arXiv:2302.01039v1 [cs.HC] 2 Feb 2023

可解释性的人机训练与增强现实协作

CHAO WANG*, ANNABELARDINELLI*, STEPHANHASLER*, THEODOROSSTOURAITIS*, DANIELTANNEBERG*, and MICHAEL GIENGER*, 本田欧洲研究院，德国

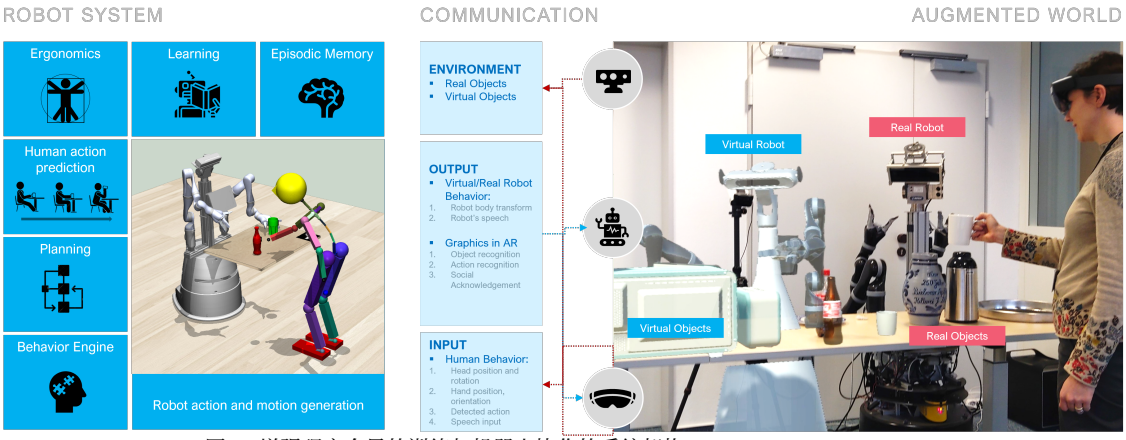

图1。增强现实介导的训练与机器人协作的系统架构。

当前，社交机器人和辅助机器人技术应用日益普及，这越来越凸显了对易于教学和交互的机器人的需求，即使是没有技术背景的用户也能使用。然而，人们往往难以了解这类机器人掌握了哪些信息，也难以评估机器人是否正在形成对任务的正确表征。增强现实（AR）有望弥合这一差距。我们展示了三个应用案例，其中增强现实设计元素提升了人机交互的可解释性和效率：1）人类通过示范教机器人完成一些简单的厨房任务；2）机器人在增强现实中向人类展示其解决新任务的计划，以供验证；3）机器人在日常活动中为行动不便的人提供帮助时，通过增强现实传达其意图。

CCS 概念： • **计算机系统组织** → **嵌入式系统**； 冗余；机器人学； • **网络** → 网络可靠性。

其他关键词和短语：可解释性、人机交互、增强现实

ACM 参考文献格式：
Chao Wang、Anna Belardinelli、Stephan Hasler、Theodoros Stouraitis、Daniel Tanneberg 和 Michael Gienger。2023。《借助增强现实实现可解释的人机训练与协作》。载于*CHI 2023*：*ACM*计算机系统中的人为因素会议，*2023*年*4*月*23*日至*28*日，德国汉堡。 ACM，纽约，美国纽约州，6页。https://doi.org/XXXXXXX.XXXXXXX

*所有作者对本研究的贡献均等。

允许为个人或课堂使用而免费制作本作品的全部或部分数字或纸质副本，但前提是不得为营利或商业利益而制作或分发副本，并且副本首页必须附有本声明和完整引文。对于本作品中由 ACM 以外其他方拥有版权的组成部分，必须尊重其版权。经注明出处后，可以进行摘要。以其他方式复制、重新出版、发布到服务器或重新分发到列表，均须事先获得具体许可和/或支付费用。请通过 permissions@acm.org 申请许可。© 2023 计算机协会。提交给 ACM 的稿件

1 引言

要让社会辅助机器人进入我们的家庭并在我们身边行动，首要条件之一是，即使非专业用户也能够直观地与机器人交互，并教会它新任务。这类系统旨在进行长期交互，因此需要足够灵活，能够处理大量任务并学习新任务。这意味着它们的运行会相当复杂，依赖多种人工智能和机器学习算法，终端用户无法理解这些算法。与此同时，机器人需要以有效的方式向用户说明自身的工作方式，使用户能够调整自己的心智模型，并了解对机器人的预期，而无需接受专门培训、阅读冗长的说明，或反复观察机器人。这样的特性将大大有助于提升公众对机器人助手的接受度和信任。近年来，这一需求已在非具身人工智能系统中显现，并由可解释人工智能（XAI）领域加以应对 [3], ；在移动机器人学中，它此前被称为可解释性或可读性 [2], ，近期则通过特定的机器人维度加以刻画[4]。事实上，纯人工智能系统通常会对机器学习算法所作出的决策或预测提供事后解释，而机器人是具身智能体，会在现实世界中自主行动，并可能与人类交互和协作。这种情境性和能动性要求采用不同类型的解释，不仅要说明过去的决策，更重要的是说明当前的评估、未来的计划和意图。这些信息需要及时传达给人类用户，并以快速、明确的方式与他们进行交互式协商。在人机交互领域，这类交互往往通过模仿人类之间的交互和社交线索来实现，例如目光接触和共同注意、语音、手势 [11], ，其理念是，行为上更像人类能够提升直观性。然而，由于这些线索为人们所熟悉，它们可能会让人误以为机器人拥有与人类对等的感知和理解能力。从这个意义上说，其他交互模态可以与这些线索相辅相成，让用户得以了解机器人的“思维”，从而弥合用户心智模型中的空白。

我们提出一种运用增强现实（AR）的概念，通过显示直观的说明性提示，分层呈现机器人感知、学习、泛化和规划，从而提升用户与机器人助手交互的体验。尽管增强现实／混合现实解决方案近年来在机器人学领域日益普及 [1] ,，但这些方案通常面向受过训练的专家，用于调试 [6], 、可视化内部状态和运动意图 [9], 、远程操作 [14],，以及简化运动轨迹的示教学习 [5]。在此，我们介绍一种机器人系统，它利用增强现实的能力，涵盖机器人学中的多种典型应用场景：通过示范学习新任务、为新问题制定计划，以及以符合人体工学的方式协助人类完成体力任务。通过利用真实物体和虚拟物体，并将类人社交线索与说明性虚拟设计元素相结合 [10],，该框架展示了如何在人类与机器人之间共享混合工作空间。这类界面让机器人的状态清晰直观地呈现在用户面前，并以用户对当前场景的人类感知为依据，从而增强交互。

2 系统架构

我们的系统由多个模块组成，分别实现增强现实眼镜的后端功能（学习、规划、预测和运动生成）以及前端界面能力（可视化和交互）（见图 1）。

后端机器人系统：实体机器人由一个躯干、两条 Kinova机械臂[1] （每条具有 7 个自由度）和一个云台组成，隶属于移动平台“Johnny” [13]。机器人的认知技能由多个 ROS[2] 节点实现，

[1]https://www.kinovarobotics.com/product/gen2-robots [2]
https://www.ros.org

这些节点彼此通信，并与前端界面通信。行为引擎负责协调机器人的社交行为。它控制机器人注视，同时发出可在增强现实环境中显示的可解释人工智能提示（参见 [12]）。它还负责调节语音交互，例如确认用户指令或提出由好奇心驱动的问题（见第 3 节）。情景记忆会收集人类教师进行示范时的世界状态观测。学习模块实现符号技能学习，将情景记忆中的示范整合到知识图谱中，并生成新的假设供用户确认（第 3 节）。规划模块利用已学习的语义技能生成高级计划，以解决新任务。这些计划仍需由人类导师检查（见第 4 节）。人类动作预测模块在协助用户时运行。机器人可以预测用户打算实现的目标，并规划辅助动作，例如将接下来需要的物体移到用户附近。人体工程学模块以符合人体工程学的最佳方式为用户生成干预措施（见第 5 节）。最后，动作生成模块（图 1）接收其他模块发出的动作指令（例如看向何处、抓取什么等），并执行相应的运动行为。

前端界面：虚拟物体、虚拟机器人和可解释人工智能提示（增强现实图形元素）通过 HoloLens[3] 显示在混合现实环境中。HoloLens 可以扫描周围环境，构建环境中物体的三维网格，并定位自身在房间中的位置，从而能够在环境中稳定地叠加图形，同时考虑与真实物体之间的遮挡。

用户可通过 HoloLens 访问共享环境（见图 1 右侧），其中包含多个真实物体和虚拟物体。物体姿态会持续发送到后端：虚拟物体的姿态由 HoloLens 发送，真实物体的姿态则由固定摄像机发送（这些物体带有基准标记）。用户可以像与真实物体交互一样与虚拟物体交互；例如，他们可以拿起面包，将其放入烤面包机的槽中，然后按下按钮。

HoloLens 还会检测用户的行为，并将其传递给后端系统。这包括用户的头部位置、朝向和语音输入。更重要的是，由于 HoloLens 能够追踪用户的手和手指，因此手部动作（例如“拾取”或“放下”）也会被检测到。通过 Microsoft MRTK SDK 4 实现人手对物体的操作；当执行“拾取/ ಹಿಡಿ持”手势时，相应的虚拟物体会附着在人的手上，而在检测到“放下”手势后，该物体会从手上释放。此外，用户的手指上还附加了碰撞体，使教师能够按下烤面包机的控制杆、打开微波炉的电源按钮以及关闭微波炉门。最后，用户行为及相关被操作物体的信息会通过 ROS 发送至后端系统（见图 1“输入”）。该系统还可以显示并为一个全息虚拟机器人设置动画，其外观几乎与实体机器人完全一致。HoloLens 接收机器人行为数据（见图 1“输出”），包括虚拟机器人的姿态和语音命令，并将其进行可视化或语音播报。此外，后端还会触发可解释人工智能提示信息的显示，这些提示信息会呈现在增强现实环境中。接下来的章节将介绍三个基于该系统架构、利用增强现实增强人机交互的使用案例。

3 用例1：用于模仿学习的可解释的人机交互 G

机器人学中的示教学习或模仿学习旨在让人类通过亲自演示任务来教会机器人新技能，就像他们会向另一个人演示那样。这种方法的主要局限在于，机器人能否理解并泛化特定的示范，以及人类用户难以了解怎样的示范才适合机器人。因此，我们开发了一种两阶段技能学习概念。在第一阶段，用户向机器人演示一项技能，机器人运用语义技能学习概念来习得该技能。

[3]https://www.microsoft.com/en-us/hololens [4]
https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/?view=mrtkunity-2021-05

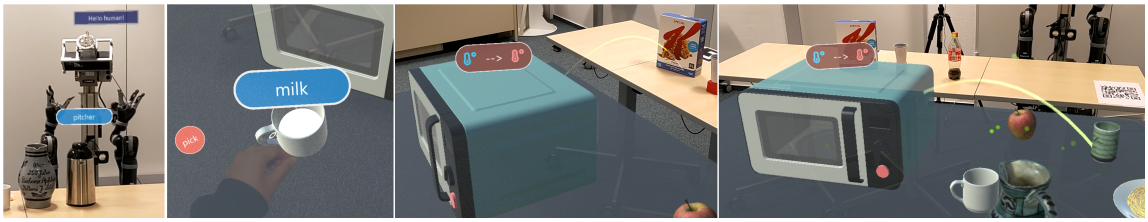

图 2。借助可解释人工智能进行示范学习：用户示范新技能时，机器人通过增强现实提示其识别出的动作和物体。随后，机器人提出相关问题，同时突出显示涉及的物体。

该技能的学习到的表示由符号构成，用于编码前置条件、动作和效果。在第二阶段，机器人会主动向用户询问有关所示范任务的问题。两个阶段均旨在运用增强现实和社交线索，增强用户对系统的心智模型。

具体而言，用户（佩戴增强现实眼镜）向机器人示范一项技能，例如与（虚拟）物体交互并提供语言解释（技能标签）。机器人跟随教师的视线，或回望教师，以显示其注意力。教师还会持续获知机器人的感知结果：每当教师注视某个物体，或识别出某个手动动作时，物体和动作标签（可解释人工智能提示）就会弹出（参见图 2 和[12]）。情景记忆会在示范技能之前和之后记录环境状态，包括物体和智能体。随后，这些观察结果会被解析为预定义的环境符号表示，由逻辑谓词组成。结合技能标签，这些观察结果被用于学习符号技能，记录该技能可应用的情境，以及执行该技能会如何改变环境。

如果学习了效果相似但涉及不同物体的技能，机器人就可以利用预定义的物体层级结构对这些技能进行泛化。这种技能的类型泛化使机器人能够在新任务中考虑新物体。技能表示的这种符号化特性，使机器人能够以直观的方式解释已学到的知识或未来的计划。

虽然这种模仿学习场景对用户来说直观且引人入胜，但示范技能仍然需要花费时间和精力。为了加快交互式学习过程，机器人具备了就技能和物体提出好奇问题的能力，也就是创建假设并呈现给用户确认或否决。与完整的技能示范相比，这使用户能够以交互式且更快速的方式向系统教授额外知识。例如，在看到一段用于学习如何使用微波炉的示范（教师加热了牛奶）后，机器人可以询问类似的物体（例如水）是否也能在微波炉中加热。在这种情况下，可解释人工智能的虚拟元素会突出显示候选物体和微波炉，并在用户回答后用红色或绿色粒子表示否定或肯定的回答（见图 2 右侧各图）。

4 用例 2：机器人规划的可视化

学习到的技能包含关于不同设备物理效果的高层知识，以及关于如何操作这些设备和操控物体的低层知识。我们使用标准的符号化 STRIPS 规划器，将这些技能组合起来，针对当前存在的物体（包括真实物体和虚拟物体）解决新颖且更复杂的任务。例如，当被要求“准备一杯冰茶”时，规划器可能会考虑先用水壶或微波炉加热一些水，再放入茶包，之后用冰箱或一些冰块将其冷却。随后，将这一高层解决方案扩展为所需的低层操作，例如摆放物体、打开门和按下按钮。

通过学习技能并将其应用于新的任务和环境，系统可以根据先前观察到的情节进行泛化。因此，生成的计划可能不可行或不符合预期，需要由人类进行验证。在此，我们提出一种交互式系统，可通过增强现实眼镜向人类展示机器人的计划

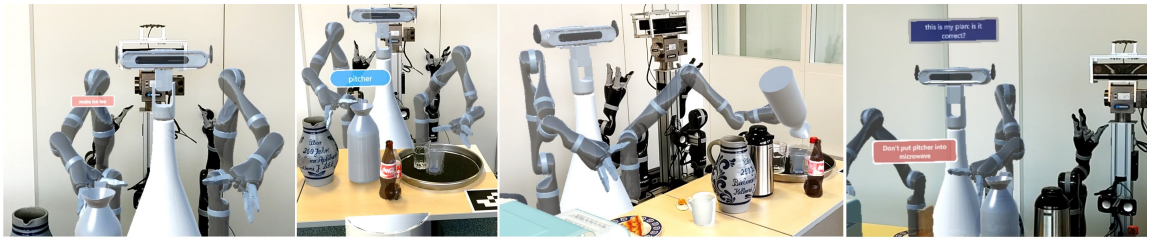

图 3. 规划阶段可视化，从左至右：用户通过语音向机器人发出指令；机器人展示其虚拟化身和真实物体的数字孪生；虚拟化身执行计划；人类对计划提供反馈。红框显示语音识别检测到的用户指令。

并在执行前进行确认（见图 3）。用户可以通过语音向机器人发出指令。随后，机器人会根据其不同层级的知识生成解决该查询的计划。执行前，机器人会请用户在增强现实中验证计划。机器人的虚拟“虚拟化身”叠加显示在机器人的实体上，真实物体的“影子”（全息孪生体）则显示在增强现实眼镜中。然后，虚拟机器人将使用虚拟物体执行计划。这样，人类便能理解机器人推理，并对其计划提供反馈，批准或修正计划。

5 用例 3：辅助用户时传达机器人的意图

人类与机器人之间的高效协作面临诸多挑战，其中包括：(i) 人类行为预测，(ii) 预测机器人动作对人类动作的影响，以及 (iii) 生成易于人类伙伴理解的可理解的机器人动作 [2]。在此，我们在操作场景中研究这些挑战：机器人协助人类完成一系列动作。一个典型例子是，如图 4 所示，一名坐在轮椅上的人类端取饮料。这一场景要求人类弯腰或伸展身体去拿玻璃杯，因此机器人助手可以提供帮助。

为了支持人类，机器人理想情况下会预测可能的人类动作与姿态序列 [7], 评估未来的人类姿态 [8], 计算能够改善人类身体状态的预期干预措施，并提前向人类传达预期干预措施的效果。为实现这些目标，我们提出一种基于模型的优化方法，对人类动作进行基于物理的预测，使机器人能够判断哪种干预措施可以改善人体姿态，进而改善其人体工学状态，并通过可解释人工智能提示向人类展示这些干预措施。

在此用例中，我们考虑一个桌面场景，系统已获知若干物体的位置以及预期的高层任务。此类任务可以是端上一杯饮料或一碗麦片。我们的系统能够：（i）预测人类动作的序列，例如拿起瓶子，然后往玻璃杯里倒饮料；（ii）评估人体姿态，例如倒饮料时人类的上身构型，并决定如何调整连续量，例如改变玻璃杯的姿势，以改善人类倒饮料时的上身构型；（iii）在执行辅助动作的同时，使用增强现实告知人类（例如使用全息图显示物体将被移至何处）。利用增强现实展示未来干预措施的结果，可以让用户理解机器人动作，从而使其能够轻松流畅地完成任务。

6 结论

我们的系统旨在提供有关机器人的关键信息，减轻用户的认知和身体负荷，使用户能够以自然的方式行动和交互，而无需持续监控和仔细检查机器人，只需

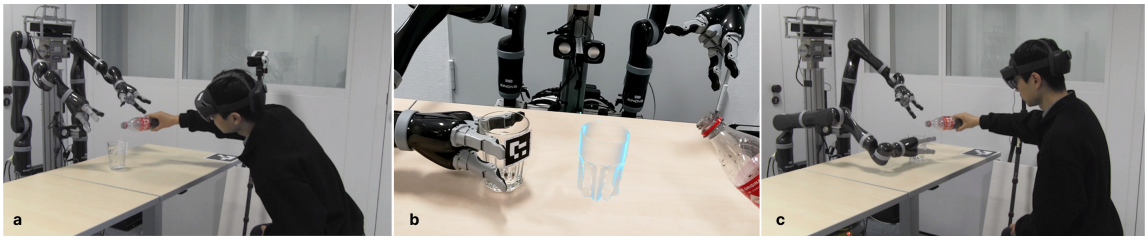

图 4。实例 (a) 展示了一个没有协助的情景：人类弯腰将液体倒入玻璃杯，姿势不便。相比之下，实例 (b) 和 (c) 是有协助情景中的连续两帧。在 (b) 中，机器人预测人类将要把液体倒入玻璃杯，并在移动玻璃杯的同时利用可解释人工智能提示揭示其干预结果。在 (c) 中，人类以舒适的姿势将液体倒入玻璃杯。

在必要时才对其进行纠正。虽然我们在此聚焦于机器人在家庭环境中提供辅助这一特定场景，但我们实现了基于增强现实的设计概念，以满足人们在与物理智能系统交互时日益增长的理解和解释需求；我们也希望借此为更广泛的 CHI 社区带来启发，为智能家居或其他以人为本的人工智能系统提供相关解决方案。不过，我们的工作重点是为非专业人士提供便利的机器人交互，并提升老年人和行动不便者的独立性，从而促进人机交互中的无障碍性和包容性。

参考文献

[1] Zahraa Bassyouni 和 Imad H. Elhajj。2021。增强现实与机器人学中的人工智能：系统综述。机器人学与人工智能前沿 8（2021年9月）。https://doi.org/10.3389/frobt.2021.724798[2] Anca D. Dragan、Kenton C.T. Lee 和 Siddhartha S. Srinivasa。2013。机器人运动的可读性与可预测性。载于*2013*年第*8*届 *ACM/IEEE* 人机交互国际会议（*HRI*）。301–308。https://doi.org/10.1109/HRI.2013.6483603[3] David Gunning、Mark Stefik、Jaesik Choi、Timothy Miller、Simone Stumpf 和 Guang-Zhong Yang。2019。可解释人工智能。*Science Robotics* 4，37（2019），eaay7120。[4] Thomas Hellström 和 Suna Bensch。2018。可理解的机器人——是什么、为什么以及如何实现。*Paladyn*，行为机器人学期刊 9，1（2018年7月），110–123。https://doi.org/10.1515/pjbr-2018-0009[5] Matthew B Luebbers、Connor Brooks、Minjae John Kim、Daniel Szafir 和 Bradley Hayes。2019。用于受限示范学习的增强现实界面。载于第*2*届面向人机交互的虚拟、增强与混合现实国际研讨会论文集（*VAM-HRI*）。[6] Alexandros Rotsidis、Andreas Theodorou、Joanna J Bryson 和 Robert H Wortham。2019。提升机器人透明度：一项关于移动增强现实的研究。载于*2019*年第*28*届 *IEEE* 机器人与人类交互通信国际会议（*RO-MAN*）。IEEE，1–8。[7] Theodoros Stouraitis、Iordanis Chatzinikolaidis、Michael Gienger 和 Sethu Vijayakumar。2020。通过双层优化实现双人协作操作的在线混合运动规划。*IEEE*机器人汇刊 36，5（2020），1452–1471。https://doi.org/10.1109/TRO.2020.2992987[8] Linda van der Spaa、Michael Gienger、Tamas Bates 和 Jens Kober。2020。预测并优化实体人机协作任务中的人体工程学。载于*2020*年 *IEEE* 机器人与自动化国际会议（*ICRA*）。IEEE，1799–1805。[9] Michael Walker、Hooman Hedayati、Jennifer Lee 和 Daniel Szafir。2018。利用增强现实传达机器人运动意图。载于*2018*年 *ACM/IEEE* 人机交互国际会议论文集。ACM，纽约，美国，316–324。https://dl.acm.org/doi/ 10.1145/3171221.3171253[10] Michael Walker、Thao Phung、Tathagata Chakraborti、Tom Williams 和 Daniel Szafir。2022。面向人机交互的虚拟、增强与混合现实：综述与虚拟设计元素分类体系。*arXiv*预印本 *arXiv:2202.11249*（2022）。[11] Sebastian Wallkötter、Silvia Tulli、Ginevra Castellano、Ana Paiva 和 Mohamed Chetouani。2021。通过社交线索实现可解释的具身智能体。*ACM* 人机交互汇刊 10，3（2021年9月），1–24。https://doi.org/10.1145/3457188[12] Chao Wang 和 Anna Belardinelli。2022。利用增强现实研究可解释的人机交互。载于第*5*届面向人机交互的虚拟、增强与混合现实国际研讨会论文集（*VAM-HRI*）。[13] Chao Wang、Stephan Hasler、Manuel Mühlig、Frank Joublin、Antonello Ceravola、Joerg Deigmoeller、Lydia Fischer 和 Pengcheng An。2021。面向办公环境中多智能体协作系统的交互设计。计算机协会，纽约，美国。[14] A.W.W. Yew、S.K. Ong 和 A.Y.C. Nee。2017。用于维护机器人远程操作的沉浸式增强现实环境。*ProcediaCIRP* 61（2017），305–310。https://doi.org/10.1016/j.procir.2016.11.183

收稿日期：2023年1月19日