美德伦理学、生成式人工智能、价值敏感设计

ACM 参考格式：
Matthew P. Lad、Louisa Conwill 和 Megan Levis Scheirer。2026。能否让聊天机器人拥有美德？探究应用于大语言模型的基于美德的设计方法论。1，1（2026年2月），7页。https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 引言与相关工作

自 ChatGPT 于 2022 [11] 年 11 月发布以来，大型语言模型（大语言模型）的受欢迎程度迅速攀升。虽然它们能带来诸多生产力方面的益处 [15], ，但也引发了社会和伦理方面的担忧。这些担忧包括环境问题 [3, 9, 16], 、对偏见的担忧 [3, 8] ，以及其他有害输出，例如助长饮食失调 [5], 、自杀 [10], 和精神病 [7], ；对技能退化的担忧 [13], ；以及对通过大语言模型幻觉传播错误信息的担忧 [2]。

人们提出了负责任人工智能（RAI）原则，旨在指导人工智能系统（包括大语言模型）的设计、开发和部署，以减少危害。这些原则包括可解释性、公平性、鲁棒性、透明度、隐私、可靠性、问责制和包容性 [1, 17]。然而，人机交互领域的既有研究指出，尽管当前许多负责任人工智能方法侧重于在有害系统开发或部署之后加以修正，但人们日益达成共识，认为应在开发早期阶段就处理伦理问题 [14]。

虽然负责任人工智能的方法对于构建合乎伦理的人工智能仍至关重要，但考量负责任人工智能之外的人类价值观，能够更全面地了解人工智能对人类繁荣的影响。尤其是在 HCI 和设计中考量宗教价值观与古代智慧，能够为人类经验带来丰富的见解 [12]。

因此，我们提出应在大语言模型开发的早期阶段就纳入伦理设计，而非事后补救，并使其体现宗教价值观。为此，我们采用“美德引导的技术设计”

作者联系信息：Matthew P. Lad，mlad@nd.edu；Louisa Conwill，lconwill@nd.edu；Megan Levis Scheirer，mlevis@nd.edu。

允许免费为个人或课堂用途制作本作品的全部或部分数字或纸质副本，前提是这些副本不得用于或分发以获取利润或商业利益，并且须附上本声明及首页的完整引文。必须尊重本作品中由作者以外的其他方拥有版权的内容。注明出处后允许摘录。若要以其他方式复制、重新出版、发布到服务器或分发至列表，则须事先获得具体许可和/或支付费用。请向 permissions@acm.org 申请许可。© 2026 版权归所有者/作者所有。出版权已授权给 ACM。手稿已提交给 ACM

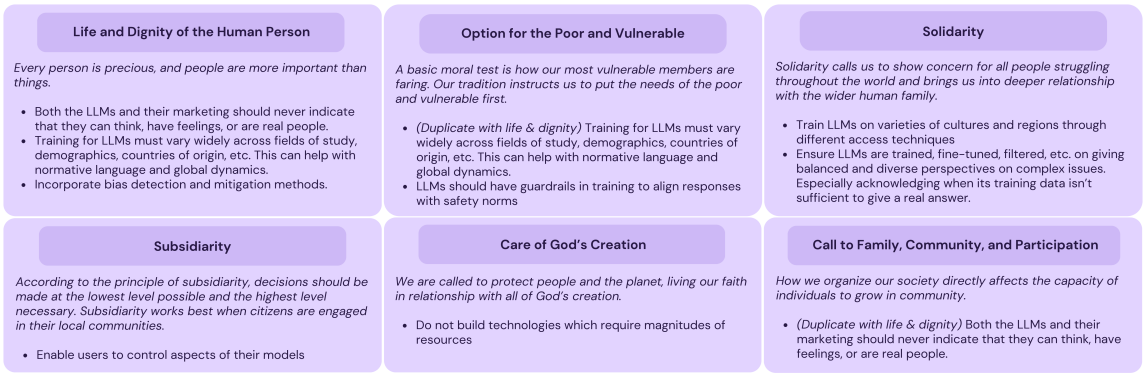


图 1。概念探究总结，详细说明天主教社会训导的六项原则如何应用于大语言模型。为简洁起见，我们仅展示概念探究中最终启发本文五种设计模式的要点。

方法，该方法发表于 CHI 2025 [6]。在本文中，Conwill 等人提出了一种方法，用于生成体现特定宗教或哲学传统美德的伦理设计模式，并以天主教社会训导（CST）的美德为例，为社交媒体提出了七种设计模式。我们以 Conwill 等人的研究为基础，为大语言模型生成并评估体现天主教社会训导美德的设计模式。

通过这项新工作，我们提出五种设计模式，以指导大语言模型的开发，使其更加符合人类繁荣。我们还希望引发关于运用美德伦理学进行大语言模型设计的讨论。

2 方法论

我们遵循 Conwill 等人所概述的三部分方法，即概念、技术和实证方法，用于记录和评估以美德为导向的设计模式。 [6] 第一步是选择所借鉴的美德伦理学传统，以及要构建的技术。为简便起见，我们采用与 Conwill 等人用于其美德传统相同的天主教社会训导原则——人的生命与尊严；家庭、社群与参与的召唤；对贫困者与弱势群体的优先选择；团结互助；辅助性；以及关爱上帝的创造——当然，我们记录的是面向大语言模型的设计模式。

在**概念探究**中，我们考察了所选美德如何适用于大语言模型的情境。对于每种美德，我们先考察其在梵蒂冈官方来源中的定义，再结合各种大语言模型使用案例进行思考，以确定如何将这种美德应用于大语言模型情境。第一作者首先进行了初步思考；随后，其他作者针对初稿提出了自己的想法，力求使最终的概念探究尽可能全面。概念探究的概要见图 1。

在**技术探究**中，我们找出已有的大语言模型，并依据概念探究分析其技术特征，记录支持或违背所选美德的特征。随后，我们根据支持或违背美德的特征模式，归纳出设计模式。第一作者提出 7 个初步模式后，我们与全体作者及一位外部评审进行了多轮讨论，最终选出并进一步完善了 5 个最有可能引发有意义讨论的设计模式。设计模式见图 2。

在这项**实证研究**中，我们开展了一项经机构审查委员会批准的访谈研究，邀请技术领域的研究人员和业界从业者评估这些设计模式。访谈采用半结构化形式，向参与者逐一展示每种设计模式（为避免影响受访者的回答，已从模式中删除其体现的后果和美德），并请他们描述自己认为采用该模式会对大语言模型产生何种积极影响。提交至 ACM 的手稿

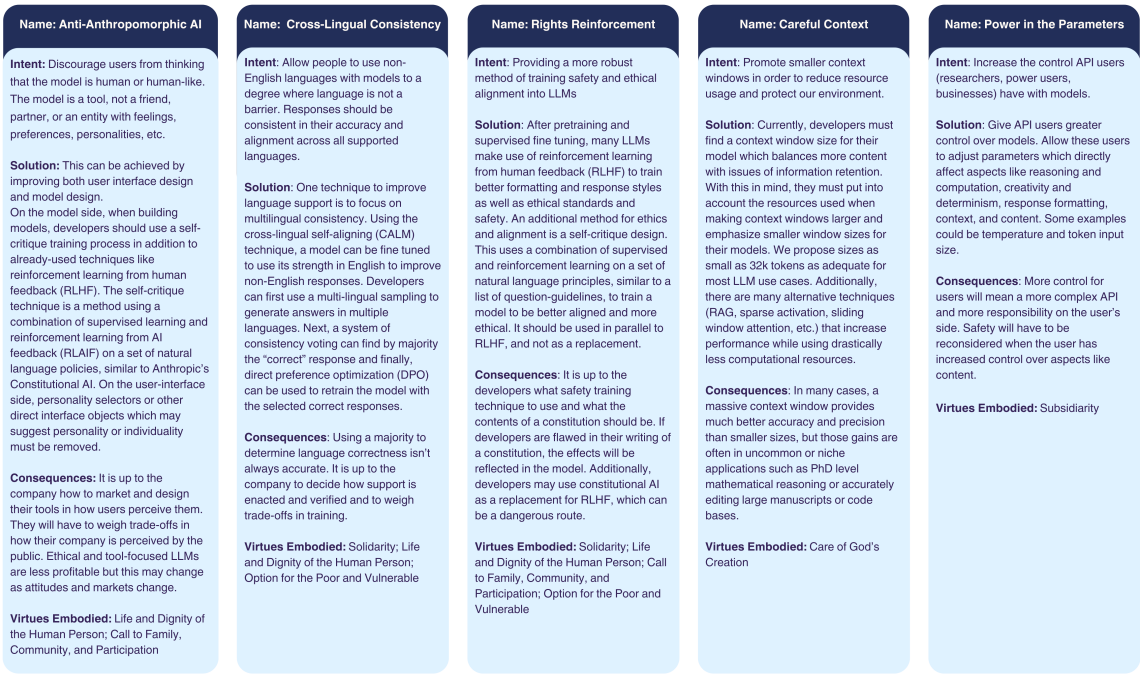


图 2。我们 通过美德引导的技术设计过程记录的面向有美德的大语言模型的五种设计模式 ss.

若采用该模式，会对大语言模型产生何种负面影响；以及他们认为该模式总体上会产生净积极、净消极还是 ambivalent 的影响。完整的访谈协议见附录。

参与者招募与人口统计信息：我们招募了一个由 13 名参与者组成的小型试点小组。与 Conwill 等人的做法类似，参与者须年满 18 岁，并且是具有计算机科学、软件工程、人机交互或相关学科背景的开发者、研究人员、设计师或研究生。这一技术背景是确保参与者充分理解大语言模型并具备更强的设计模式评估能力所必需的。参与者通过口头传播以及相关 Slack 团队等大学沟通渠道招募。参与完全基于志愿原则。完整的参与者人口统计信息见附录。

分析方法：与 Conwill 等人的做法类似，我们使用反思性主题分析 [4] 来分析回答，以归纳方式对参与者关于每种设计模式优缺点的观点进行编码。第一作者和第二作者分别独立对每份访谈记录进行编码。随后，整个研究团队共同讨论、辩论并综合这些编码，寻找价值观和关注点方面的模式，并将编码合并为更高层次的主题。最终，我们针对参与者所重视的模式特征形成了五个主题，针对他们对这些模式的担忧也形成了五个主题。

立场声明： 本文所有作者均为白人、美国人和罗马天主教徒。第一作者为男性，第二和第三作者为女性。

3 结果

我们为每种设计模式的提供方法与 Conwill 等人相同：以及正面回应的数量减去净负面回应的数量，再除以参与者人数，

	Net Positive	Ambivalent	Net Negative	Score
Anti-Anthorpomorphic AI	9	4	0	0.69
Cross-Lingual Consistency	9	3	1	0.62
Rights Reinforcement	12	1	0	0.92
Careful Context	7	5	1	0.46
Power in the Parameters	10	2	1	0.69

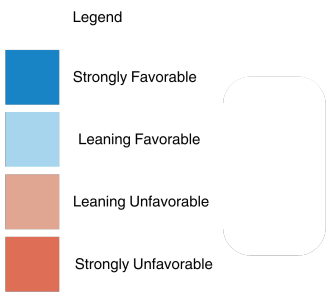


图 3. 将每种模式描述为净正面、净负面或态度矛盾的参与者总数，以及表明该模式总体上被参与者视为强烈有利、偏向有利、偏向不利或强烈不利的得分。所有设计模式均为有利，且除谨慎上下文外，所有模式均为强烈有利。

VALUES	ANTI-ANTHROPOMORPHIC AI	CROSS-LINGUAL CONSISTENCY	RIGHTS REINFORCEMENT	CAREFUL CONTEXT	POWER IN THE PARAMETERS
ROBUSTNESS & ACCURACY (Enables models to do their jobs better)	✓	✓	✓		
SAFETY (Healthy norms and behaviors of users)	✓		✓		
NEW PATHWAYS (Users and developers will develop new pathways of research and use)				✓	✓
IMPROVED ACCESS & CONTROL (Users will be able to tailor LLMs to their own needs)		✓		✓	✓
NOT WASTEFUL (Encourages power and resource management)					

图 4. 参与者在相应拟议设计模式中所看重内容的总体主题（绿色对勾）。

得出 -1 到 1 之间的分数，其中 -1 表示一致反对，1 表示一致赞成。好感度评分见图 3。大多数模式都明显受到好评，而谨慎上下文则倾向于获得好评。

参与者为设计模式确定的价值包括**鲁棒性与准确性**、**安全性**、**新途径**、**改善访问与控制**以及**不浪费**。图 4 显示了每种设计模式提及了哪些价值。

鲁棒性与准确性 强调模型提供更好回答的任何潜力。一位参与者认为，跨语言一致性模式“会……有可能帮助模型总体上……给出正确回答。”另一个被提及的方面是需要客观且情绪色彩较少的回答。P2 将其描述为人类倾向的减少，并说道“…… *[大语言模型]* 有时会对想法过于奉承和积极。我认为去除一些类人倾向，可能会带来……更好的建议。”

安全性 被描述为与大语言模型聊天机器人互动时应遵循的健康行为和规范。这包括聊天机器人不应成为“某个人的替代品”（P8），不应强化可能导致自杀等不健康行为的负面想法（P6），并且应以更透明的方式作出回应，强调信息来自“人工智能模型，而非该领域的专家”（P1）。多位参与者提到了用聊天机器人取代治疗师的危险。

新途径指的是鼓励开辟新的研究方向、改进工作流程，以及强调用户教育的必要性。参与者提到，谨慎上下文模式可以“鼓励开展更多研究并使用替代技术”（P2），并且“可以迫使 *[用户]* 正确界定问题”（P8）。针对较小上下文窗口的培训

CONCERNS	ANTI-ANTHROPOMORPHIC AI	CROSS-LINGUAL CONSISTENCY	RIGHTS REINFORCEMENT	CAREFUL CONTEXT	POWER IN THE PARAMETERS
ONE SIZE FITS NONE (Cultural context matters and translation of words and values are not universal)	✕	✕	✕		
JAILBROKEN (Any safety measure will include loopholes)	✕				✕
MULTIPLE MODES FOR MULTIPLE USES (Users may have domain specific needs, outside of the norms)		✕	✕	✕	
DOING TOO MUCH (Too many post-hoc changes can make tools worse overall)		✕	✕	✕	
IMPLEMENTATION CONCERNS (Technical challenges to implementing virtue-driven design patterns)	✕	✕		✕	

图 5. 参与者关注点的总体主题及相应的拟议设计模式（红色 X）。

甚至可以通过改善大语言模型的环境影响，赋予其更积极的“形象”（P9）。参数中的力量设计模式可以帮助“知道自己在做什么的人”从大语言模型中发挥更大的能力（P1）。

改善访问与控制指的是不同用户能够根据自身目的定制大语言模型。这包括扩展大语言模型支持的语言，以提升可访问性：“如果有人……能够用另一种语言更好地表达自己的想法，那么他们就能更好地使用这些大语言模型”（P9）。谈到定制时，P2 表示，参数中的力量设计模式可以“通过调整温度、输入词元大小等参数，让模型更有效，从而更好地适应特定场景。”

不浪费指设计模式能够节省电力和资源，总体上更环保，同时也能节省时间（例如加快推理速度）。参与者谈到谨慎上下文模式时表示，“它肯定能减少 *[*大语言模型*]*对我们基础设施的消耗……甚至可能让它们……快一些”（P1），尤其考虑到“模型通常需要大量的计算资源”（P7）。P8 还指出，谨慎上下文模式会“迫使 *[*它们*]* 更好地界定问题*[]*”，从而带来更高效的解决方案。

设计模式方面的担忧包括一**刀切并不适用**、**越狱**、**多模式多用途**、**做得太多**以及**实施方面的担忧**。图 5 展示了针对哪些设计模式提到了哪些担忧。

一**刀切并不适用**指的是人们担心设计模式会强化主导文化的语言或文化规范。这包括翻译不准确，或文化规范被固化在回答中的风险：“我认为，考虑本地化很重要，在这种情况下，也要考虑文化差异……如果你想用英语数据来改进非英语回答，你……使用英语，或者……可能会以美国文化作为这些回答的基础……使用 *LLM* 并以自己的语言交流的其他文化中的人，可能会在回答中发现美国文化标记”(P2)。这一主题还涵盖了对伦理多元主义和对齐的担忧。不仅“我的道德观可能和别人的不同”(P8)，而且编码特定的伦理标准也可能迫使系统表现出并非所有用户都认同的行为：“通过设置这些 *[*伦理*]*防护措施，你……会影响人们使用这些系统的方式”(P8)。

越狱指的是这样一种担忧：意图总体上积极的安全措施或设计模式会存在漏洞。参与者对“参数中的权力”表达了担忧，“当给予用户更多控制权时，滥用的可能性就会更大……其中一些参数可能会……让用户更容易越狱模型或规避部分防护措施”（P6）。关于反拟人化人工智能设计模式，参与者担心该模式会因人类不可避免的拟人化而无法按预期发挥作用，或者反过来说，

人们可能比信任人类更信任人工智能系统：“也许强调这一事实——是这个计算系统为你生成了这段输出——反而会让人们更加信任它”（P6）。

多模式多用途指的是一种不允许满足用户特定领域需求的设计模式。一个例子是，谨慎上下文模式在正当使用场景中不允许提供足够大的上下文窗口：“如果你……需要更大的上下文窗口，而……却没有增加窗口的选项，那就是个问题。”（P10）。此外，权利强化模式可能会无意中阻碍正当且合乎道德的使用场景：“如果我现在要问 ChatGPT 一些更黑暗的情况，它会回应说，嘿，想这些事情可能不太好。但如果我是警察调查员，我可能需要在调查工作的背景下了解其中一些更黑暗的情况。”（P5）。最后，跨语言一致性模式可能需要针对关键使用场景进行人工、领域特定的核验：“在医疗或法律环境中使用它……对于非英语语言，……准确且经过人工核验的措辞是强制要求……你需要领域专家或人工翻译人员。”（P12）。

做得过多是指担心这些设计模式会导致准确性或用户体验变差。例如，较小的上下文窗口可能会要求用户“给出更多提示才能完成事情”（P3），也可能在“窗口大小的上限处”引发幻觉（P13）。参与者还担心这些设计模式在准确性或伦理标准方面承诺过高、兑现不足，其中包括“*[*公司*]* 做着他们认为能提升安全性和伦理对齐的事情……结果却给出了不真实的答案”（P2）。

实施方面的担忧是指认为在技术上存在局限，无法按预期方式实施设计模式。参与者指出，技术可行性、计算时间增加、稀释、数据隐私以及意料之外的环境影响，都可能是各种设计模式在实施过程中面临的担忧。

4 讨论与结论

无论是在参与者的价值观还是担忧中，参与者都表达了希望摆脱通用大语言模型，转而使用更小且针对特定用途定制的模型。具体而言，“改善访问与控制”这一价值与可定制性有关，而“不浪费”则与较小模型有关。“一刀切并不适用”和“多模式多用途”这两项担忧强调了为不同使用场景创建专用大语言模型的必要性。综合考虑这些价值观与担忧，可以得出这样的设计启示：针对特定使用场景开发的较小型专用模型，比当今普遍使用的通用模型更受青睐。

大多数参与者认为这些设计模式对大语言模型有积极影响，但他们也经常提到，每种设计模式都存在权衡，并非毫无争议地只有积极作用。参与者还经常表示，即使设计模式中提出的伦理准则并不完美，有一些伦理准则总比完全没有好。综合来看，这表明尽管伦理的实施可能颇具挑战，大语言模型用户仍然重视以某种方式落实伦理原则，即使落实得并不完美。鉴于权利强化是最受青睐的拟议设计模式，下一步研究可尝试实施这一模式，以探明其中的权衡和实施方面的担忧是否确实存在。

将美德引导的技术设计应用于大语言模型，暴露出该方法面临的一些挑战，而 Conwill 等人在生成社交媒体设计时并未遇到这些挑战。首先，大语言模型的设计更加通用，通常表现为聊天机器人。社交媒体平台的设计种类更多，因此更容易生成多样化的设计模式。其次，尽管社交媒体在技术上已经相当成熟，大语言模型问世较晚，目前仍处于算法开发阶段。这使得不少参与者表达了“实施方面的担忧”，不确定拟议设计模式中的算法能否奏效。这凸显了进一步推进大语言模型算法开发的重要性，因为这有助于将伦理理想转化为可行的实现方案。

参考文献

[1] 2025。微软的负责任人工智能。https://www.microsoft.com/en-us/ai/responsible-ai[2] 2025。什么是人工智能幻觉? https://www.ibm.com/think/topics/ai-hallucinations[3] Emily M Bender、Timnit Gebru、Angelina McMillan-Major 和 Shmargaret Shmitchell。2021年。论随机鹦鹉的危险：语言模型会不会过于庞大？载于*2021*年 *ACM*公平、问责与透明度会议论文集。610–623。[4] Virginia Braun 和 Victoria Clarke。2019年。反思反思性主题分析。《运动、锻炼与健康质性研究》11卷4期（2019年），589–597。[5] Ryuhaerang Choi、Taehan Kim、Subin Park、Jennifer G. Kim 和 Sung-Ju Lee。2025年。私密却具社交性：大语言模型聊天机器人如何支持并挑战进食障碍康复。载于*2025*年 *CHI*计算机系统中的人为因素会议论文集（日本横滨）（*CHI '25*）。计算机协会，纽约州纽约市，美国，第642号文章，共19页。doi:10.1145/3706598.3713485[6] Louisa Conwill、Megan K. Levis、Karla Badillo-Urquiola 和 Walter J. Scheirer。2025年。《公益设计模式：运用美德伦理学的智慧打造更好的技术》。载于*2025*年 *CHI*计算机系统中的人为因素会议论文集（日本横滨）（*CHI '25*）。计算机协会，纽约州纽约市，美国，第831号文章，共23页。doi:10.1145/3706598.3713546[7] Rachel Fieldhouse。2023年。人工智能聊天机器人会引发精神病吗？科学研究怎么说。*Afr. J. Ecol* 61卷（2023年），226–227。[8] Ishita Gupta、Ishika Joshi、Adrita Dey 和 Tapan Parikh。2025年。“因为律师都是男性……”：考察大语言模型印地语生成中的隐性性别偏见。载于*2025*年 *ACM*公平、问责与透明度会议论文集（希腊雅典）（*FAccT '25*）。计算机协会，纽约州纽约市，美国，3254–3264。doi:10.1145/3715275.3732208[9] Nanna Inie、Jeanette Falk 和 Raghavendra Selvan。2025年。CHI 的碳排放有多“昂贵”？人机交互研究中生成式人工智能的碳足迹及应对之道。载于*2025*年 *CHI*计算机系统中的人为因素会议论文集（日本横滨）（*CHI '25*）。计算机协会，纽约州纽约市，美国，第206号文章，共29页。doi:10.1145/3706598.3714227[10] Rob Kuznia、Allison Gordon 和 Ed Lavandera。2025年。家属在针对 OpenAI 的诉讼中称，CHATGPT 怂恿一名大学毕业生自杀。https://www.cnn.com/2025/11/06/us/openai-chatgpt-suicide-lawsuit-invs-vis[11] Bernard Marr。2023年。ChatGPT 简史：我们如何走到今天。https://www.forbes.com/sites/bernardmarr/2023/05/19/a- short-history-of-chatgpt-how-we-got-to-where-we-are-today/[12] Khushnood Naqshbandi、Kristina Mah 和 Naseem Ahmadpour。2022年。为信仰、宗教与灵性在人机交互中的亲社会实践留出空间。《*Interactions*》29卷4期（2022年），62–67。[13] Patrick L. Plaisance。2025年。人工智能去技能化：使用聊天机器人会侵蚀我们的人性吗？https://www.psychologytoday.com/us/blog/virtue-in-the-media-world/202512/ai-de-skilling-will-chatbot-use-corrode-our-humanness[14] Devansh Saxena、Ji-Youn Jung、Jodi Forlizzi、Kenneth Holstein 和 John Zimmerman。2025年。人工智能错配：在人工智能开发前识别潜在的算法危害。载于*2025*年 *CHI*计算机系统中的人为因素会议论文集。1–23。[15] Olivia Shone。2025年。大型语言模型的五大关键特性与优势。https://www.microsoft.com/en-us/microsoft-cloud/blog/2024/10/09/5- key-features-and-benefits-of-large-language-models/[16] Emma Strubell、Ananya Ganesh 和 Andrew McCallum。2019年。自然语言处理深度学习的能源与政策考量。载于计算语言学协会第*57*届年会论文集。3645–3650。[17] Cole Stryker。2025年。什么是负责任人工智能? https://www.ibm.com/think/topics/responsible-ai

《让聊天机器人拥有美德是否可行？将基于美德的设计方法论应用于大语言模型的探究》附录

MATTHEW P. LAD、LOUISA CONWILL 和 MEGAN LEVIS SCHEIRER

ACM 引用格式：
Matthew P. Lad、Louisa Conwill 和 Megan Levis Scheirer。2026。《让聊天机器人拥有美德是否可行？将基于美德的设计方法论应用于大语言模型的探究》附录。1，1（2026年2月），2页。https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 参与者人口统计信息

表1详细描述了参与者的人口统计信息。

2 访谈协议

我们依据这组基础问题进行了半结构化访谈。如果参与者不太了解什么是设计模式，我们会向他们读出以下内容：

在软件工程中，设计模式是一种分享设计知识的工具，尤其用于分享解决常见问题的设计方案。设计模式概述了常见设计问题、其背景以及该模式提出的解决方案。本次访谈中，我们将向您介绍为大型语言模型（LLM）创建的各种设计模式。这些模式旨在引导大型语言模型的开发，使其有益于人类，并避免当今人工智能（AI）所带来的诸多隐患。

2.1 访谈问题

在向参与者展示设计模式之前，我们向每位参与者提出了以下问题。

预先提问（基本人口统计信息）：以下问题均可选择是否回答 e

- 您的性别是什么?
- 您如何描述自己的种族/族裔?
- 您的宗教背景是什么?

问题 1： 您的教育和职业背景是什么，尤其是在技术方面？**问题 2：** 在日常工作中，您会做出或曾经做过设计决策吗？（例如，在软件层面或用户体验层面）

问题 3： 您通常如何做出设计决策?
问题 4： 做出设计决策时，您会优先考虑什么?
问题 5： 您使用设计模式的经验如何?
问题 6： 根据您的设计模式使用经验，您觉得它们有哪些帮助（如果有的话）？

Auth作者的联系信息：Matthew P. Lad，mlad@nd.edu；Louisa Conwill，lconwill@nd.edu；Megan Levis Scheirer，mlevis@nd.edu。

允许出于个人或课堂用途免费制作本作品的数字或纸质副本的全部或部分，但前提是不得为营利或商业利益而制作或分发副本，并且副本须包含本声明及首页上的完整引文。必须尊重本作品中由作者以外的其他方拥有版权的组成部分。注明出处后，允许摘录。若要以其他方式复制、重新发布、发布到服务器或分发至列表，则须事先获得特定许可和/或支付费用。请通过 permissions@acm.org 申请许可。© 2026 版权归所有者/作者所有。出版权已授权给 ACM。手稿已提交至 ACM

表 1. **参与者人口统计信息**

ID#	性别	种族/族裔	宗教信仰	教育经历-ence	技术工作经历-验
P1	Male	未回答	天主教徒	计算机科学理学学士	网络安全 实习期、信息技术工作、医疗设备维修
P2	女性	白人	天主教徒	计算机科学理学学士，目前攻读博士学位计算机科学专业学生	软件工程实习
P3	Male	白人	天主教徒	计算机科学理学学士，目前攻读博士学位计算机科学专业学生，目前神学硕士生	软件工程实习
P4	Male	白人	天主教徒	计算机科学理学学士，现攻读理学硕士计算机科学专业学生	未报告
P5	Male	白人	新教徒（长老会）	计算机科学理学学士	数据科学与分析sis
P6	Male	白人	天主教徒	目前就读计算机科学学士/硕士博士生	软件工程
P7	Male	西班牙裔/拉丁裔	天主教徒	计算机科学学士和博士	博士后研究计算机科学
P8	Male	西班牙裔/拉丁裔	天主教徒	计算机科学理学学士，目前攻读博士学位计算机科学专业学生	软件工程
P9	Male	白人	天主教徒	目前就读于CS	软件工程实习
P10	Male	白人	归正宗（新教）	计算机科学理学学士，目前攻读博士学位计算机科学专业学生	本科生 re-在计算机科学中搜索
P11	Male	白人	天主教徒	计算机科学理学学士	IT支持
P12	Male	南亚裔	天主教徒	计算机科学理学学士和理学硕士，现-任计算机科学博士生	机器学习实习
P13	Male	白人	天主教徒	理学学士，主修 数学， 理学硕士教学、目前攻读博士学位工程学专业博士生	高中数学教学数学、计算机科学和机器人学

参与者看过每种设计模式后，都会被问及以下四个问题。如有必要，还会被问及后续问题。

问题 7：关于该设计模式的运作方式，您有什么需要澄清的问题吗（如果有）？**问题 8：**您认为该设计模式在哪些方面（如果有）能够改进大型语言模型？**问题 9：**您认为该设计模式在哪些方面（如果有）无法改进大型语言模型？**问题10：**您认为这种模式对大型语言模型的总体影响会是更积极、更消极，还是喜忧参半？