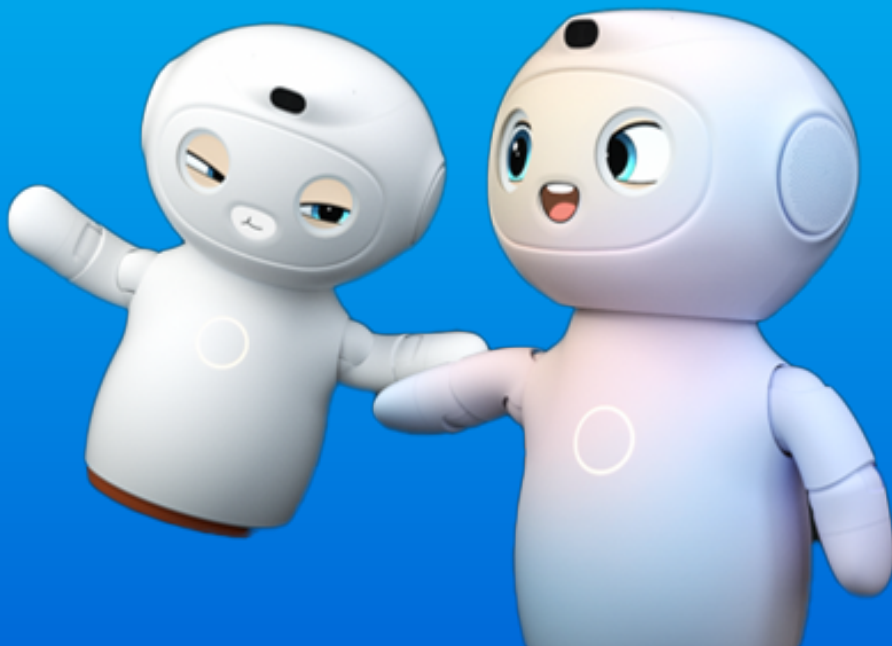


# 教育陪伴 机器人



## 案例研究

### 项目名称

华为教育陪伴机器人

### 实施机构

华为

### 项目类型

人工智能 (AI) 系统

### 项目情况

经过各个领域的专家团队（包括人工智能、机器人技术、游戏和动画设计、人因工程设计、数据安全、伦理学、教育学、发展心理学和认知心理学等）多阶段的设计过程调整，教育陪伴机器人已经可以更好地回应主要用户群体——3-12 岁儿童的需求和权益。在迭代设计时，团队贯彻以儿童为中心的设计宗旨，让儿童参与其中并为儿童服务，从而提高产品的可及性、透明度、公平性和非歧视性。同时，在产品设计之初构建预防性措施，将隐私和安全机制纳入机器人设计的全过程。

### 遵循联合国儿童基金会《人工智能为儿童—政策指南》的指导原则

项目旨在：

- + 促进儿童的发展和福祉
- + 注重平等对待每一名儿童，保护儿童免受歧视
- + 确保儿童参与并为儿童服务
- + 保护儿童的数据和隐私
- + 确保儿童安全

### 地点

中国及全球其他国家和地区

### 发布时间

2021 年

# 案例研究介绍

在芬兰政府的支持下，联合国儿童基金会于 2020 年发布了《人工智能为儿童——政策指南》，并于 2021 年发布 2.0 版本，鼓励商界在各自领域进行试点，并公开分享研究结果。本文是两个案例研究之一，聚焦联合国儿童基金会《人工智能为儿童——政策指南》应用情况，并以案例研究的形式展现其应用结果。这些案例研究是由中国传媒大学在中国网络社会组织联合会的指导下和联合国儿童基金会驻华办事处的支持下开展的。案例研究由埃莱奥诺·鲍威尔斯（Eleonore Pauwels）执笔。每个案例研究都说明了如何将指南中所述的、“以儿童为中心”的原则嵌入人工智能系统相关产品的开发、设计和推广的全过程中。

相关人工智能系统产品由企业生产，并为企业所有。联合国儿童基金会未参与该过程的任一阶段。本项目提及特定公司并不意味着联合国儿童基金会认可该公司的政策和做法。

联合国儿童基金会鼓励开源学习平台的设计、制作和传播。

# 概述

人工智能、机器人技术、教育和认知心理学的融合催生了新的应用，这些应用有望助力儿童提升学习和创造力、情感和社交能力以及逻辑思维和解决问题的能力。不过，这些创新性的应用也涉及与公平性、非歧视性、包容和参与、隐私和安全相关的伦理问题。

由于数据集或人工智能模型设计中存在偏差，基于数据和人工智能的技术可能会再现和加剧社会中的歧视和边缘化模式。如果设计不当，智能玩具和机器人伙伴可能会出现偏见或歧视行为。

保护儿童隐私是设计智能玩具时的另一个复杂挑战。跨文化和跨学科的研究发现，儿童有时会过度信任机器人并向其倾诉，分享私人和敏感信息。在家中和教室中与年幼用户互动时，陪伴机器人还会捕捉大量的视频和音频数据。联合国儿童基金会《人工智能为儿童——政策指南》指出，“如果缺乏充分的数据保护措施，这些数据可能会被出售至第三方，并且可能永远与儿童关联，或将对其未来接受高等教育或就业产生潜在影响”。

在此背景下，通过设置“以儿童为中心”的操作界面、体验活动、文本内容和工具，教育陪伴机器人能够个性化、自适应地陪伴儿童进行学习和游戏，并增强数据保护能力。在联合国儿童基金会《人工智能为儿童——政策指南》的启发下，教育陪伴机器人的开发过程运用了多项策略：

**与跨学科专家合作，开发了一个认知和心理学框架，用于按年龄分类评估儿童的发展需求。**

这一框架可以支持人工智能系统确定与儿童对话、游戏和学习的类型，从而更好地融入日常活动或家庭生活中。机器人提出的学习和社交活动也将父母和其他监护人纳入其中。

**实施了以儿童为中心的迭代设计，以提高安全性、可靠性、公平性和非歧视性。**

机器人整合了适合不同年龄段儿童的内容，禁止数字营销，仅限受信任的联系人访问。通过训练算法建立的数据集能确保客观和公正，避免偏见、刻板印象和成见的影响。

**研发团队进行了包括隐私保护在内的跨学科合作，以了解、预见并预防人工智能和机器人的技术交汇可能对儿童产生伤害的意外后果。**

在设计过程中，研发团队为机器人内置了隐私和安全保护机制。在设计和开发的各个阶段都考虑了现实生活和网络世界中的风险。

# 背景和项目起源

儿童与人工智能机器人的互动方式与成人不同。例如，儿童可能会把数字学习视为一种“体验之旅”，在这个过程中，他们会基于信任、分享等多方面的感官和情感体验，与机器人“教师”或“玩伴”进行互动。研究表明，由人工智能驱动的机器人有很大的可能会影响儿童行为和幸福感<sup>1</sup>。在人工智能机器人的支持下开展的个性化学习，可以满足儿童的独特需求、偏好和兴趣，适应他们的敏感性和个性。这样的教育伙伴可以帮助他们塑造个人的学习轨迹。

人工智能改变学习、文化、休闲和娱乐，但许多数字产品，包括智能和辅助技术，并非专为儿童而设计，在设计过程中未充分考虑儿童需求。这些产品在设计过程中缺乏必要的提前考量，无法为儿童赋能或让他们参与，同时未纳入风险预防措施，无法主动保护他们免受伤害和剥削<sup>2</sup>。这是研发团队跨学科领域的专家（包括人工智能、机器人技术、游戏和动画设计、人因工程设计、数据安全、伦理学、教育学、发展心理学和认知心理学等领域）共同努力得出的判断。

**以儿童为中心的人工智能产品的设计应该是一个跨学科的严谨过程，需要分析不同年龄段儿童的认知能力和心理特征，从而建立一个强大的认知框架来指导人工智能主导下的人机交互。**

研发团队认识到幼儿的成长速度不同，有些孩子在早期阶段就十分早熟。因此，团队在设计人工智能系统时，让机器人尝试适应每个儿童用户的学习进展和学习能力，并提供个性化的学习建议。机器人的多感知模型可以支持丰富多样的互动形式，包括适用于幼儿的目光接触、触摸和手势，以及适用于年龄稍大些的儿童的语言表达、讲故事和幽默感。

**在设计人工智能教育机器人和智能陪伴时，促进公平性和非歧视性是一项复杂但至关重要的研究和设计工作。**

研究表明，学龄前儿童可以从成人的口头和非口头线索中习得偏见，而这种社会模仿或学习可能导致种族和其他偏见。其他研究解释了为什么一些话题可能会变得敏感，例如在谈论发型时会间接涉及种族，谈论家庭时会涉及性取向，以及谈论假期时会涉及宗教信仰。联合国儿童基金会《人工智能为儿童——政策指南》中强调，由于人工智能存在偏见而导致的排斥和歧视可能对儿童造成严重的有害影响：“这种排斥可能对儿童产生长期影响，影响他们一生中的许多重要决定”<sup>3</sup>。

# 关于人工智能系统

教育陪伴机器人采用了多人交互模式，孩子们可以与监护人或其他孩子一起与机器人进行互动。这一模式旨在促进儿童社会情感能力的发展，为家长提供可以激发儿童兴趣的儿童友好资源（叙事 / 动画、对话话题、绘本），并对儿童用户的进步、兴趣和学习情况进行全面管理。

为了更好地回应 3-12 岁儿童的需求和权利，不同领域的专家团队参与到了教育陪伴机器人的设计过程中，这款教育陪伴机器人推出了一种全新的概念，力求最大限度地丰富儿童的社交互动，同时维护儿童权利，践行儿童保护原则。为了更好地满足每个儿童的需求，促进他们从社交到情感和智力的发展，这款人工智能陪伴机器人能提供个性化的建议，帮助儿童在值得信赖的环境中学习，发展社会情感能力和想象力，培养批判精神和逻辑思维能力。

在设计初期，人机交互（HRI）专家开发了一个多模态感知和决策程序——人工智能教育引擎，用来指导机器人行为，优化与儿童的交互活动。儿童心理学家、学者和教育工作者都参与了这一阶段的改进完善。他们共同开发了一个框架，可以按照年龄分类分析和评估儿童的认知、情感和社交能力及需求。

例如，年幼的儿童可能对探索和学习感兴趣，但如果活动时间过长或要求过高，他们很快会感到沮丧。4 岁到 6 岁的儿童开始表现出日益复杂的行为模式，大脑和身体功能的协调性也有所提升。他们变得越来越有同理心，能够理解他人的情绪和观点，因此可以通过讲故事和做游戏来增加社交活动，激发他们的兴趣。

设计过程的第二个关键点在于**机器人和人工智能软件感知、解释和理解儿童的情绪、语言、动作和意图，以及周围的家庭环境的能力**。例如，通过编程，让机器人的行为符合儿童的健康日常作息安排（包括睡觉时间、上学时间、刷牙时间、锻炼时间、玩耍时间、远离屏幕的时间），并与家庭生活活动紧密融合。人工智能陪伴机器人还可以组织需要父母或其他监护人参与的游戏、对话和阅读活动。

第三个方面，研发团队与学术界、教育界和商业界的伙伴合作，策划了一个以儿童为中心的大型内容库，内容包括互动游戏、益智游戏、阅读和学习材料和课程，以及对话中要探讨的各类主题。这一过程也注重参考外部专家意见，包括幼儿园和学校老师、自适应学习和教育领域的专家，以及来自中国科学院心理研究所和东南大学等学术机构的专家。受儿科学和发展心理学的启发，项目还强调了“游戏的力量”对于帮助儿童学习和沟通的重要性。

# 实施以儿童为中心的要求

教育陪伴机器人开发团队重点关注联合国儿童基金会《人工智能为儿童——政策指南》提出的要求，尤其是下列原则：促进儿童的发展和福祉，确保包容、多样性和儿童参与，注重平等对待每一名儿童，保护儿童免受歧视。

## 促进儿童的发展和福祉

研发团队认识到，陪伴机器人不能、也不应该取代家长和其他监护人的作用，因此在开发教育陪伴机器人时，研发团队重点关注帮助家长促进子女的发展和福祉，包括参与儿童日常生活中的重要活动和情境，改善对话中的聆听和沟通能力，鼓励在家中开展能够促进父母和幼儿之间互动的活动。

在构建机器人学习和发展框架时，研发团队明确了一系列重点领域，包括身体发展、认知发展和社会情感发展，并对中国、美国部分代表性的州（包括伊利诺伊州、哥伦比亚特区、威斯康星州、华盛顿州、特拉华州）、英联邦部分代表性国家以及中国香港地区儿童早期学习和发展标准中的重点领域进行了比较和系统研究。

## 注重平等对待每一名儿童，保护儿童免受歧视

为了防止人工智能的使用造成的潜在歧视，研发团队致力于让机器人能够支持不同社会、经济和文化背景的儿童开展各类教育和休闲活动，支持其社会情感发展。在数据收集和算法模型训练过程中，工程师们尝试整合各类儿童教育情境和行为学习策略，确保以有助于消除模型训练过程中潜在偏见的方式，收集、清理和标注用于训练机器人的算法数据集。

在产品发布之前，研发团队通过测试，并加强与外部合作伙伴的协作，对算法模型进行了评估和审计。例如，工程师们构建了一个完全独立的数据集，专门用于迭代测试，以发现并防止潜在的偏见。在整个设计周期中，工程师须采用高质量的数据集来确保数据的可用性、可靠性和完整性。

虽然工程师们都接受过有关数据和算法公平性的相关培训，但为了减少潜在偏见，在机器人设计过程中，也对所收集数据进行人工监督和手动干预。



## 确保儿童参与并为儿童服务

通过与教育学、发展心理学、认知心理学、人因设计、动画和游戏设计等领域的专家合作，研发团队打造了一款具有包容性的陪伴机器人。

它具有使用多模式交互（通过声音、视觉、触摸信号、手势和情感表达）的能力，以便视障儿童（或他们的监护人）依靠语音引导和触摸来启动和参与活动，如要求机器人读书或大声讲故事。大多数情况下，与机器人互动不需要太高数字素养。通过触摸、语音和视觉信号就可以操作，以便数字技能较低的监护人使用。

与成年人相比，儿童说话时在发音、情感表达、语言逻辑、重复音节的使用以及断句方面存在显著差异。为了应对这些挑战，研发团队面向大量儿童进行了测试，引入了适合儿童发展、发音、言语逻辑和情感的儿童语音识别框架，从而极大地改进了人工智能识别语音和与儿童对话的能力。研发团队制定了一套适用于不同年龄段儿童的语音助手标准，以帮助机器人尝试进行不同长度的中文或英文对话。

在中国多个城市和地区，研发团队一共对超过 3500 名 3-12 岁用户进行了测试。测试的方法各不相同，包括私立、公立小学和幼儿园的集体体验活动，以及焦点小组、调查等。测试对象涵盖首都、一二线城市的儿童用户，包括低收入和中产家庭用户，也包括城市和郊区的用户。

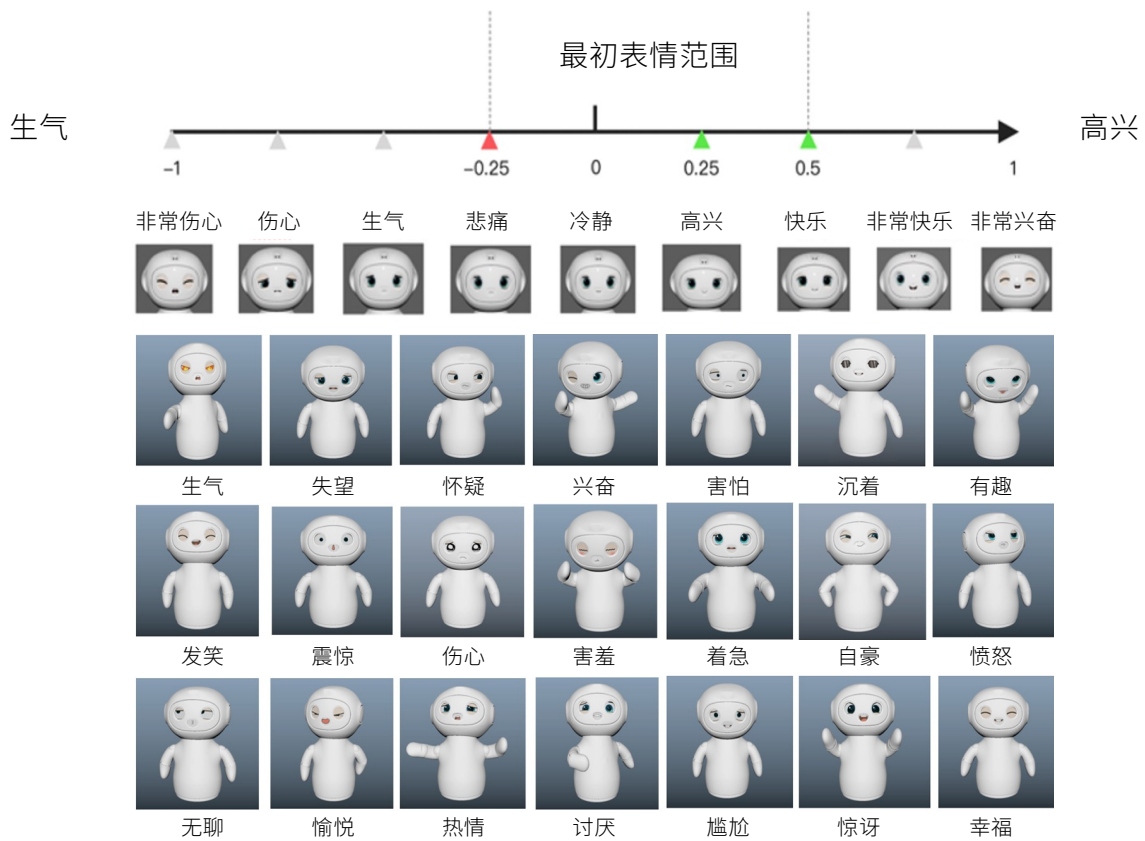
### ■ 对不同音色设计进行测试

为了确保机器人的声音受到儿童的欢迎，研发团队对 300 多名儿童进行了调查，并根据调查结果开发了适合不同年龄段儿童的中英文双语声音。

### ■ 对不同情绪和表达进行测试

随着儿童年龄的增长，他们的情感也逐渐变得更加多样和复杂。然而，他们往往缺乏充分管理自己情感表达的能力，并且可能容易做出冲动反应。经证明，陪伴机器人能够感知和表达情感，并对儿童表现出同情心。它们可以表达多达七种自身的情感。目前，有 2000 多种面部表情可供选择。例如，它们可以做出超过 16 种快乐的表情，如微笑、大笑、兴奋和狂喜（见图 1 和图 2）。

图 1 和图 2：华为陪伴机器人情感秀



## 保护儿童数据和隐私

研发团队根据《中华人民共和国个人信息保护法》《中华人民共和国网络安全法》和《中华人民共和国未成年人保护法》制定了教育陪伴机器人的儿童隐私保护政策。针对儿童隐私和数据保护，研发团队开展了尽职调查，包括但不限于以下内容：

### 透明度、同意和家长控制：

在与机器人互动之前，家长必须阅读并同意产品的隐私政策，了解人工智能陪伴机器人的隐私保护功能。隐私保护功能方面，家长会了解到所有数据采集能力（视频和音频采集）都会通过机器人上的灯光信号进行显示，并且可以通过触摸进行启用和禁用，或通过家长或其他监护人的安全手机远程操控。机器人不需要记录个人信息就可以与孩子互动（机器人甚至可以通过视觉算法分析识别孩子的年龄范围）。不过，为了提供更加个性化的体验，在最初“问候阶段”，监护人可以帮助儿童用户提供一些信息，如昵称以及对主题、对话和活动的偏好。机器人作为一个封闭系统运行，无法自动下载外部应用或程序。家长可以在手机上管理儿童对机器人的使用情况，例如，查看来电和去电，为机器人设置可信任联系人列表以接听或拨打电话。同时，家长可以查看机器人和儿童用户之间的活动类型，也可以使用手机清理聊天记录、删除游戏。



### ■ 数据安全：

华为公司表示，个人信息等用户数据是使用经过认证的数据安全技术进行管理的，如匿名化和加密技术。依据中国国家标准《信息安全技术 - 个人信息安全规范》（GB/T 35273-2020），华为根据重要性和敏感性对个人信息进行分类。华为公司表示，其数据安全机制已通过相关国际认证程序。

### ■ 第三方分享：

华为公司表示，他们仅将用户信息分享或转让给少数提供适龄互动内容（例如，绘本和游戏）的附属公司。工作人员会在相关信息传输前致电家长，家长可以选择同意或拒绝此类数据传输。华为会对可能接收数据的附属公司进行安全评估，并签订协议确保用户个人信息安全。华为公司表示，会使用特定技术对敏感信息进行匿名化和假名化处理。

### ■ 杜绝广告和营销内容：

华为禁止针对为儿童设计的服务进行个性化广告投放和直接营销。如需使用儿童个人信息以提供个性化体验或改善华为的教育服务，必须征得家长或其他监护人同意。

### ■ 视觉和音频数据采集：

华为公司表示，通过视觉技术采集的儿童数据仅会在机器人的数字系统上进行本地处理，不会上传到云端。语音命令（例如唤醒机器人时）需要通过云端进行识别，但华为确认不会长期存储此类音频数据。



## 确保儿童安全

除了上述数据和隐私保护措施外，预防安全和人身风险的措施还包括使用通过相关电子认证的环保材料。机器人的柔软材质、重量以及用于碰撞检测和运动控制的算法技术也都经过测试，尽可能减少翻倒或夹住手部等物理风险。

# 未来挑战

研发团队已经明确了有待进一步深入探讨和解决的挑战，其中之一就是联合国儿童基金会关于在开发儿童友好的人工智能系统时应具备透明度、可解释性和问责制度的要求。跨学科的协作设计加深了人们对人工智能、自主性和儿童权利融合时出现的伦理考虑的认识。虽然这些考虑可能并不一定适用于在封闭系统中运行的华为人工智能教育机器人，但它们仍然是跨领域和跨行业环境中值得讨论的考虑因素。

随着不同类型的拟人化机器人和设备学会参与儿童的日常活动，人工智能的伦理问题仍然存在，尤其是在陪伴儿童进行协作活动时。儿童有时可能会过度信任机器人或者认为他们具有人类情感。因此，在设计社交和教育机器人时必须特别谨慎，以保障儿童的权利。儿童与机器人互动领域遇到的一个重要的新挑战是保护和报告的责任问题。正如加州大学伯克利分校人权中心的研究团队在《2019年人工智能与儿童权利备忘录》<sup>4</sup>中指出的：在涉及报告儿童与智能玩具分享自杀念头、表现出自伤行为的迹象，或透露自己正在受虐待等情况时，企业的责任应如何界定？企业有哪些义务向有关当局报告此类信息，并采取哪些预防和保护措施？对于研发团队来说，另一个挑战是为父母制定一份易于理解的隐私政策和同意书，确保简单明了。

总体而言，虽然要赋予机器人权限和自主性，华为的研发团队认为有必要将伦理理念融入机器人设计中，并通过开放式商业对话建立一套专业的伦理指导，为用户提供积极的解决方案，力求预见、防范和缓解未来在透明度、可解释性和问责制度方面可能面临的挑战。

## 尾注

<sup>1</sup> 有关儿童神经和认知发展的信息，请参阅 Dion-Dosti E、Paquet N、Lasund M 等人。儿童多感官统合与神经发育 [J]. 脑科学, 2015 年, 5 (1) : 32-57。另请参见 P. Simoens、C. Mahieu、F. Ongenaes、F. De Backere、S. De Pestel、J. Nelis 等人, “机器人物联网: <https://ieeexplore.ieee.org/document/7776603> 辅助社交机器人的情境感知和个性化干预 (短论文)”, IEEE, 第 204-207 页, 2016 年 10 月。有关方法论和伦理观点, 请参阅 Tolksdorf NF、Siebert S、Zorn I、Horwath I、Rohlfing KJ (2020), 在幼儿园环境中应用机器人的伦理考虑: 从宏观角度出发, Int J Soc Robot, 13: 1-12。另请参见 Serholt S、Pareto L、Ekström S、Ljungblad S (2020), 儿童机器人交互中的故障与修复: 对小学教室中与机器人学生的复杂交互的研究, Front Robot AI, 7: 46。另请参见 Belpaeme T、Kennedy J、Ramachandran A、Scassellati B、Tanaka F (2018), 教育社交机器人: <https://www.science.org/doi/10.1126/scirobotics.aat5954> 回顾, Science robotics, 3(21)。另请参见 Belpaeme T、Vogt P、Van den Berghe R、Bergmann K、Göksun T、De Haas M、Kanero J、Kennedy J、Küntay AC、Oudgenoeg-Paz O 等人 (2018), 设计社交机器人作为第二语言导师的指南, <https://pubmed.ncbi.nlm.nih.gov/30996752/> Int J Social Robot, 10(3):325-341。

<sup>2</sup> 将人为因素和需求应用于系统和服务的设计和开发。

<sup>3</sup> (第 22 页), 联合国儿童基金会全球洞察与政策办公室, “《人工智能为儿童——政策指南》(2.0 版)”, 2021 年。

<sup>4</sup> (第 55 页), 加州大学伯克利分校人权中心研究团队, 人工智能和儿童权利备忘录, 2019 年 4 月 30 日。