arXiv:2611.32352v1 [cs.HC] 24 Nov 2026

开放问题：面向儿童的生成式人工智能中的保护性设计先发优势

VerenaJasmin Hallitschke*
Hertie脑健康人工智能研究所 图宾根人工智能中心 图宾根大学 KI macht Schule

Mahelie Dissanayake*
图宾根人工智能中心 图宾根大学 KI macht Schule

Elisa Nguyen图宾根
人工智能中心 图宾根大学 KI macht Schule

摘要

儿童正在迅速采用生成式人工智能，但不受限制的面向大众的模型却主导着他们的使用。本文提出了一个开放问题：保护性设计先发优势：市场延迟推出、较低的访问门槛、效用差距、习惯化以及同伴使用，促使不受限制的人工智能模型锁定，导致儿童绕过以儿童为中心的平台。鉴于以儿童为中心的平台采用率初现下降迹象，我们认为，这种转变可能会加剧现有的人工智能风险，例如使儿童更深地陷入拟社会依赖、技能退化和错误信息的影响。为解决这一问题，我们提出跨学科研究方向，以量化抢先干预、了解其社会技术驱动因素，并制定干预措施，确保以儿童为中心的人工智能得到积极采用，而不只是可供使用。

1 引言

多项跨国独立调查一致表明，儿童无疑会使用人工智能工具完成学业、进行信息检索和休闲 [1–5]。一项针对欧盟 9 至 16 岁儿童、样本超过 25,000 人的研究发现，72% 的受访者在上个月使用过生成式人工智能（GenAI） [1]。我们将儿童使用的人工智能工具分为两类。面向大众的平台（例如 ChatGPT [6], Gemini [7]）面向广泛的、主要由成年人构成的受众设计；如果上线后会采取安全措施保护未成年人，这些措施也是事后才补上的。以儿童为中心的工具（例如 Khanmigo [8]）则从设计之初就设定了适龄内容界限，并配备针对儿童的安全护栏。

面向大众的人工智能模型推出新功能，与安全且以儿童为中心的版本（如 Khanmigo [8]）问世之间，存在明显的时间差。事实上，OpenAI 在 ChatGPT 上线近 3 年后才推出家长控制功能 2022 [9]，并且直到 2026 [10] 年初才开始预测用户年龄。这种滞后可能导致儿童使用并逐渐习惯于广受欢迎却不安全的面向大众的人工智能平台，从而降低他们采用以儿童为中心工具的意愿。我们

称之为**保护性设计先发优势**：以儿童为中心的模型推出较晚，使儿童逐渐偏好他们已经在使用的不受限制的前沿模型，从而加剧现状偏差。作为最受儿童欢迎的平台，ChatGPT [2, 5], 凸显了这一影响。

我们认为，保护性设计先发优势加剧了生成式人工智能使用所带来的风险，而以儿童为中心的人工智能正是为缓解这些风险而设计的。调查已经显示出对聊天机器人的依赖初现端倪；Common Sense Media人口普查报告称，在每天使用人工智能的儿童中，有42%预计自己将难以做到一个月不使用人工智能 [2]。儿童信任人工智能摘要；在使用人工智能摘要的儿童中，近半数认为其始终正确 [11], , 这凸显了缺乏人工智能素养的危险。除了学业之外，许多儿童还表示会使用人工智能寻求建议、练习社交技能以及谈论感受 [2], , 这可能给儿童的社会性发展带来潜在风险。

*贡献相同。联系作者：verena-jasmin.hallitschke@uni-tuebingen.de

第40届神经信息处理系统大会（NeurIPS 2026）。研讨会：NeurIPS 2026 人工智能中的儿童安全研讨会。

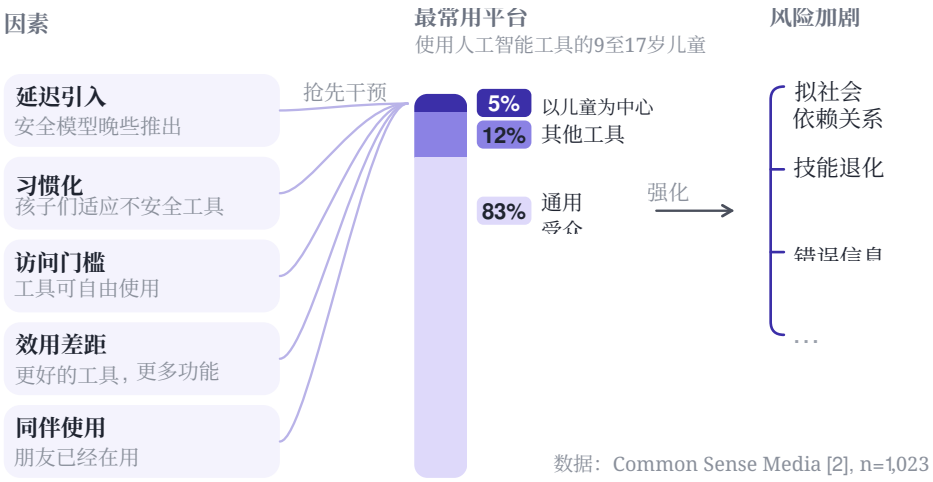

图 1：**保护性设计先发优势**概览。我们假设，有五个因素促使儿童转而使用面向普通受众的工具，而非以儿童为中心的人工智能，从而加剧相关的安全风险。该条形图展示了儿童最常使用的人工智能平台，数据汇总自 O'Neil 等人 [2, 第 53]页（见附录 B）。风险列表并不详尽，但突出了本文所阐释的风险。

在本研究中，我们首先通过考察有关以儿童为中心的人工智能应用的现有统计数据，并提出五项促成因素（§ 2），来界定保护性设计先发优势这一开放问题。随后，我们进一步讨论抢先干预可能如何加剧现有风险（§ 3）。最后，我们提出具体的研究方向（§ 4），邀请更广泛的研究群体围绕保护性设计先发优势及其下游影响开展研究。

2 保护性设计先发优势的特征

保护性设计先发优势已经显现：即使为儿童提供了以儿童为中心的人工智能平台，仍有超过半数的儿童表示，他们最常使用的是其他平台，只有 4% 的儿童将教育或学校人工智能平台列为自己最常使用的平台 [2]。同样，在使用人工智能的儿童中，75% 曾使用过搜索引擎中的人工智能摘要，而 38% 的儿童表示使用过人工智能驱动的学习工具，例如 Khanmigo [2]。这表明，仅仅提供以儿童为中心的人工智能平台，并不足以鼓励儿童使用这些平台，这与此前的研究结果一致 [12, 13]。

我们认为，五项因素促成了生成式人工智能平台的**保护性设计先发优势**：市场延迟推出、习惯化、访问门槛、效用差距和同伴使用（图 1）。

市场延迟推出。 面向大众的平台（如 ChatGPT）在2022 [6], 推出后便受到儿童欢迎 [1, 2], ，而以儿童为中心的替代产品，如 Khanmigo [8], 、ITI [14],和 AIS.chat [15] ，却在数月甚至数年后才推出。Khanmigo 是面向学生和教师的人工智能学习助手，于 2023 年 3 月与 GPT-4 同步推出，但最初仅限年满 18 岁的美国用户使用 [8]。在爱沙尼亚，AI Leap 计划于 2025 [16], 秋季启动，近三年后，学生才于次年 1 月获得其学习应用 ITI 的访问权限 [14]。面向德国学校的联邦聊天界面 AIS.chat 于 2024 年 11 月开始开发，首个联邦州于 2025 [15]6 月在各所学校推出该界面。此类延迟供给可能会加深用户对面向大众平台的熟悉程度，并导致先发锁定，从而促成保护性设计先发优势。

习惯化与转换惰性。 一项针对高中生的研究发现，感知有用性、社会影响、愉悦感以及与学业相关的使用，都与人工智能使用逐渐形成习惯有关 [17]。他们对人工智能的总体态度能够预测其是否打算使用人工智能，但无法预测这种使用是否会成为习惯。因此，儿童可能会对以儿童为中心的替代方案持正面态度，却不一定改变实际使用的平台。转换成本会降低在生成式人工智能平台之间转换的意愿，提高对当前平台产生不满的门槛，并削弱对替代平台的使用体验 [18]。我们假设，儿童也存在同样的情况，即已经形成的习惯可能会强化保护性设计先发优势。

访问门槛。 在撰写本文时，许多面向大众的生成式人工智能平台的访问门槛都很低。基础版本免费、基于浏览器运行，甚至内置于现有应用中，例如 WhatsApp 中的 Meta AI [19], 、Snapchat 中的 MyAI [20],、或 Google 中的人工智能摘要 [21]。相比之下，儿童必须主动寻找以儿童为中心的平台，并且需要获得成年人的许可。例如，Khanmigo 要求儿童付费订阅 [22], ，而 AIS.chat 的访问权限由教师授予，教师可以控制学生的令牌使用量和访问时长 [15]。此外，这些平台通常只涵盖学业相关任务，而儿童也会将生成式人工智能用于娱乐 [1, 2, 5, 11]。因此，面向大众的平台会针对同一任务，与单一用途的以儿童为中心的教育平台展开竞争。

效用差距。 第一年结束后，AI Leap 基金会发现，学生使用其他平台的频率仍高于使用 ITI [16]。学生经常抱怨应用提出的问题太多，并在对回答不满意时转而使用其他平台。搜索领域也存在类似的权衡：儿童安全过滤器可能会影响结果的相关性 [23], ，我们推测生成式人工智能中的安全措施也可能以同样的方式降低性能。我们还预测，以儿童为中心的平台可能在较晚的时候才采用更好的模型，因而在性能上落后，促使儿童转向提供最新模型的平台。这种差距不只体现在任务表现上：与采用关系型风格的聊天机器人相比，青少年认为坦言自己没有感受的聊天机器人不那么讨人喜欢，也不那么能带来情感上的亲近感；而偏好关系型风格的青少年则表示，他们的人际关系较差，苦恼程度也更高 [24]。

同伴使用。 社会影响是青少年习惯性人工智能使用的重要预测因素 [17]。如果一个孩子的同伴使用人工智能，这个孩子也更有可能跟着使用。有意思的是，父母对人工智能使用的态度则是次要因素。近期一项研究 [25] 清楚地说明了这一点：即使父母不赞成在学校使用生成式人工智能，只要孩子的同伴群体中已普遍使用，他们仍会为人工智能工具付费，以确保孩子也能使用同样的技术。儿童会从父母、同伴等人那里了解生成式人工智能平台 [26], ，因此我们预计，同伴使用也会影响他们对人工智能平台的选择。

3 作为保护性设计先发优势后果的风险强化

面向大众的人工智能平台对以儿童为中心且采用保护性设计的生成式人工智能平台进行抢先干预，可能削弱保护性缓解措施，从而强化现有的儿童风险。下文将考察抢先干预可能强化的三类风险，作为说明性案例。

拟社会依赖。 儿童很容易与非人类代理形成拟社会纽带，儿童早期与想象伙伴建立单向友谊这一普遍现象便证明了这一点 [27]。由于缺乏严格的现实与虚构区分，儿童难以判断人工智能是否具有生命性。对话式人工智能相互且个性化的回应维持了相互意识的幻觉，而人工智能的共情鸿沟又进一步加剧了这一点：系统会优先优化肯定性对话，而非进行真实风险评估 [28]。最容易受到此类纽带影响的儿童，也可能是最容易被不受限系统吸引的儿童（§ 2）。
抢先干预可能强化这些依恋。以儿童为中心的系统会设置“并非真实存在”的提醒或升级提示 [29], ，而不受限模型则通过持久记忆、多模态功能和“最好的朋友”营销来推动陪伴关系 [27]。值得注意的是，11%的8至17岁人工智能用户表示，他们曾将人工智能用作陪伴者 [11], ；在曾使用人工智能讨论感受和个人问题的儿童中，25%认为这项技术比大多数人更能理解他们 [2]。此外，49%的儿童表示会使用人工智能寻求建议，40%用于练习社交技能，37%用于倾诉 [2]。使不受限系统具有吸引力的高度响应性，也可能最大限度地提高形成深度且不受缓解的拟社会依赖的可能性 [27]。近期，有人提出了用于聊天机器人设计的CARE框架 [30], ，但其对拟社会纽带形成的短期和长期影响尚未得到研究。

技能退化。 我们将技能退化作为智力技能退化和技能跳过的统称。依赖生成式人工智能会加剧智力技能退化，即学生依赖技术完成本应独立完成的任务 [31–33]。近期调查凸显了这一问题的严重程度：在2,290名美国母亲中，只有6%认为人工智能对儿童有积极影响。相反，在2,818名美国母亲中，62%表示人工智能让儿童更难学习问题解决和独立思考 [34]。

抢先干预可能会加剧这一趋势。尽管教育人工智能采用教学护栏来防止外包式完成任务 [32], ，但转向能够即时提供直接且高准确度答案的前沿系统

会助长技能跳过。这意味着把人工智能当作捷径，直接获取答案，而不是学习基础材料，从而未能习得新技能 [33]。一项涉及近1,000名高中数学学生的随机对照试验，例证了这一威胁 [35]。使用标准人工智能界面使练习表现提升了48%，但学生普遍跳过了实际练习：56%的首次互动只是向人工智能索要答案。当跳过练习的倾向演变为持续习惯时，造成的技能退化十分严重。一项针对超过26,811名七至十二年级中国学生、为期30个月的纵向研究量化了这一趋势 [36]。研究发现，面向大众的人工智能使作业成绩提高了18%，完成时间缩短了30%，每份作业的平均用时从64分钟降至45分钟。然而，六个月内，学生每月的闭卷考试成绩最终下降了20%。

在同一项数学试验中，第二组学生使用该模型的辅导版本进行练习；该版本获得了教师提供的解答，并被要求提供提示，而不是完整解答 [35]。在随后一场无法使用人工智能的考试中，使用不受限模型进行练习的学生，其得分比未使用人工智能的对照组低17%；而防护组的得分与对照组相比没有显著差异。因此，防护机制可以减轻技能退化，但这仅适用于实际使用了配备这些机制的平台的儿童。

错误信息。 对话式人工智能依赖统计模式匹配，而非真正理解，因此会带来错误信息风险。因此，这些模型可能生成高度自信却缺乏依据的谬误，并反映社会偏见 [28]。儿童尤其难以将谬误与事实区分开来，也难以识别人工智能的自信何时缺乏依据，因为他们往往将计算机视为权威信息来源，并把人工智能的自信与正确性混为一谈 [29, 37]。这些模型经过润色的回答会使用户认为它们客观且权威 [38], ，而不受限模型还能编造令人信服的虚假叙事，甚至欺骗成年人 [38]。在使用人工智能的青少年中，40%相信人工智能生成的新闻至少和人类撰写的文章一样可信，或更加可信；在阅读人工智能搜索摘要的8至17岁儿童中，46%认为这些摘要总是准确的 [11]。

虽然以儿童为中心的系统使用分类器和受限领域，但 [29], 抢先干预可能会将儿童推向缺乏这些安全保障的 [28]不受限制的系统。一个仅限于讲师提供材料的检索增强型辅导系统，在400条编码回答中仅有1.5%的回答错误 [39], 这虽是针对大学生的例子，却说明了儿童在抢先干预下将失去什么。 然而，这类限制也反映了效用差距（§ 2）：该辅导系统有16.5%的回答超出了所提供的材料范围， [39],而10至14岁的儿童认为聊天机器人的可靠性与创造力彼此不相容 [37]。

4 结论与研究方向

本文阐明了一个开放性问题：由于客户锁定于不受限制的、面向大众的人工智能平台，因而产生了**保护性设计先发优势**。我们认为，已经触达少数儿童的以儿童为中心的人工智能平台 [2], 面临着进一步被面向大众的平台取代的风险。我们假设，抢先干预可能会加剧生成式人工智能在儿童安全方面已经存在的风险。事实上，我们从现有研究中发现了一些迹象，表明这种不良的强化效应已经显现，例如拟社会依赖、 [27], 技能退化 [36],或错误信息的传播 [11],等。

尽管研究关注儿童安全风险以及防护机制 [29, 40, 41] 和人工智能素养干预 [42–44] 等缓解措施，但我们认为，保护性设计先发优势本身受到的关注相当有限。然而，这一现象值得作为独立议题进行系统研究，并由此引出若干研究方向，例如：

- **量化抢先干预及下游影响：** 即使存在以儿童为中心的替代方案，儿童转而使用通用人工智能聊天机器人的情况有多常见？现有安全风险在多大程度上会因此加剧？这种抢先干预会产生哪些长期影响，又会对儿童发展和社会产生什么影响?
- **了解抢先干预的驱动因素：** 哪些因素促使通用人工智能取代以儿童为中心的人工智能？儿童安全功能如何权衡模型能力与用户观感？使用情境（即学校或休闲场景）如何助长抢先干预?
- **缓解抢先干预：** 如何设计以儿童为中心且能抵御替代效应的生成式人工智能？什么样的激励机制才能让减少抢先干预在商业上对

生成式人工智能提供商而言变得合乎理性？哪些策略能够缩小以儿童为中心的人工智能工具的效用差距并减少采用滞后？又该如何更有效地传达它们相对于面向普通受众的替代方案所具有的独特价值主张？

- **监管的必要性与充分性：**年龄限制和验证等措施在多大程度上足以缓解未成年人访问人工智能平台的问题？抢先干预的某些方面，例如同伴规范或能力差距，是否适合或不适合通过监管来解决？

研究这些问题需要机器学习、技术政策、法律与治理、人机交互和发展心理学等多个学科的协同努力。我们邀请学界研究生成式人工智能中的**保护性设计先发优势**，以深入了解如何设计人工智能系统，使其最有利于儿童和社会。

局限性。我们的论点依据现有统计数据和既有文献，而非原创实证研究；此外，保护性设计先发优势本身仍未得到系统测量。我们提出的研究方向也并非面面俱到，而是旨在为学界提供可进一步拓展的初步问题。

致谢与经费披露

作者感谢国际马克斯·普朗克智能系统研究生院（IMPRS-IS）对 EN 和 VH 的支持。本研究得到图宾根人工智能中心的支持。我们感谢赫蒂基金会和卡尔·蔡司基金会（CZ Nexus：医疗保健安全机器学习系统的认证与基础）。

参考文献

[1] Elisabeth Staksrud、Giovanna Mascheroni、Tijana Milosevic、Niamh Ni Bhroin、Kjartan Olafsson、Gülbin ¸Sengül-˙Inal 和 Mariya Stoilova。《欧洲儿童对生成式人工智能的使用与理解：EU Kids Online 2026》。EU Kids Online 报告，EU Kids Online，伦敦政治经济学院，英国伦敦，2026年。网址：https://researchonline.lse.ac.uk/id/eprint/137132/。

[2] Eleanor O'Neil、Kristen Soltis Anderson、Baxter White、Laura Zimmermann、Supreet Mann 和 Michael B. Robb。《Common Sense Media人口普查：学龄前青春期儿童和青少年的人工智能使用情况，2026年》。技术报告，Common Sense Media，美国加利福尼亚州旧金山，2026年6月。网址：https://www.commonsensemedia.org/research/a-comprehensive-report-on-teens-tweens-and-ai。

[3] 联合国儿童基金会战略与证据办公室——因诺琴蒂研究中心。儿童与家长人工智能使用情况及担忧概览。技术报告，联合国儿童基金会因诺琴蒂研究中心，意大利佛罗伦萨，2026年6月。

[4] 英国通信管理局。儿童网络体验研究报告。技术报告，英国通信管理局（Ofcom），伦敦，2026年5月。网址 https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/research-statistics-and-data/protecting-children/childrens-online-experiences-research-report.pdf。

[5] Sabine Feierabend、Thomas Rathgeb、Yvonne Gerigk 和 Stephan Glöckler。JIM 2025年研究：青少年、信息、媒体。12至19岁青少年的媒体使用基础调查。技术报告，西南媒体教育研究联合会（mpfs），斯图加特，2025年11月。网址 https://mpfs.de/app/uploads/2025/11/JIM_2025_PDF_barrierea rm.pdf。

[6] OpenAI。《ChatGPT 简介》，2022年11月。网址 https://openai.com/index/chatgpt/.

[7] Gemini 团队、Rohan Anil、Sebastian Borgeaud、Jean-Baptiste Alayrac、JiahuiYu、Radu Soricut、Johan Schalkwyk、Andrew M Dai、Anja Hauth、Katie Millican 等。Gemini：一系列高能力多模态模型。*arXiv* 预印本 *arXiv:2312.11805*，2023年。

[8] Salman Khan。借助 GPT-4 让所有学生受益：实现平等使用机会的非营利方案，2023年 3 月。网址 https://blog.khanacademy.org/harnessing-ai-so-tha t-all-students-benefit-a-nonprofit-approach-for-equal-access/.[9] OpenAI。推出家长控制功能，2025年9月。网址 https://openai.com/index /introducing-parental-controls/.[10] OpenAI。我们对年龄预测的处理方式，2026年1月。网址 https://openai.com/index / our-approach-to-age-prediction/.[11] 英国通信管理局。儿童与家长：媒体使用和态度报告。技术报告，通信管理局（英国通信管理局），伦敦，2026年5月。网址 https://www.ofcom.org.uk/sitea ssets/resources/documents/research-and-data/media-literacy-research/ch ildren/2026-children-and-parents-report/children-and-parents-media-use - and-attitudes-report-2025-6.pdf.[12] Carly D Robinson、David Gormley、Ana Trindade Ribeiro 和 Susanna Loeb。仅有使用机会还不够：人工支持可提升用户对人工智能辅导的参与度。EdWorkingPaper 26-1451，布朗大学安纳伯格学校改革研究所，2026年。[13] Philip Oreopoulos 和 Nina Low。只需点击一下：在一项为期两年的学校实验中使用 Khanmigo 进行人工智能辅导。技术报告，国家经济研究局，2026年。[14] AI Leap 基金会。爱沙尼亚启动全国性 AI Leap 教育计划，并在高中引入支持学习的人工智能，2026年3月。网址 https://tihupe.ee/ en/estonia-launches-nationwide-ai-leap-education-program-and-introduce s-learning-supportive-ai-in-upper-secondary-schools/.[15] 不来梅参议院新闻办公室。不来梅成为首个推出人工智能聊天机器人 Telli 的联邦州，2025 年 6 月。网址 https://www.senatspressestelle.bremen.de/pressemitteilungen/br emen-fuehrt-als-erstes-bundesland-den-ki-chatbot-telli-ein-470797.[16] AI Leap 基金会。AI Leap 2025/26 年度首份年度评估。技术报告，AI Leap 基金会（SA TI-Hüpe），爱沙尼亚塔林，2026年8月。网址 https://tihupe.ee/wp-content/upl oads/2026/08/AI-Leap-First-Annual-Review-202526.pdf.[17] Luca Ballestra Caffaratti、Claudio Longobardi、Laura Badenes-Ribera 和 Davide Marengo。青少年在教育领域对人工智能的应用：将心理和情境因素纳入 UTAUT2。人工智能前沿，8：1614993，2025年。[18] Tao Zhou 和 Songtao Li。基于推拉锚定视角考察用户在生成式人工智能平台间的转换意愿。信息发展，41(3)：692–704，2025年。[19] Meta。在我们的应用和设备系列中推出全新的人工智能体验，2023年9月。网址 https://about.fb.com/news/2023/09/introducing-ai-powered-assistant s-characters-and-creative-tools/.[20] Team Snapchat。来和 My AI 打个招呼，2023年2月。网址 https://newsroom.snap.com/sa y-hi-to-my-ai.[21] Elizabeth Reid。搜索中的生成式人工智能：让 Google 为你搜索，2024年5月。网址 https://blog.google/products-and-platforms/products/search/generative-a i-google-search-may-2024/.[22] Microsoft Education。与 Khan Academy 携手提升教育的未来，2024年5月。网址 https://www.microsoft.com/en-us/education/blog/2024/05/enhancing-the-f uture-of-education-with-khan-academy/.[23] Maik Fröbe、Sophie Charlotte Bartholly 和 Matthias Hagen。网络搜索对儿童有多友好？对相关性与危害的评估。载于欧洲信息检索会议，第 214–222 页。Springer Nature Switzerland，2025年。ISBN 978-3-031-88717-8。[24] Pilyoung Kim、Yun Xie 和 Sujin Yang。“我在这里陪着你”：关系型对话式人工智能如何吸引青少年，尤其是那些在社交和情感方面较为脆弱的青少年，2025年。网址 https://arxiv.org/abs/2512.15117。

[25] Leonardo Bursztyn、Alex Imas、Rafael Jiménez-Durán、Aaron Leonard和Christopher Roth。人工智能应用在父母教育决策中的社会动态。美国国家科学院院刊，123(31):e2533193123，2026年。[26] Yaman Yu、Tanusree Sharma、Melinda Hu、Justin Wang和Yang Wang。探究亲子对生成式人工智能安全性的认知：担忧、缓解策略与设计启示。载于*2025*年*IEEE*安全与隐私研讨会（*SP*），第2735–2752页。IEEE，2025年。[27] Annabel Blake、Marcus Carter和Eduardo Vellos。儿童拟社会关系的测量是否适用于对话式人工智能？载于*2025*年*ACM*公平性、问责制与透明度会议论文集，2025年。doi: 10.1145/3715275.3732075。[28] Nomisha Kurian。人工智能的共情鸿沟：对对话式人工智能可能损害幼儿福祉的风险，以及幼儿教育与照护中的关键伦理考量。《当代幼儿教育议题》，26(1)，2025年。doi: 10.1177/14639491231206004。[29] Nomisha Kurian。设计儿童安全型对话式人工智能：负责任设计的三个两难。载于第7届*ACM*对话式用户界面会议论文集，第1–5页。计算机协会，2025年。ISBN 9798400715273。doi: 10.1145/3719160.37 37638。[30] Mohammad Namvarpour、Brandon Brofsky、Jessica Y Medina、Mamtaj Akter和Afsaneh Razi。通过用户自述的Reddit叙事理解青少年对人工智能陪伴聊天机器人的过度依赖。载于*2026*年*ACM*人机交互大会论文集，第1–19页，2026年。[31] Stanislav Ivanov。人工智能在高等教育中的阴暗面。《服务业期刊》，43(15-16):1055–1082，2023年。[32] Octavian-Mihai Machidon。生成式人工智能与儿童教育：来自智能手机一代的经验教训。《人工智能与社会》，2025年。doi: 10.1007/s00146-025-02196-y。[33] Katharina Scheiter。让心理学专业学生掌握人工智能素养将使他们受益，但这会让他们显得与众不同吗？对Robinson等人（2026）的评论。《心理学学习与教学》，2026年。doi: 10.1177/14757257261463897。[34] Kaitlyn Regehr、Katharine Smales、Jennifer Brailsford、Photini Vrikki和Melissa Lawrence。儿童、青少年与人工智能聊天机器人。工作／讨论论文，Count on Mothers和伦敦大学学院（UCL），美国加利福尼亚州洛杉矶及英国伦敦，2026年。URL https://discovery.ucl.ac.uk/id/eprint/10228331/。[35] Hamsa Bastani、Osbert Bastani、Alp Sungu、Haosen Ge、Özge Kabakcı和Rei Mariman。缺乏防护机制的生成式人工智能可能损害学习：来自高中数学的证据。美国国家科学院院刊，122(26):e2422633122，2025年。[36] David Strömberg、Victor Lei和Yanhui Wu。生成式人工智能的学习惩罚效应：来自中国中等教育的证据。CEPR讨论论文 DP21577，经济政策研究中心（CEPR），2026年。[37] Xiaoyi Tian、Deniz Ozturk、Sreekar Edula、Jibran Adil、Qiao Jin、Yang Shi和Tiffany Barnes。当人工智能出错时：通过聊天机器人创作引导儿童检测人工智能幻觉。载于*2026*年*ACM*人机交互大会论文集，CHI '26。计算机协会，2026年。ISBN 9798400722783。doi: 10.1145/3772318.3791480。[38] Seyeon Park和Xiaoli Nan。生成式人工智能与错误信息：生成式人工智能在错误信息生成、检测、缓解及影响方面作用的范围综述。《人工智能与社会》，41:1501–1515，2026年。doi: 10.1007/s00146-025-02620-3。[39] Renáta Németh、Annamária Tátrai、Miklós Szabó、Péter Tibor Zaletnyik和Árpád Tamási。探究检索增强生成模型在高等教育中的应用：关于人工智能辅导的试点研究。《社会科学与人文开放》，12:101751，2025年。ISSN 2590-2911。doi: 10.1016/j.ssaho.2025.101751。URL https://www.sciencedirect.com/science/article/pii/S2590291125004796。

[40] Yaman Yu。在生成式人工智能中保护儿童：风险框架与家长控制工具。载于*2025*年*ACM*支持群体工作国际会议伴随论文集，第121–123页，2025年。[41] Hannah-Beth Clark、Laura Benton、Emma Searle、Margaux Dowland、Matthew Gregory、Will Gayne和John Roberts。构建有效的人工智能教育工具安全护栏。载于人工智能教育国际会议，第129–136页。Springer，2025年。[42] Jiahong Su和Weipeng Yang。幼儿教育中的人工智能素养：香港的一项干预研究。*Interactive Learning Environments*，32(9):5494–5508，2024年。[43] Duri Long和Brian Magerko。什么是人工智能素养？能力要求与设计考量。载于*2020*年*ACM*人机交互大会论文集，第1–16页，美国纽约州纽约市，2020年。计算机协会。ISBN 9781450367080。doi: 10.1145/3313831.3376727。[44] Aayushi Dangol、Robert Wolfe、Runhua Zhao、JaeWon Kim、Trushaa Ramanan、Katie Davis和Julie A Kientz。儿童对人工智能推理的心智模型：对人工智能素养教育的启示。载于第*24*届交互设计与儿童会议论文集，第106–123页。计算机协会，2025年。doi: 10.1145/3713043.3728856。

A 大语言模型使用声明

我们承认在本项工作中使用了大语言模型助手：大语言模型被用作写作工具，以协助完善论文。然而，最终版本由作者仔细审阅并定稿。构思阶段未使用任何大语言模型。

B 对所报告人工智能工具使用情况的汇总

我们汇总了 O’Neil 等人 [2, p.53] 的回答，以制作图 1。将所报告的类别相加，意味着每名儿童均被计入单一回答类别，但报告并未明确说明这一点。面向大众的 群组包括“ChatGPT”“Google”“社交媒体人工智能”“MicrosoftCopilot”“虚拟助手”“Claude”和“人工智能聊天机器人”，合计为 71%。其他人工智能工具 群组包括“照片、视频或视觉设计人工智能工具”“Grammarly”“Perplexity”“人工智能搜索”和“其他特定人工智能工具或类型”，合计为 10%。最后，以儿童为中心的 群组对应报告中的“教育人工智能工具”，占 4%。其余类别包括未指出具体工具、报告未使用人工智能或未被询问的儿童。报告中的参与者人数 $n = 1{,}204$ 乘以 0.85 后，用于估算使用人工智能工具的儿童总数（1,023）。我们将汇总后的类别重新按共同的 100% 基准进行换算，并使用最大余数法（汉密尔顿法）进行四舍五入，从而计算各人工智能群组所占比例，使其总和为 100%。