arXiv:2303.02160v1 [cs.HC] 2 Mar 2023

像我一样导航：理解人们如何评价电子游戏中的类人 AI

Stephanie Milani
smilani@andrew.cmu.edu 卡内基梅隆大学 宾夕法尼亚州匹兹堡，美国

Arthur Juliani 微软研究院 纽约，纽约，美国

Ida Momennejad 微软研究院 纽约，纽约，美国

Raluca Georgescu
微软研究院 剑桥，英国

Jaroslaw Rzpecki
Monumo 剑桥，英国

Alison Shaw Ninja Theory 剑桥，英国

Gavin Costello
Ninja Theory 剑桥，英国

Fei Fang 卡内基梅隆大学 匹兹堡，宾夕法尼亚州，美国

Sam Devlin 微软研究院 剑桥，英国

Katja Hofmann 微软研究院 剑桥，英国

摘要

我们旨在了解人们如何评估电子游戏中由人类和人工智能（AI）智能体生成的导航行为的类人性。为此，我们提出了一种新型人工智能智能体，旨在生成更像人类的行为。我们收集了数百份众包评估，将该智能体和基线人工智能代理生成的导航行为与人类生成的行为进行类人性比较。我们的智能体通过了图灵测试，而基线智能体未能通过。所谓通过图灵测试，是指人类评判者无法通过量化方式区分人类和人工智能智能体进行导航的视频。为了了解人们认为哪些特征构成了类人导航，我们对这些评估的理由进行了广泛分析。本研究揭示了人们在目标导向的电子游戏导航情境中所认为的类人特征，这是进一步改善人机交互的关键一步。

CCS 概念

• **应用计算** → **电脑游戏**；• **计算方法** → 强化学习；• **以人为中心的计算** →人机交互中的实证研究。

关键词

人类受试者研究，可信 AI，游戏，导航

允许为个人或课堂用途免费制作本作品部分或全部内容的数字副本或纸质副本，前提是不得为牟利或商业利益制作或分发副本，并且副本须附有本声明及首页上的完整引文。必须尊重本作品第三方组件的版权。如需用于其他用途，请联系所有者/作者。*CHI '23，2023*年*4*月*23*日至*28*日，德国汉堡© 2023 版权所有者/作者所有。ACM ISBN 978-1-4503-9421-5/23/04。
https://doi.org/10.1145/3544548.3581348

ACM 引用格式：
Stephanie Milani、Arthur Juliani、Ida Momennejad、RalucaGeorgescu、Jaroslaw Rzpecki、Alison Shaw、Gavin Costello、Fei Fang、Sam Devlin 和 Katja Hofmann。2023。像我一样导航：理解人们如何评价电子游戏中的类人 AI。载于 *2023*年计算机系统中的人为因素会议论文集（*CHI '23*），*2023*年*4*月*23*日至*28*日，德国汉堡。ACM，美国纽约州纽约市，18页。https://doi.org/10.1145/ 3544548.3581348

1 引言

游戏被视为人类社会互动最古老的形式之一 [49, 79]。纵观历史，人们一直通过游戏开展重要的文化与社会联结活动 [45], 、将其作为教学与学习工具 [15, 44], , 并从中获得乐趣 [41]。如今，电子游戏已成为一种流行的结构化游戏形式，吸引玩家沉浸于引人入胜的虚拟世界中。这种沉浸感对于提升游戏的乐趣至关重要 [11, 53]。

为了提升这种沉浸式体验，游戏设计师专注于创造可信的非玩家角色（NPC），使其能够以多种方式与玩家互动。可信度的一个重要组成部分是类人性 ——也就是说，NPC 能够像人一样行动。由于许多玩家与非玩家角色的互动对游戏至关重要，因此 NPC 的行为必须可信，才能维持沉浸感 [14, 35, 85]。事实上，电子游戏玩家认为，与更像人类的智能体对战会更有趣 [72]。传统上，游戏开发者设计 NPC 时，会让它们遵循一套预先设定的行动。不过，这种方法既耗时又具有挑战性，因此促使设计师转向人工智能（AI），寻求其在 NPC 设计方面的帮助。随着这一转变，有必要开展研究，了解人们认为人工智能智能体具备哪些类人特征。

与此同时，人工智能研究人员已将实现复杂的类人行为视为一项关键里程碑， [8, 21, 71]以开发能够与人类灵活协作的智能体

应用于共享的人类-AI环境 [9, 25, 84] 及各种机器人应用 [64]。仅仅让智能体在指定任务上表现出高水平的熟练度，并不足以实现这一目标。例如，人工智能驱动车辆必须表现得足够像人类，以便人类驾驶员能够在其周围解读其行为、预测其行动并作出反应 [27]。因此，了解哪些行为会影响人们对类人性的感知，是实现人工智能体普遍类人行为的基础性第一步。

在这项工作中，我们通过识别并理解电子游戏中的类人行为是什么，为开发类人智能体这一目标做出贡献。为了界定研究范围，我们专注于一款 3D电子游戏，智能体必须从一个地点导航至另一个地点。这种导航形式在许多电子游戏中都很普遍，因此成为游戏开发者关注的重点 [3, 18]：在具身游戏中，玩家必须从一处移动到另一处，以达成目标或探索世界。更广泛而言，这被认为是具身生物智能的基础 [26, 59], ，因此也受到认知科学家的关注 [56, 62] ，以及对智能行为感兴趣的研究者的关注 [87]。它也一直是人机交互领域的重点 [80] ，因为人们（或机器人）如何在现实、增强或完全虚拟的空间中导航，是一个重要议题。

为了研究电子游戏中的导航，我们采用了近期提出的人类导航图灵测试（HNTT） [16], ，在该测试中，人类评判者需要指出两段视频中哪一段展现了更像人类的行为。随后，评判者需要说明做出该判断的理由，并指出对选择的确信程度。在那项工作中，作者将类人性评估的准确性与随机水平进行了比较，但没有采用统计检验来明确判定某个智能体是否通过了 HNTT。根据他们的评估，所研究的两个人工智能智能体都未通过 HNTT。因此，打造能够通过 HNTT 的智能体仍是一项尚待解决的挑战。为此，我们设计了一种新型智能体，以通过 HNTT。为了协助设计类人智能体，我们考察了先前工作中两个基线智能体产生的行为 [16]。基于这些发现，我们运用简单直观的技术，设计了新型的 奖励塑形智能体。

随后，我们在亚马逊土耳其机器人（MTurk）上针对 HNTT 开展了一项行为研究，以调查我们的智能体和基线智能体的行为。为判断智能体是否通过 HNTT，我们提出了一项明确的判定标准：通过统计检验，判断人类评判者区分人类行为与智能体行为的能力是否显著不同 于随机水平。随后，我们验证了先前研究的结论：两个基线智能体不够类人，因为它们未能通过 HNTT。相比之下，人类评判者无法可靠地区分我们的奖励塑形 智能体与由人控制的智能体的行为。据我们所知，这是首个通过 HNTT 的智能体。

为理解这些评估，我们分析了开放式回答，以确定人们认为哪些特征能够代表人类与人工智能的导航行为。我们对回答进行标注，并用编码概括其中给出的理由。根据这些标注，我们发现，人们描述类人行为与非类人行为的方式存在关键差异。具体而言，我们发现，人们在描述类人行为和非类人行为时会采用相同的高层次特征，但这些特征是否存在会显著影响他们的判断。

根据分析结果，我们总结了开发和评估人工智能智能体类人性时需要考虑的事项。例如，明确智能体的最终用途对于界定何为类人至关重要，也有助于据此设计研究。总之，我们作出以下贡献。

（1）我们提出了一种新颖的奖励塑形 智能体，其导航行为更像人类。（2）我们开展了一项行为研究，以评估：a）人们能否可靠地区分人工智能智能体产生的行为与人类产生的行为；b）人们认为哪些特征能够体现类人行为。（3）我们对所得数据进行了全面分析。我们提出了一项明确的判定标准，用于判断智能体是否通过 HNTT；根据这一指标，我们发现只有我们的奖励塑形 智能体通过了 HNTT。我们分析了自由形式回答，以确定人们认为哪些特征能够代表类人行为。（4）基于研究结果，我们提出了开发和评估类人 AI 的具体建议。

2 相关工作

研究人员采用了多种方法来应对在游戏中开发可信的人工智能智能体这一挑战，包括从示范中学习 [33, 38, 52], [5, 20, 52, 强化学习 89], [54, 77] [73] 等。我们专注于强化学习，因为它提供了一套普遍适用的算法，可用于学习在各种场景（包括但不限于现代游戏环境）中控制智能体 [3, 24, 43, 75, 81]。作为生成导航行为的一种方法，它也具有显著优势 [3]。具体而言，强化学习的应用可能有助于实现更复杂的导航能力（例如抓钩移动或传送），并使游戏设计师免于采用最流行的替代方法来生成此类行为时所需的繁重工作 [51]。

在强化学习中，智能体通过最大化奖励（或分数）来学习完成任务；奖励会告知智能体其表现如何。尽管智能体通过最大化奖励学会有效导航，但它们不会考虑自身行动的风格 [3]。从业者明确指出，若要在商业游戏开发中采用这些方法，控制行动风格至关重要 [34]。举例来说，近期在现代游戏中击败人类世界冠军选手的强化学习方法表现出异常行为 [30] ，使得人类与人工智能组成混合团队时的协作游戏效果大打折扣 [6]。仅仅最大化任务专属的奖励信号，不太可能造就类人智能体。

奖励塑形 [58, 86] 是一种简单而强大的技术，能够让从业者清晰地指定期望的智能体行为。这种方法通过设计奖励信号，为智能体提供密集反馈。对于没有机器学习背景的人来说，这是一种直观的控制智能体行为的方式：只需指定目标，无须花时间优化那些不直观的超参数。此外，奖励塑形可与任何强化学习算法搭配使用，因此可以根据需要替换底层算法。我们利用奖励塑形来生成更像人类的行为。



图1：研究参与者所观察到的导航任务（左：屏幕截图），以及电子游戏关卡小地图的细节（右）。智能体在主地图之外的岛屿上生成，主地图显示在右侧小地图的下半部分。它们必须跳到主区域，并导航至目标位置。左侧屏幕截图中的浅蓝色容器表示目标位置。

目前尚无用于评估类人 AI 的标准指标体系。一种范式是使用人类判断的代理指标来衡量与人类的相似性。有些研究会衡量人工智能的任务表现 [77, 89], , 但这一指标不足以作为与人类相似性的代理指标。另一些研究会评估人工智能智能体预测接下来的人类动作 [33]或使其行为与人类保持一致的能力 [54, 76], , 但这些指标并未纳入实际的人类评估。它们没有评估人们能否准确区分人工智能玩家和人类玩家，而这对于评估游戏 [22] 及其他领域 [88]中的类人性至关重要。

涉及人类评估的研究往往采用小规模问卷，以了解人们对类人性的看法 [20, 52]。这些研究通常只对影响此类看法的具体特征进行初步探究，而且一般不包含图灵测试 [39], , 而图灵测试是解决这些问题的成熟框架 [22]。采用图灵测试的研究往往不会探究相关行为，也不会提供具体指标 [55], ; 或者，它们侧重于评估游戏行为的整体范围 [5, 20, 52]。由于这类游戏及其产生的行为十分复杂，要向游戏设计师提出具体建议颇具挑战。相比之下，我们专注于一种具体但广泛使用的行为：点到点导航。为了开展评估，我们采用了近期提出的人类导航图灵测试[16]的设置；不过，我们进一步提出并开展了更深入的评估，考察人们对人工智能行为和人类行为的判断。

3 背景与初步准备

我们采用了先前研究中的导航任务 [16] , 并在同一款现代3A游戏中开展实验。我们首先更详细地介绍这款游戏，然后概述导航任务。

3.1 电子游戏

为了在我们的研究中重复使用智能体和人类生成的视频，我们选择了与先前研究相同的游戏。这是一款多人在线战斗游戏，包含13个可自定义角色，每个角色都拥有特殊能力。这款游戏常被拿来与其他热门团队动作游戏比较，例如《守望先锋》和DotA。玩家分为两支队伍，每队四人，相互对抗。游戏设有两种游戏模式。一种模式要求玩家占领并守卫地图上的特定地点（称为目标）；另一种模式则要求玩家收集名为细胞的物品，并将其存放到地图上的激活平台中。游戏的团队机制、目标平衡和角色自定义带来了独特的多人游戏体验，因此非常适合用于研究人工智能行为和人机交互。

游戏的核心机制是目标导向导航：玩家必须从一个地点移动到另一个地点，以收集强化道具或细胞，在投放平台处于激活状态时前往那里，并与其他玩家进行战斗。因此，地点之间的导航代表了游戏中最常见任务的一种抽象。为了使我们能够专注于导航特有的特征，我们采用了游戏的简化版本，排除了其他复杂机制和目标。

3.2 导航任务

我们采用与以往研究相同的方式构建导航任务：一个游戏角色必须导航至目标位置。图1左侧的截图展示了该位置，由三个蓝色容器标出。前往目标是主游戏中的一项子任务，玩家必须在导航至目标位置以收集细胞或提升生命值的同时，设法抵御其他玩家。

玩家移动之前，导航目标会随机出现在16个可能位置中的一个，右侧图1中的绿色十字标记了这些位置。随后，玩家会在主地图之外的一座岛屿上生成（显示在小地图的下方区域），并必须利用可用的跳跃区域跳入地图的主区域。进入中央区域后，玩家便可移动至目标位置。

HNTT 要求人类评判者从两种导航行为中，判断哪一种更接近人们在现实中的导航方式。这种表述旨在体现某个智能体相较于另一智能体有多令人信服 [48], ，而这不同于图灵测试的另一种解读：判断一个实体是由人类还是人工智能智能体控制的。我们选择这种表述，是因为我们希望创造令人信服的非玩家角色，为玩家带来沉浸式的游戏体验。相反，我们并不希望欺骗玩家，让他们误以为某个智能体由真人控制。

3.3 基线智能体

先前的研究 [16] 使用两种智能体类型开展了研究：一种符号型智能体和一种混合型智能体。当参与者看到这两种智能体时，他们识别人类玩家的准确率高于随机水平，这意味着人们并不认为它们的行为足够像人类。因此，我们在实验中将这些智能体用作基线，下面将介绍其主要细节。

为了朝目标位置前进，智能体会从预先设定的一组动作（称为动作空间）中选择动作。该动作空间包含 8 种可能的动作：不做任何操作、向前移动，以及向左或向右移动（分别为 30 、 45 和 90 度）。为便于训练，智能体会收到稠密奖励信号，以鼓励其成功导航至目标。该信号由以下几项组成：每步 -0.01 的惩罚，以鼓励智能体高效抵达目标；因智能体可能跌落地图而死亡时一次性扣除 -1；接近目标时获得增量奖励；以及抵达目标时获得 +1 奖励。我们发现，该奖励信号仅包含鼓励智能体尽快成功抵达目标的各项内容。

这两种智能体的主要区别在于它们作为输入的观测内容。符号型 智能体仅接收语义化的低维表示作为输入；混合型 智能体还会接收图像输入。有关基线智能体的更多详情，感兴趣的读者可参阅附录A.1和Devlin等人 [16]。

4 构建一个更像人类的人工智能

为了帮助设计我们的奖励塑形智能体，我们分析混合型和符号型智能体，以找出可能影响先前类人性评判的特征。基于这项分析，我们为 HNTT 引入一种新型智能体：奖励塑形智能体。

4.1 设计我们的奖励塑形智能体

该智能体在混合型 智能体的基础上进行了两项关键改动，以促进类人行为的学习。具体而言，我们在奖励信号中引入了额外项，并扩展了智能体可用的动作空间。为检验我们的改进是否会导致人们对类人性的感知产生差异，我们将奖励塑形 智能体的所有其他组件固定为与混合型智能体相同。

由于符号型 和混合型 智能体此前表现出非类人行为，我们检查了它们生成的导航示例，并归纳出三类问题行为。智能体会：

P1. 大幅摆动摄像机角度或突然转向，**P2.** 频繁撞上墙壁，以及**P3.** 有时移动得比预期更慢。

为纠正这些行为，我们采用奖励塑形 [63] ，加入与期望或不期望行为对应的项。我们引入以下各项。首先，为应对**P1**，对于超过设定的 0.15 差异阈值的快速摄像机角度变化，我们加入摄像机角度差异惩罚。其次，为解决**P2**，每次发生墙壁碰撞时，我们施加 -0.05 的惩罚。第三，为应对**P3**，如果两步之间的移动距离低于特定环境的 220 个地图单位阈值，我们施加 -0.01 的惩罚。我们根据先前的训练奖励以及专家对各组成部分相对重要性的评估来选取这些数值。

为了鼓励更平滑的控制并避免突然转向，我们采用了类似于动作空间塑形的方法 [37] ，通过为智能体引入额外的可用动作来实现这一点。直观而言，我们预计引入更精细的控制将带来更流畅的导航。我们将动作空间从原先的 8 个动作扩展到 14 个。除了“不执行任何操作”和“向前移动”这两个动作外，我们还加入了向左和向右转动的 6 种角度，而基线方法使用的是 3 种。该智能体更新后的转向角度列表为：左右两侧各为 18、36、45、54、72 和 90 度。

这两个组成部分共同构成了奖励塑形智能体的新颖之处。我们以相对不依赖特定算法的方式设计该智能体，让缺乏深度强化学习专业知识的人也能更容易使用。因此，这两个组成部分可以应用于任何最先进的深度强化学习算法。根据底层算法的不同，具体数值，尤其是奖励信号中各项所使用的数值，可能需要进行不同设置。不过，我们认为，调整这些数值比指定针对特定算法的复杂参数更直观。

4.2 生成高质量导航

我们对所有智能体进行训练，使其在导航任务上达到相近的表现水平（训练设置的详情见附录 A.2），以确保任务技能不会造成类人性方面的感知差异。我们通过到达目标所需的步数来衡量任务熟练度。每一步相当于约 5 秒的实时游戏时间。图 2 证实，这些智能体模型确实代表了在复杂三维游戏中学习导航的最先进技术。

在训练过程中，奖励塑形 智能体和混合型 智能体的方差高于符号型 智能体。由于这些智能体还必须从像素中学习，其学习任务比符号型 智能体（它只接收符号输入）更具挑战性。因此，我们预计智能体在学习这一更复杂任务时，训练过程中的方差会更高。重要的是，所有智能体都学会了可靠地抵达目标，这一点从训练接近结束时的表现可以看出。如今，一个熟练的智能体大约需要 60 步完成任务（相当于实时游戏中的约 12 秒）。这一结果确保了

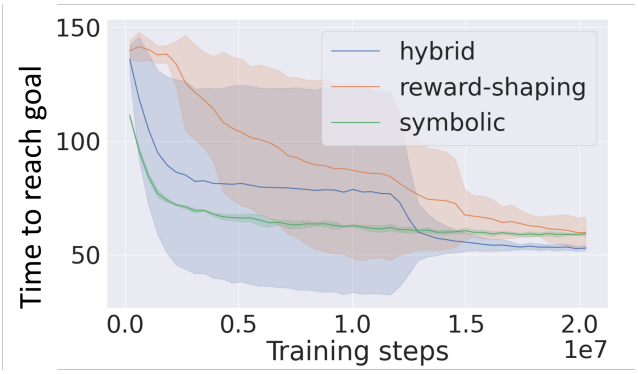


图 2：混合型、符号型和奖励塑形智能体均成功学会了导航。该图展示了解决任务所需的平均时间（y 轴）随智能体训练时长的变化情况。阴影区域表示标准差。对于奖励塑形，N=3；对于混合型和符号型，N=4）。所有曲线均采用 200 步的滚动窗口进行平滑处理。重要的是，平均而言，所有智能体都逐渐收敛到约 60 步解决任务（约相当于游戏内时间的 12 秒）。相比之下，智能体刚开始时平均需要约 140 步（约相当于游戏内时间的 28 秒）才能解决任务。智能体在该任务上的初始表现，与采取随机动作的智能体解决任务所需的时间相近。主要结论是，表现差异并不能解释人们感知到的类人性差异，而任务表现的标准指标不足以评估类人性。

对类人性的评估存在差异，并非源于智能体完成任务的能力不同。

5 实验设计

为了了解人们认为哪些特征能够体现类人性，我们与人类参与者开展了一项行为研究。我们的设置与先前的工作 [16]密切相关；不过，我们也进行了重要扩展，包括通过众包平台（MTurk）收集更多参与者的评估，并获取额外数据以进行更全面的分析。为确保内容完整，我们在此详细介绍完整的研究设计。

5.1 实验任务

我们要求每位人类参与者通过完成一份包含 6 个 HNTT 试次的问卷来担任评判者。在每个 HNTT 试次中，评判者会看到两段并排播放的视频刺激材料，内容是人或智能体完成导航任务。观看这些视频后，评判者需要回答三个问题：他们认为哪段视频中的导航行为更像现实世界中的人类、对其回答的理由，以及他们对该判断的确定程度。具体而言，参与者回答了以下问题：

（1）**哪个视频中的导航方式更像人类在现实世界中的导航方式？**评判者点击了下方的按钮

研究	数量 参与者	数量 试次
人类 vs.混合型	50	6
人类 vs.符号型	50	6
人类 vs. 奖励塑形	92	6

表1：各项研究测试的条件，以及每种条件下的试次数量。值得注意的是，人类与混合型以及人类与符号型的研究是对先前工作的重复实验 [16]，旨在验证改用众包平台的效果。

他们认为哪段视频中的导航行为更像人类。此决定为强制二选一。(2) **你认为为什么会这样？请提供与该视频相关的具体细节。**评判者在问题下方的文本框中以自由形式响应回答了这个问题。(3) **你对自己的选择有多确定？**评判者通过五点李克特量表回答了这个问题，选项从极其确定到极其不确定。

为减少连续观看多个视频所产生的受试者学习效应，我们没有向评判者透露哪些视频是人工智能生成的。换句话说，参与者完成每项任务后，最终并不知道哪些视频是人类生成的。

5.2 实验流程

我们完成了3项研究；每项研究都将一个人类控制的智能体与不同的人工智能智能体进行比较。在每项研究中，所有评判者都观看了相同的6个试次。每位评判者看到的试次顺序均为随机排列。每个试次中，两段视频的播放顺序也经过随机处理，因此无法根据播放顺序推断哪段视频是人类生成的。表1概述了每项研究所测试的条件。

每位参与者首先阅读介绍页面，其中包含任务所需的说明（完整文本见附录B）。随后，他们填写知情同意书，并阅读背景页面，了解有关该电子游戏的简要信息。他们回答了一系列问题，以评估自己对任务的理解程度以及对电子游戏的熟悉度。最后，参与者完成了6个HNTT试次。图3展示了理解度和熟悉度问题的截图（a），以及一个HNTT试次示例（b）。

5.3 导航视频生成与采样

研究的一个关键部分是向人类评判者展示的视频。对于人类生成的导航数据和视频，我们使用了先前研究发表的公开样本 [16].[1] 我们从其按照“研究1”方案发布的 40 个样本中抽取了人类视频。为了生成 AI 导航数据，我们选取每个智能体最近保存的版本。然后，我们新建一个会话，并在游戏中部署该智能体 100 次，共生成 100 个

[1]数据依据 MSR-LA 许可证使用。许可证详情见原作者的 GitHub：https://github.com/microsoft/NTT。此链接包含本研究中使用的所有数据，包括来自我们的奖励塑形智能体的数据。

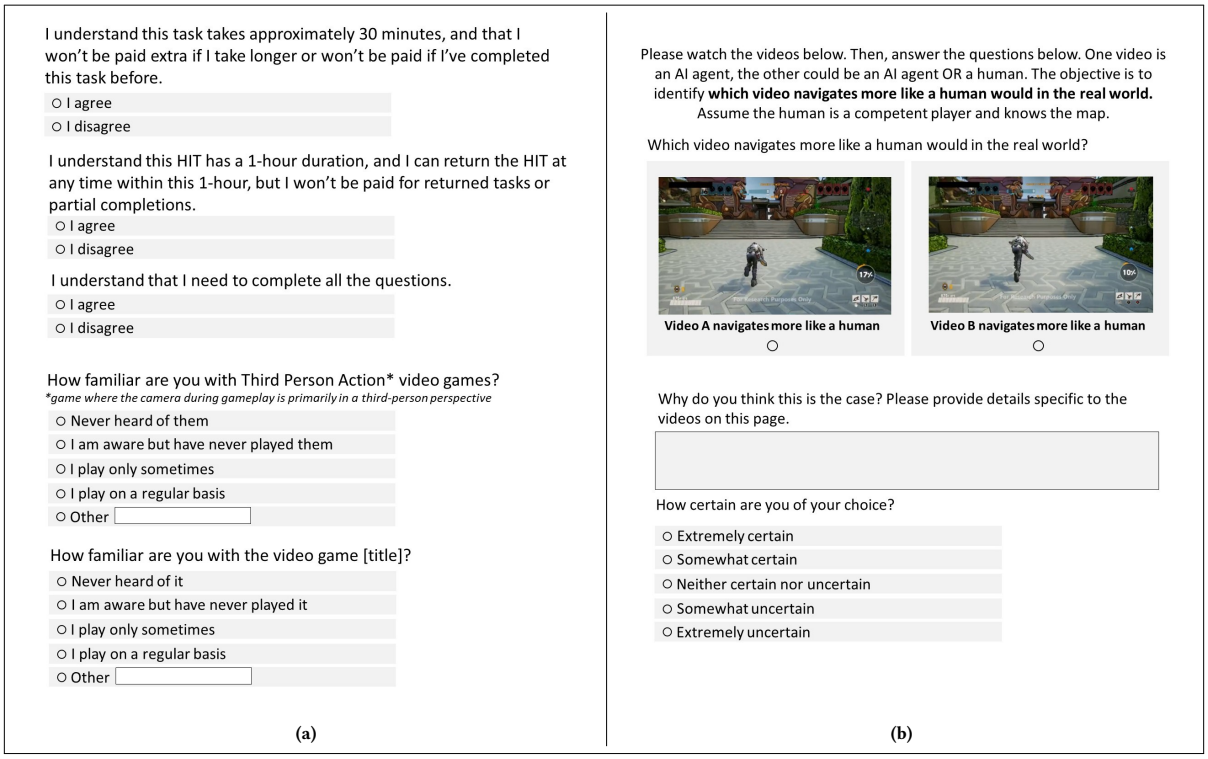


图3：HNTT调查问题的截图。(a)中的截图显示了理解度和熟悉度问题（每位参与者只需回答一次）。我们评估参与者对任务所需时间、对任务完成要求的理解程度、对第三人称动作电子游戏的熟悉度，以及对问卷中所用电子游戏的熟悉度。(b)中的截图展示了一个HNTT试次。我们请参与者回答人类相似性问题、说明理由，并指出自己有多确定。

个智能体导航视频。为了生成研究中使用的视频刺激材料，我们从录制内容中进行均匀随机抽样。

我们采取了多项措施来统一视频，并确保所有条件都受到相同程度的测量噪声影响。首先，我们检查了各条件下光照变化是否一致。其次，我们设计了刺激材料的呈现时长，以确保参与者在所有试验中都有充足时间观看并给出有意义的回答。因此，我们没有使用过长的视频，并排除了短于10秒的视频（后处理前），因为先导研究表明，这些视频太短，无法评估导航质量。第三，由于目标位置可能因游戏控制的初始化过程而异，我们将人类视频中的目标位置与人工智能智能体视频中的目标位置进行了匹配。因此，不同研究中使用了不同的人类视频。第四，我们采用了先前研究 [16]中的后处理步骤，包括遮盖身份识别信息、添加"仅供研究用途"水印，以及剪掉人类视频末尾的几秒。我们进行最后一项修改，是为了纠正数据收集过程造成的影响：人类玩家手动结束录制，导致视频末尾多出几秒。

5.4 其他实验控制

MTurk众包平台 [60] 因具有可扩展性而被广泛用于数据收集和研究，前提是研究人员采取适当的质量控制措施 [32]。在此，我们详细说明为进行质量控制而设定的研究纳入标准。

我们为问卷参与者设置了以下 MTurk 要求：所在地为美国，年龄为 18 岁或以上，语言为英语。我们没有收集人口统计信息或任何其他个人身份信息。为了招募经验更丰富的 MTurk 工作人员，我们设置了以下人类智能任务 (HIT) 资格条件：HIT批准率高于 98%、已批准HIT数量超过 500，并要求具备一项防止重复作答的资格。为鼓励参与者提供高质量回答，我们为每份高质量回答提供额外报酬。我们审查了开放式回答，以找出低质量回答或疑似由机器人提交的回答；例如，对于错别字过多、答案为复制粘贴内容或措辞毫无意义的回答，我们将其排除在分析之外。所有完成该 HIT 的参与者均获得了报酬，即使其回答被判定为低质量也不例外。低质量回答不享有额外报酬。我们支付的平均时薪为 15 美元。我们的研究已获机构审查委员会 (IRB) 批准，并取得了每位参与者的知情同意。

我们在知情同意书中列出了研究详情，并说明了参与者可能面临的风险。

6 分析

我们的主要目标是运用定量和定性测量来评估智能体的类人性。为量化人类评判者区分类人和非类人智能体的能力，我们分析其准确率得分和自我报告的不确定性。为确定影响其对类人性看法的因素，我们采用定性方法。我们构建并使用编码来概括开放式回答中提及的原因，并比较这些编码在不同情境下出现的频率。

6.1 评估类人性

我们首先旨在确定哪些智能体符合我们提出的判定标准，从而通过 HNTT。由于现有研究表明，评估能力会因专业知识而异，我们希望确定这一现象是否也适用于我们的研究情境。最后，我们希望探究评估智能体时自我报告的不确定性与准确性之间的关系。我们提出以下研究问题：

研究问题1. 哪些智能体会被评判为具有人类特征？
研究问题2. 评判者在评估人类特征时，其准确性是否会随着他们对游戏的经验而提高？**研究问题3.** 人类评判者的准确性与其自我报告的不确定性之间有何关系？

为回答**研究问题1**，我们提出一项明确的判定标准，以确定智能体是否足够具有人类特征，并将问题形式化为：人类评估者是否无法区分智能体行为与人类行为？我们将这一标准实现为一项统计检验，用于判定人类评判者区分人类行为与智能体行为的能力是否显著异于随机水平。具体而言，我们通过自助抽样（一种非参数方法）计算人类与智能体比较结果中位数的95%置信区间，以实施这项检验。如果95%置信区间包含0.5（随机水平的一致性），则该智能体通过HNTT。

对于**研究问题2** 和**研究问题3**，我们将感兴趣的变量与准确性进行比较。我们将准确性定义为参与者能够识别出人类生成的行为比人工智能生成的行为更具人类特征。为回答**研究问题2**，我们将准确性与参与者自我报告的对动作游戏总体及研究所用特定游戏的熟悉度进行比较。为回答**研究问题3**，我们考察评判者自我报告的不确定性及其与准确性的关系。

6.2 评估类人性特征

为分析与类人性评估相对应的特征，我们提出了以下研究问题：

研究问题4。 人们在描述类人行为和非类人行为时，是否存在关键差异？智能体是否通过 HNTT，会不会影响这种差异？

RQ5. 人们用于评估类人性的特征与其准确评估类人性的能力之间存在何种关系？

我们从混合型 智能体和奖励塑形 智能体研究中选取了部分回答进行分析。选择这两项研究，是为了比较未通过 HNTT 的智能体与通过 HNTT 的智能体（见第 7.1 节）。首先，我们随机抽取了 55 条回答来计算初始一致性，并将其称为一致性样本。移除含义模糊或无法归入任何编码的回答后，该样本缩减至 53 条。随后，我们通过随机抽样，为每项研究中的每位评审者抽取三条自由形式回答，构建分析样本。为尽量减少偏差，我们在抽样前打乱了回答顺序。我们移除了含义模糊或无法归入任何编码的回答，最终得到一个包含 395 条回答的数据集，用于分析类人特征。

我们采用成对编码方法对数据进行标注。对数据更熟悉的标注员根据既有研究提出了初始编码列表 [91]。按照既定记法 [4],，我们有一组 I 个项目（或回答），由 $C = 2$ 名编码员将其标记为至少一个 K 类别。我们将每个标签分解为较像人类或较不像人类 $H =$（更多、更少），并在适用时量化其方向 $D =$（更多、更少）。例如，如果我们将项目 i 标记为动作平滑度，我们会记录评判者是否认为该行为具有类人性，以及他们是否认为其动作更加 + 或不那么 – 流畅。

随后，两名标注员共同讨论编码的含义，并对一组5个回答进行联合编码。表2展示了一个编码回答的示例。之后，两名标注员分别使用初始编码对一致性样本进行编码。标注员也可以选择将回答标记为其他，并提供具体示例，以便在出现其他主题时修订编码。两名标注员反复会面，讨论分歧并完善编码。经过多轮讨论、独立编码和分歧解决后，标注员确定了用于标记完整样本的编码集合（表3）及其纳入标准。

由于我们的目标是设计类人 AI 智能体，因此我们希望找出 AI 设计者可采用的编码。因此，在确定编码时，我们优先选择指代具体行为的编码，而非更宽泛的编码。例如，碰撞规避行为可以编码为目标导向；但我们只将其编码为碰撞规避。这一流程在优先考虑可用于设计智能体的具体、较低层级行为的同时，也促进了类别的独立性。编码时，标注员首先考虑某项响应是否可归入较低层级的编码，必要时再转向更宽泛的编码。附录 C 对这一流程作了更详细的说明。

标注员在一致性样本上取得了0.84的总体平均标注者间一致性。我们使用二元Cohen's kappa κ [13] over K, D和 H计算标注者间一致性，具体定义如前所述。更多详情请参见表4。确定编码列表后，标注员对数据子样本进行了划分，使其中25%的数据（99个条目）存在重叠。我们在表4中报告了重叠样本的Cohen's kappa系数，以确保

响应	自由形式响应	更像人类	较不像人类
B	视频B中的角色跑步时沿直线前 他 视频 A 中的角色跑步时沿 绕 等。	动作平滑度 +; 目标导向 +	碰撞规避 –; 目标导向 –

表2：对"哪个视频在现实世界中的导航方式更像人类？"这一问题的编码回答示例。最左侧一栏表明，评判者认为视频B展现出更像人类的行为。高亮文本展示了标注过程。评判者指出，更像人类的角色会沿直线移动（更像人类的标注代码：动作平滑度 +），并朝目标前进（更像人类的标注代码：目标导向 +）；而他们认为较不像人类的角色会绕圈（较不像人类的标注代码：目标导向 –），并撞上物体（较不像人类的标注代码：碰撞规避 –）。

标注代码	简写	定义	关键词和短语	示例片段
.....的流畅度 移动	流畅	智能体导航的质量 或摄像机运动	流畅、卡顿、直线， 急转、平稳、流畅	运动非常 更加流畅
目标导向	goal	智能体行为的目标导向程度 似乎	意图、专注， 知道该往哪里走	刻意的摄像机运动 运动
碰撞 规避	规避	智能体是否规避碰撞	碰撞、规避、撞击 撞上障碍物	撞上箱子
环境 感知性	感知性	智能体是否理解和/或 与环境进行恰当的互动	探索、沿着路径前进， 收集强化道具	忽略所有 生命值/法力值/等
直觉	直觉	评判者无法准确指出具体行为	自然、感觉， 似乎是	只是一种感觉
自我指涉	自我指涉	与评判者自身的关系 移动或游戏	像我一样玩	[像]我一样移动 用那种.....视角

表3：标注代码定义。用于标注自由形式回答的编码列于最左侧。中左列显示了这些编码对应的简写，后文将使用这些简写。中间列简要定义了每个编码。中右列列出了标注员用来判断某个回答是否包含特定编码的关键词和短语。最右侧列提供了一个回答片段示例，该片段将被标注为包含相应编码。尽管所列示例相当清晰，自由形式回答通常包含更多模棱两可的内容。

我们对编码的理解没有过度拟合一致性样本中的具体示例。

我们对标注代码和纳入标准进行了更详细的讨论。表3列出了这些定义，以及帮助我们判断每个代码是否适用的短语。对于每个代码，我们都提供了一个示例，帮助读者了解常见回答可能是什么样的。动作平滑度指智能体导航或摄像机运动的质量。该代码既考虑即时的顿挫动作，也考虑随时间延续的之字形运动行为。目标导向指智能体行为看起来有多么有意图。我们会纳入涉及某种感知目标的行为描述，即便该目标并非主要目标。我们纳入碰撞规避这一代码，因为它是机器人学界长期以来的研究领域 [67]。该代码指有意改变方向以避开潜在碰撞的行为。环境响应性旨在捕捉智能体与游戏环境之间的关系、对情境的理解以及对规范的遵循。在现实场景中，这可能表现为某人沿着小路行走而不是踩草地，或在行人信号灯允许时过马路。任何提及某种非特定感觉、认为某种行为更像人类的回答，都归为直觉。我们纳入这一代码，以捕捉参与者能够辨认出自己认为更像人类的行为、却难以表达出来的情况。

最终，我们将自我指涉纳入代码，用于捕捉评判者将智能体行为与自身游戏经历联系起来的情况。

在迭代编码过程中，标注员评估了产生分歧的可能原因。解决错误及其他容易处理的问题后，标注员认定，剩余分歧源于个人在解读含糊的自然语言回答时存在差异。这意味着，在解决分歧时，不能将任何一位标注员视为更正确的一方。因此，标注员决定采用以下分歧解决方案：当两位标注员对某个条目的至少一个标签存在分歧时，我们随机选择一位标注员作为正确方，并采用其标签。

7 结果

我们首先呈现第6.1节所述分析的结果；具体而言，我们证明奖励塑形智能体通过了 HNTT，而其他智能体则未能通过。随后，我们通过突出人类评判者感知人工智能与人类玩家时的典型行为和关键差异，呈现第6.1节所述分析的结果。我们发现，人们倾向于利用

标注编码	方向	Cohen 的 κ 一致性样本	Cohen 的 κ 重叠样本
流畅度	更多 +	0.90	0.79
移动	更少 -	0.64	0.64
目标导向	更多 +	0.82	0.78
	更少 -	0.82	0.63
碰撞规避	更多 +	1.00	1.00
	更少 -	0.64	0.73
环境感知性	更多 +	0.82	0.93
	较少 -	0.73	0.67
直觉		1.00	0.87
自我指涉		1.00	1.00
	平均值	**0.84**	**0.78**

表4：各编码的 Cohen's κ 分数。对于一致性样本中的所有编码，两位标注员的 Cohen's κ 分数平均为0.84 。按照 Cohen 提出的解释标准，我们在每个类别上都达到了至少中等一致性，并且在标注一致性样本时，有7 个类别（共10 个）达到了近乎完全一致。标注重叠样本时，两位标注员在所有编码上的 Cohen's κ 分数平均为0.78。按照 Cohen 提出的解释标准，我们在每个类别上都达到了至少较高的一致性。这两种设置下的总体一致性仅略有下降，表明我们的编码过程具有较强的通用性。

智能体	准确率中位数（四分位距） [95% 置信区间]
符号型	0.83（0.67 − 1.00） [0.67, 1.00]
混合型	0.83（0.67 − 1.00） [0.83, 1.00]
奖励塑形	0。50（0.33 − 0.67） [0.50, 0.50]
智能体	不确定性中位数（四分位距）
符号型	2.17 (1.67 − 2.42)
混合型	1。92 (1.33 − 2.25)
奖励塑形	2.17 (1.75 − 2.67)

表5：准确性和不确定性的完整汇总统计。我们展示每个智能体的准确率中位数（四分位距=Q1-Q3），将其作为集中趋势和离散程度的非参数测量；并报告每个智能体的人类与智能体比较的95%置信区间和不确定性中位数（四分位距=Q1-Q3）。根据我们提出的指标，只有奖励塑形 智能体通过了 HNTT。

在描述类人行为时，他们对高层次特征的看法相似。然而，他们对人工智能能力的看法可能会影响他们认为人工智能智能体在多大程度上表现出这些特征。

7.1 类人性分析

只有 奖励塑形 智能体通过了 HNTT。 表5展示了完整的汇总统计数据，这些数据是基于完整数据集计算得出的

问卷。每次自助法计算均运行 10000 次迭代。根据我们的判定标准，符号型 和混合型 基线智能体未通过 HNTT。评判者的准确性中位数为 0.83（符号型 智能体，95%置信区间=[0.67, 1.0]）和 0.83（混合型 智能体，95%置信区间=[0.83, 1.0]），这表明他们区分智能体与人类的准确性显著高于随机水平。相比之下，我们的奖励塑形 智能体通过了这项类人性测试：准确率中位数的 95% 置信区间包含 0.5（随机水平的一致性）。这一结果表明，评判者无法稳定地区分奖励塑形 智能体与人类玩家（奖励塑形 智能体，准确率中位数=0.50, 95 =[0 50, 0 50] % CI . .）。

由于各试次的样本量不同（人类与混合型 条件、人类与符号型 条件各有 50 个样本；人类与奖励塑形 条件有 92 个样本），我们通过将奖励塑形 智能体的数据子抽样至 50 个样本来验证结果，然后运行自助抽样程序 100 次。我们发现，每次自助抽样中计算出的置信区间（CI）始终包含 0.5，即随机水平的一致性。平均准确率中位数为 0.50，方差为 0.00；平均 CI 为 [0.44, 0.63], ，其下限方差为 0.01，上限方差为 0.00。因此，我们回答**研究问题1**：根据这一拟议指标，奖励塑形 智能体是唯一被判定为类人的智能体。

游戏熟悉度与准确评估人工智能智能体类人性的能力之间不存在关联。 对于每项研究，我们都进行多元线性回归分析，以检验特定游戏熟悉度和一般游戏熟悉度是否能显著预测评估类人性的准确性。对于所有智能体，自我报告的熟悉度与准确性均不存在关联。对于符号型 智能体，拟合的回归模型为：

$$\text{accuracy} = 0.68 - 0.01(\text{specific game familiarity}) + 0.03(\text{general game familiarity}).$$

整体回归模型在统计学上不显著（$R^2 = 0.01$, $F\ 2, 47$ $()$ $= 0.21$, $p = 0.814$）。进一步分解结果后发现，特定游戏熟悉度（$\beta = -0.01$, $p = 0.878$）和一般游戏熟悉度（$\beta = 0.03, p = 0.525$）均无法预测准确性。

对于混合型 智能体，拟合的回归模型为：

$$\text{accuracy} = 0.67 - 0.03(\text{specific game familiarity}) + 0.06(\text{general game familiarity}).$$

整体回归模型在统计学上不显著（$R^2 = 0.06$, $F\ 2, 47$ $()$ $= 0.26$, $p = 0.261$）。我们发现，特定游戏熟悉度无法显著预测准确性（$\beta = -0.03, p = 0.377$）。一般游戏熟悉度也无法显著预测准确性（$\beta = 0.06$, $p = 0.109$）。

接下来关注奖励塑形 智能体，拟合的回归模型为：

$$\text{accuracy} = 0.40 - 0.02(\text{specific game familiarity}) + 0.04(\text{general game familiarity}).$$

整体回归再次未达到统计显著性（ $R^2 =$ $0.02\ F\ 2, 89\ = 0.94\ p\ = 0.393$ $()$, , ）。这一结果对特定游戏熟悉度（$\beta = -0.02$, $p = 0.522$）和一般游戏熟悉度（$\beta = 0.04$, $p = 0.191$）均成立。

这些发现表明，游戏熟悉度通常不能预测这项特定任务的准确性，从而回答了**研究问题2**。相比之下，先前研究已表明，评估类人性的能力与对研究领域的熟悉度之间存在关联。我们认为，这一结果之所以不同，是因为我们研究的是一个相对简单的场景，大多数人对什么构成类人性都有明确的先验认知。通过步行或跑步来导航是大多数人每天都会进行或观察到的活动，这意味着即使我们不熟悉需要导航的游戏，也很可能对类人运动有着强烈的内在感知。相比之下，我们假设，在完整游戏场景中，游戏熟悉度能够预测准确性，这凸显了在更复杂的场景中评估类人性的重要性，并应将其作为下一步研究重点。

人类评判者在评估奖励塑形**智能体时表现出较少的虚假信心。**我们评估参与者不确定性的中位数；数值越低表示越确定，数值越高表示越不确定。表5展示了这项分析的结果。参与者在评估我们的奖励塑形智能体（中位数=2.17，四分位距=(1.75-2.67)）和符号智能体（中位数=2.17，四分位距=(1.67- 2.42)）时，报告的不确定程度相近。相比之下，参与者在评估混合型智能体时报告了更高的确定性（中位数=1.92，四分位距=(1.33-2.25)）。

与混合型智能体相比，人们对自己评估符号型和奖励塑形智能体的结果更没有把握；然而，在符号型和混合型智能体存在的情况下，参与者更准确地检测出了人类生成的行为。这一结果令人意外，因为它表明，自我报告的不确定性与准确评估并不一定相关。换言之，参与者可能会对自己评估智能体类人性的能力表现出虚假信心。我们认为，参与者对符号型智能体的评估可能不那么确定，是因为视频时长存在差异：符号智能体的视频平均时长为8.3秒，而混合型智能体的视频时长为15.3秒。参与者可能没有足够的时间了解智能体，因而无法准确评估它。综合来看，准确性和不确定性结果表明，当参与者看到奖励塑形智能体的行为时，与看到混合型智能体生成的行为时相比，他们对自身评估能力表现出的虚假信心较少。这一结果回答了**研究问题3**。

7.2 类人性分析

在所有图表中，我们使用表3中注明的编码简写版本，以及 + 和 – 记号。 + 和 – 记号表示编码的程度或方向。例如，smooth+ 表示参与者提到了更多的平滑运动，而 smooth – 表示参与者提到了更少的平滑运动。

人类评判者在评估类人行为时，依赖相似的高层次特征。图4展示了参与者用来描述类人和非类人行为的编码。我们考察某项编码被使用的次数占

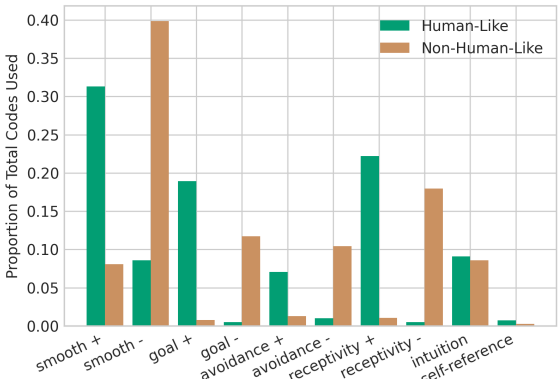
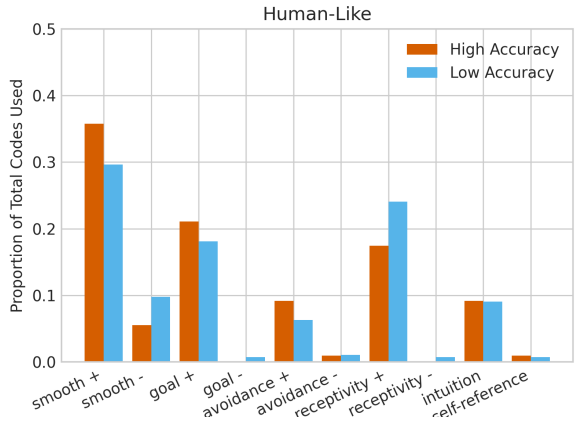


图4：用于描述类人行为和非类人行为的编码。我们比较了人类评判者在评估类人性时，用于描述类人行为和非类人行为的编码比例。人们更常将类人行为描述为更加平滑、更具接受性、对环境更具响应性，并且更具目标导向性。相比之下，参与者更常将非类人行为描述为不那么平滑、接受性较低、对环境的响应性较弱，并且目标导向性较低。

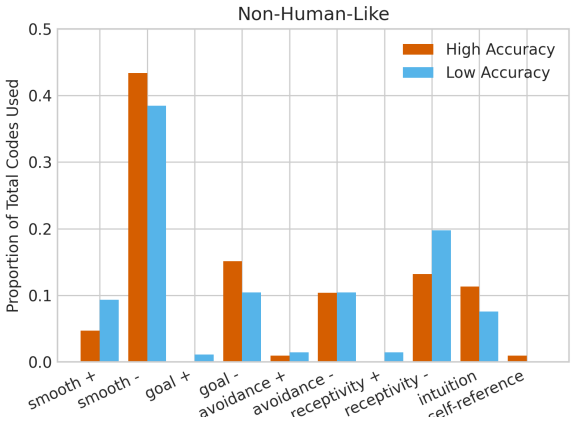
描述类人或非类人行为所用全部编码的比例（类人和非类人编码比例之和应为1）。评判者在描述类人行为时，往往依赖相似的高层次特征。总体而言，他们最常提及以下高层次编码：动作平滑度、环境响应性和目标导向性。根据行为是否被评定为类人，我们对回答进行分类后发现，人们更常将类人行为描述为动作更流畅、对环境更敏感且更能响应环境，以及更具目标导向性。相比之下，参与者更常将非类人行为描述为动作不够流畅、对环境不够敏感且响应不足，以及目标导向性较弱。描述类人和非类人行为时，他们对直觉和自我指涉的依赖程度相近。

我们根据智能体类型对这些回答进行了研究，但没有发现所得编码比例之间存在差异。这一结果支持了这样一种观点：人们对于什么构成类人行为，可能具有相对稳定的信念。因此，这一推理有时才有用：换句话说，只有当人工智能没有被设计得比人更不生硬时，寻找动作更生硬的智能体才有意义。因此，我们得出结论：尽管人们依赖不同的具体特征来判断类人性，但这些总体特征在不同人工智能智能体之间相对稳定，这回答了**研究问题4**。

更准确地评估类人性的评判者，对特征的看法与评估准确性较低的人类评判者不同，尽管他们对高层次特征的依赖程度相近。我们将参与者分为两组：高准确性



(a) 低准确度和高准确度评判者用于描述类人行为的编码。我们比较人类评判者在评估类人性时，用于描述类人行为的各类编码所占比例。



(b) 低准确度和高准确度评判者用于描述非类人行为的编码。我们比较人类评判者在评估类人性时，用于描述非类人行为的各类编码所占比例。

图 5：用于描述类人行为和非类人行为的编码，进一步按高准确性和低准确度评判者分组。我们比较了用于描述类人行为（左）和非类人行为（右）的编码比例。我们进一步按高准确性和低准确度评判者对这些编码进行细分，以确定更准确的个体是否依赖不同特征来解释其判断。有趣的是，我们发现评判者对相似特征的依赖程度各不相同。

（超过 80% 的回答认为，更像人类的智能体与人类生成的视频相符）和低准确度（不超过 80% 的回答认为，更像人类的智能体与人类生成的视频相符）。我们考察了各组参与者更常使用哪些编码来描述类人性。图 5 展示了这一细分结果。尽管高准确度和低准确度评判者通常依赖相似的特征，但依赖程度有所不同。例如，两类评判者在描述类人行为时，使用的编码中有 40% 涉及动作平滑度这一高层级编码。同样，在描述他们认为不像人类的行为时，两类评判者使用的编码中约有 49 % 涉及动作平滑度这一高层级编码。这些结果表明，他们依赖这一特征来解释行为的倾向并无差异。不过，高准确度评判者在描述类人行为时更常提及动作流畅。相比之下，低准确度评判者更常将平滑度视为非类人行为的特征。这一结果进一步支持了这样一种观点：人们对人工智能能力的看法可能会影响其评估。在我们的研究中，低准确度参与者似乎普遍认为人工智能智能体比人更有能力（能够实现更平滑或更“完美”的导航）。

低准确度评判者更常将类人智能体描述为对周围环境表现出更强的感知性和响应性。对于较低的感知性，也出现了类似的模式，人们会以此作为其不具类人性的理由。这一结果表明，低准确度参与者可能会错误地将某些行为归因于与环境互动。例如，一位错误地认为视频 A 更像人类的评判者声称：

> 视频 A 选择了通往终点更明显的路线，而 B 则选择了可能最长的路线。人类通常会选择最简单的路线。

有趣的是，在评估类人行为时，高准确度评判者和低准确度评判者运用直觉和自我指涉的频率相近。结合此前结果——对类人性的评估会受到关于人工智能能力的刻板印象影响——这一发现表明，有些参与者的直觉更准确，是因为他们的直觉与人工智能智能体的实际能力相符。

8 讨论与未来方向

尽管本研究的范围有限，我们的研究结果仍有助于未来在类人智能体的设计与评估方面开展工作。

8.1 局限性

我们的研究专门评估智能体在第三人称视角下执行点到点导航行为时的类人性。尽管这种导航形式存在于许多场景中，例如驾驶模拟器中的行人导航 [83]，但现实世界和虚拟环境中还存在许多其他形式的导航。每种形式都有其独特的挑战，并需要采用不同策略来设计类人行为。虽然我们的研究并未涵盖所有类型的导航，但它为评估智能体在一种特定导航类型中的类人性提供了宝贵的起点。我们识别出的编码具有足够的普适性，可作为研究人员分析不同导航形式的起点。例如，碰撞规避是许多涉及多种导航类型的领域中普遍存在的一项特征，

例如驾驶和跑步。未来的研究应考虑拓展这些评估，以便更全面地理解如何设计行为更像人类的智能体。

此外，对自由形式回答的分析显示，评判者对人类相似性问题有不同的理解。一些评判者将移动直接与现实世界中的人类导航联系起来。一位评判者说道，

> 在现实生活中，人类几乎肯定不会跳下像视频 A 中的角色那样远，且不会严重伤害自己。

然而，另一些人则将这种移动与人类玩家在电子游戏中会如何操控智能体联系起来。另一位评审提到，

> ……在视频 A 中，玩家短暂地撞上了一堵墙，然后才重新调整方向。这是人们分心并短暂移开视线时会做的事。

为探究这一分歧，我们对一致性样本进行了标注，记录受试者回答问题时采用的是哪种理解：现实世界中的人类导航、电子游戏导航，还是不明确。两位标注员在这项标注上的一致性很高（Cohen's kappa系数：$\kappa = 0.94$）。受试者究竟在回答哪个问题，大多并不明确（53 份回答中有 40 份）。然而，11 份回答涉及电子游戏导航，只有 2 份明确涉及现实世界中的人类导航。我们推测，加入游戏熟悉度问题，会促使受试者认为问题询问的是电子游戏特有的导航，而非一般的类人导航。在未来的研究中，我们建议研究设计者明确告知参与者所问的问题究竟是哪一个：可以额外询问参与者对问题另一种理解下的回答，以形成鲜明对比；也可以描述希望参与者设想自己身处的情境。

8.2 设计和评估类人 AI

我们的研究表明，只有经过设计、能够展现更像人类行为的智能体通过了我们对类人性的测试，这凸显了在设计智能体时明确纳入这些目标的重要性。不过，究竟何为类人性，需要设计者仔细考量。人类评判者对人类相似性问题的不同理解进一步印证了这一点。一种理解是，智能体的表现仿佛由人操控；另一种则是，智能体展现出更逼真的行为。这两种观点在不同情境下都可能有用。

当设计师希望将开发过程中的部分环节自动化（例如游戏测试）时，打造看起来像是由人类控制的智能体就更加重要。在电子游戏的自动化游戏测试中，能够像真实用户一样行动的人工智能智能体 [23, 68]，可以帮助电子游戏设计师加快迭代开发过程，同时减轻游戏玩家全面评估新内容的负担。待移动相关的明显错误得到修正后，用户再提供反馈即可，这或许能提升他们参与反馈过程的乐趣。在共享自主控制 [2]，中，开发出表现得像用户本人的智能体，将有助于更无缝地将半自主控制与用户输入相结合。例如，我们发现评判者特别提到了横向移动，认为这是人们在电子游戏中会采取的行动。横向移动是一种战术性的侧向动作，人们在现实世界中导航时并不会这样做。加入这些游戏特有的移动方式，可能会让智能体看起来更像是由人类控制，尤其是在经验丰富的玩家看来。此类智能体还能让遭遇中断（例如网络问题）的玩家继续玩云游戏 [57]。系统检测到中断时，可以接管控制并开始模仿类人行为。用户可以无缝地重新接管控制权。对于因其他原因（例如行动障碍）而需要游戏内辅助的玩家，也可以将此功能作为一项选项。反过来说，如果目标是提升沉浸感，那么实现更逼真的导航就至关重要。

在我们的研究中，我们专注于实现更逼真的导航。为此，我们确定了一组评判者用来评估类人性的高层次特征，例如动作平滑度。因此，游戏人工智能设计师可以先专注于调整这些特征。正如我们的奖励塑形智能体所示，可以通过简单的技术来实现这些特征，并使用自动化图灵测试 [16]进行评估。在处理了最常被提及的特征后，设计者便可以进一步关注更细致的方面，例如避免智能体走进水坑，以体现更贴近现实世界的导航。

此外，我们采用了一种第三人称图灵测试，参与者通过观看智能体导航的视频来进行测试。尽管暂停、倒回和重放视频提供了一种提问方式，但这种方式完全基于观察，缺少典型图灵测试所采用的基于干预的方法。基于干预的方法可以包括更改摄像机视角、通过对抗性干扰打断人工智能智能体的预定路径等。这些交互形式可能会带来不同的见解。

然而，部署互动性更强的测试也存在一些缺点，尤其是在大规模开展时。招募一位人类评估者和一位人类玩家进行互动，需要他们同时有空以提供实时反馈。一种解决方案是开展线下研究，但这类研究难以扩大规模并加以实施。例如，在本研究进行期间，由于全球大流行病仍在持续，我们无法开展线下研究，也无法向远程参与者分发专有游戏版本。未来的研究可以利用游戏串流技术的进步，从而开展使用专有游戏版本的互动式远程研究。这一解决方案还可以纳入此前关于同步招募参与者的研究成果 [7, 82]。不过，构建能够整合这些不同技术的架构，可能需要投入大量工程工作。

重要的是，先前的研究表明，让玩家和智能体置身于同一虚拟空间中，并采用更直接的方式对智能体进行询问，所能带来的洞见有限 [78]。事实上，纳入游戏内评估来判断对象是人类还是机器人，会产生一个副作用：新增的游戏机制使部分玩家不得不在游戏玩法和可信度评估之间有所取舍 [28]。这种注意力分散会导致结果不可靠，因此其他研究者开始采用这些评估的第三人称变体 [5, 16, 70]。因此，我们认为，以下流程可用于评估类人智能体。设计者可以先采用第三人称图灵测试，评估特定行为的类人性。随后可根据得出的特征，设计一组

智能体，使其表现出不同的行为，而这些行为取决于参与者最普遍的看法。例如，如果参与者认为平滑运动是类人性的特征，智能体就可以移动得更加平滑。玩家随后可以选择自己想在游戏中互动的角色，从而根据自身的主观体验和喜好定制游戏。这种方法或许能在不损害评估过程完整性的前提下，更可靠地洞察智能体设计的成效。它还能让游戏玩家能够掌控自己的体验。

8.3 迈向更通用的类人智能体

尽管为本研究创建的特定智能体可能无法推广到不同的游戏中，但这也是人工智能领域普遍面临的一项开放挑战 [46, 61]。因此，我们提出一些建议，说明如何利用设计师和玩家的反馈（例如，通过用户研究）更高效、更有效地训练类人智能体。

不同技能水平的人类评判者在描述类人性时存在差异，这表明不同的人对何为类人行为有着不同的理解。这支持了这样一种观点：游戏中非玩家角色的可信度很大程度上取决于玩家的先前信念和预期。这一发现与熟悉度 [29]这一基本原则相符，该原则强调用户在现实生活中的个人经验和知识，也凸显了以玩家为中心的 设计和定制化的重要性 [74]。我们不应打造一成不变的类人智能体，而应努力理解玩家的信念，并据此量身定制他们的体验。

转向更复杂的环境时，还会引入额外的困难。对类人性的评估会变得更加主观，并因个人差异和文化因素而有所不同 [50]。这一结果凸显了让不同群体参与人工智能智能体评估的重要性，从而更全面地了解人们如何看待这些智能体。在游戏的语境下，这可以表现为利用参与式设计方法 [65] 让游戏玩家参与人工智能智能体本身的设计。在征得玩家同意后，我们可以使用从人类反馈中学习领域的技术 [10, 31, 90], ，为人们提供更多渠道来表达他们希望人工智能智能体具备的特性。借助这些技术，玩家可以通过以下形式向智能体提供训练数据：对智能体生成的成对演示表达偏好、展示期望的行为等。这有助于确保人工智能智能体的设计能够考虑不同群体的需求和偏好。

这种方法也可以帮助减轻电子游戏设计师的负担：在复杂领域中，手动指定奖励信号往往很有挑战性 [12, 42, 47]。部分原因在于期望行为的复杂性：正如我们所展示的，类人行为具有多面性，需要针对多个目标进行优化。此外，描述任务时，有时也很难准确写下我们的具体意图。例如，我们该如何构造奖励信号，来体现“在电子游戏中建造一栋与周围房屋风格相同的房子？ [69]”这项任务？为这项任务设计奖励信号时，我们需要编码什么样的建筑才算房子、风格中哪些组成部分最需要模仿，以及哪些结构才算房子。哪些组成部分对模仿风格最为重要，以及哪些结构算作房子。人们可以很快理解这条指令的意图，但要明确表达这种隐含的理解却颇具挑战。

因此，未来一个令人振奋的研究方向是开发更有效的向人类学习的技术，评估用户对这些技术的体验，并将其整合到灵活、易用的工具中。该工具还有助于将这项工作拓展到更普遍的游戏场景。为更轻松地推动这一研究方向，可以将类人性的评估纳入常用的游戏引擎，例如 Unity [36]。借助该工具，游戏开发者可以使用先前研究中经过验证的指标和基准，轻松评估其人工智能智能体的类人性。此外，该工具还可以包含预训练的类人人工智能智能体库，供开发者以此为起点开展自己的工作。例如，开发者可以利用预训练的类人导航代理执行导航任务，同时为其他任务开发自己的算法。使用该工具可以帮助开发者节省时间和精力，使他们能够快速、轻松地创建更可信、更具吸引力的智能体，从而提升玩家体验。

9 结论

在这项工作中，我们旨在了解人们在3D电子游戏的导航情境中，如何评估人类和人工智能生成行为的类人性。为此，我们设计并实现了一种新型人工智能智能体，用以生成类人导航行为。我们针对大规模的人类生成导航行为开展了一项研究，比较了三个人工智能智能体，其中包括我们新设计的奖励塑形智能体。研究发现，我们提出的智能体通过了图灵测试，而另外两个智能体则未能通过。我们还进一步考察了人们在评估这些智能体时给出的理由，并发现人们在判断与人类的相似性时，会依据相似的高层次特征。在这一情境下，我们推测，对这些智能体的评估准确性存在差异，更多是源于人们对人工智能系统能力的固定看法，而非他们对游戏这一评估领域的熟悉程度。最后，我们讨论了这项工作的局限性，为电子游戏设计师提出了具体的设计考量，并指出了未来研究中若干关键方向。

通过强调设计考量与挑战，我们希望本文能够呼吁人机交互和人工智能领域的研究者融合各自的视角与技术开展研究。构建更通用的类人智能体，需要仔细设计智能体本身及其评估方案。开发可融入游戏及其他环境的工具，有助于快速迭代这些设计，并整合这些不同的技术。总的来说，我们希望研究者能够开发并评估具有类人行为的智能体，从而改善人与人工智能智能体之间的交互。

致谢

我们感谢 Evelyn Zuniga、Guy Leroy、Mikhail Jacob、Mingfei Sun 和 Dave Bignell 对本项目早期研究的贡献并提出反馈。我们还要感谢 Cecily Morrison、Youngseog Chung 和 Max Meijer 提供的宝贵

意见和反馈。此外，我们感谢 CHI 匿名评审提供的详尽意见；他们的建议显著改进了本文。共同作者 Fang 获得 NSF 资助项目 IIS-2046640（CAREER）的部分支持。

参考文献

[1] Martín Abadi、Ashish Agarwal、Paul Barham、Eugene Brevdo、Zhifeng Chen、Craig Citro、Greg S. Corrado、Andy Davis、Jeffrey Dean、Matthieu Devin、Sanjay Ghemawat、Ian Goodfellow、Andrew Harp、Geoffrey Irving、Michael Isard、Yangqing Jia、Rafal Jozefowicz、Lukasz Kaiser、Manjunath Kudlur、Josh Levenberg、Dandelion Mané、Rajat Monga、Sherry Moore、Derek Murray、Chris Olah、Mike Schuster、Jonathon Shlens、Benoit Steiner、Ilya Sutskever、Kunal Talwar、Paul Tucker、Vincent Vanhoucke、Vijay Vasudevan、Fernanda Viégas、Oriol Vinyals、Pete Warden、Martin Wattenberg、Martin Wicke、Yuan Yu 和 Xiaoqiang Zheng。2015。TensorFlow：异构系统上的大规模机器学习。[2] Peter Aigner 和 Brenan McCarragher。1997。利用势场将人类纳入机器人控制。在机器人与自动化国际会议论文集，第1卷。IEEE，291–296。[3] Eloi Alonso、Ubisoft La Forge、Maxim Peter、David Goumard 和 Joshua Romoff。2020。面向 AAA 电子游戏导航的深度强化学习。在*NeurIPS* 现实世界强化学习挑战研讨会。NeurIPS，加拿大蒙特利尔，1–13。[4] Ron Artstein 和 Massimo Poesio。2008。计算语言学中的编码者间一致性。计算语言学34，4（2008），555–596。[5] Christian Arzate Cruz 和 Jorge Adolfo Ramirez Uresti。2018。HRLB2：一种基于强化学习的可信机器人框架。应用科学8，12（2018），2453。[6] Christopher Berner、Greg Brockman、Brooke Chan、Vicki Cheung、Przemysław Dębiak、Christy Dennison、David Farhi、Quirin Fischer、Shariq Hashme、Chris Hesse 等。2019。利用大规模深度强化学习玩《Dota 2》。*arXiv* 预印本 *arXiv:1912.06680*1（2019），1–66。[7] Michael S Bernstein、Joel Brandt、Robert C Miller 和 David R Karger。2011。两秒内汇集众包：实现实时众包驱动的界面。在第*24*届 *ACM* 用户界面软件与技术年度研讨会论文集。33–42。[8] Rodney A Brooks、Cynthia Breazeal、Robert Irie、Charles C Kemp、Matthew Marjanovic、Brian Scassellati 和 Matthew M Williamson。1998。智能的另类本质。*AAAI/IAAI* 1998（1998），961–968。[9] Micah Carroll、Rohin Shah、Mark K Ho、Tom Griffiths、Sanjit Seshia、Pieter Abbeel 和 Anca Dragan。2019。学习了解人类对人机协同的效用。神经信息处理系统进展32（2019）。[10] Paul F Christiano、Jan Leike、Tom Brown、Miljan Martic、Shane Legg 和 Dario Amodei。2017。根据人类偏好进行深度强化学习。在神经信息处理系统进展。4299–4307。[11] Georgios Christou。2014。电子游戏中沉浸感与吸引力的相互作用。《人类行为中的计算机》32（2014），92–100。[12] Jack Clark 和 Dario Amodei。2016。现实环境中的错误奖励函数。https: //blog.openai.com/faulty-reward-functions [13] Jacob Cohen。1960。名义量表的一致性系数。教育与心理测量20，1（1960），37–46。https://doi.org/10.1177/001316446002000104 10 1177 001316446002000104 [14] 2011 arXiv:https://doi.org/ . / David Conroy、Peta Wyeth 和 Daniel Johnson。 . 面向游戏人工智能设计的类玩家行为建模。在第八届计算机娱乐技术进展国际会议论文集。ACM，纽约，纽约州，1–8。[15] Sara De Freitas。2018。游戏是有效的学习工具吗？教育游戏综述。教育技术与社会期刊21，2（2018），74–84。[16] Sam Devlin、Raluca Georgescu、Ida Momennejad、Jaroslaw Rzepecki、Evelyn Zuniga、Gavin Costello、Guy Leroy、Ali Shaw 和 Katja Hofmann。2021。导航图灵测试（NTT）：学习评估类人导航。在第*38*届国际机器学习会议论文集。PMLR，线上，1–10。[17] Prafulla Dhariwal、Christopher Hesse、Oleg Klimov、Alex Nichol、Matthias Plappert、Alec Radford、John Schulman、Szymon Sidor、Yuhuai Wu 和 Peter Zhokhov。2017。OpenAI 基线。https://github.com/openai/baselines。[18] Hippolyte Dubois、Patrick Le Callet 和 Antoine Coutrot。2021。电子游戏体验中的导航困难可视化。在*2021*年第*13*届多媒体体验质量国际会议（*QoMEX*）。IEEE，77–80。[19] Logan Engstrom、Andrew Ilyas、Shibani Santurkar、Dimitris Tsipras、Firdaus Janoos、Larry Rudolph 和 Aleksander Madry。2019。深度强化学习中的实现至关重要：关于近端策略优化和信赖域策略优化的案例研究。在国际表征学习会议。[20] Nobuto Fujii、Yuichi Sato、Hironori Wakama、Koji Kazai 和 Haruhiro Katayose。2013。评估在生物约束下自主习得的电子游戏智能体的类人行为。在计算机娱乐技术进展国际会议。施普林格，61–76。

[21] Yolanda Gil 和 Bart Selman. 2019. 美国人工智能研究的20年社区路线图. arXiv 预印本 arXiv:1908.02624 (2019). [22] Astrid Glende. 2004. 通过电脑游戏的智能体设计. 载于第42届东南地区年会论文集. 414–415. [23] Stefan Freyr Gudmundsson, Philipp Eisen, Erik Poromaa, Alex Nodet, Sami Pur- monen, Bartlomiej Kozakowski, Richard Meurling 和 Lele Cao. 2018. 使用深度学习进行类人游戏测试. 载于2018 IEEE计算智能与游戏会议 (CIG). IEEE, 1–8.[24] William H Guss, Cayden Codel, Katja Hofmann, Brandon Houghton, Noboru Kuno, Stephanie Milani, Sharada Mohanty, Diego Perez Liebana, Ruslan Salakhut- dinov, Nicholay Topin, 等. 2019. NeurIPS 2019竞赛：利用人类先验进行样本高效强化学习的 MineRL 竞赛. arXiv 预印本 arXiv:1904.10079 (2019). [25] Rotem D Guttman, Jessica Hammer, Erik Harpstead 和 Carol J Smith. 2021. Play for Real (ism)：使用游戏预测现实世界中的人机交互. ACM 人机交互学报 5 (2021), 1–17. [26] Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser 和 Edvard I Moser. 2005. 内嗅皮层中空间地图的微观结构. 《自然》436, 7052 (2005), 801–806. [27] Simon Hecker, Dengxin Dai, Alexander Liniger, Martin Hahner 和 Luc Van Gool. 2020. 使用语义地图和注意力学习准确的类人驾驶. 载于2020 IEEE/RSJ智能机器人与系统国际会议 (IROS). IEEE, 新泽西州皮斯卡塔韦, 2346–2353. [28] Philip Hingston. 2010. 一种用于机器人的新型图灵测试设计. 载于2010 IEEE计算智能与游戏会议论文集. IEEE, 新泽西州皮斯卡塔韦, 345–350. [29] Vita Hinze-Hoare. 2007. 人机交互 (HCI) 原则的回顾与分析. arXiv 预印本 arXiv:0707.3638 (2007). [30] Sean D Holcomb, William K Porter, Shaun V Ault, Guifen Mao 和 Jin Wang. 2018. DeepMind 及其阿尔法围棋零人工智能概述. 载于2018 年大数据与教育国际会议论文集. 67–71. [31] Ahmed Hussein, Mohamed Medhat Gaber, Eyad Elyan 和 Chrisina Jayne. 2017. 模仿学习：学习方法综述. 《ACM计算调查》 (CSUR) 50, 2 (2017), 1–35. [32] Panagiotis G Ipeirotis, Foster Provost 和 Jing Wang. 2010. 亚马逊土耳其机器人上的质量管理. 载于 ACM SIGKDD 人类计算研讨会论文集. ACM, 纽约市, 纽约, 64–67.[33] Athul Paul Jacob, David J Wu, Gabriele Farina, Adam Lerer, Hengyuan Hu, Anton Bakhtin, Jacob Andreas 和 Noam Brown. 2022. 使用 KL正则化搜索建模高水平且类人的游戏玩法. 载于国际机器学习会议. PMLR, 9695–9728. [34] Mikhail Jacob, Sam Devlin 和 Katja Hofmann. 2020. “它很笨重且耗时很长”——在商业游戏中创建智能体的挑战与机遇. 载于 AAAI人工智能与交互式数字娱乐会议论文集，第 16 卷. AAAI出版社, 帕洛阿尔托, 加利福尼亚州, 88–94. [35] Magnus Johansson. 2013. 非玩家角色会梦见电羊吗. 一篇关于玩家、非玩家角色、沉浸感和可信度的论文（斯德哥尔摩大学计算机与系统科学系博士论文）(2013). [36] Arthur Juliani, Vincent-Pierre Berges, Ervin Teng, Andrew Cohen, Jonathan Harper, Chris Elion, Chris Goy, Yuan Gao, Hunter Henry, Marwan Mattar, 等. 2018. Unity：一个面向智能体的通用平台. arXiv 预印本 arXiv:1809.02627 (2018). [37] Anssi Kanervisto, Christian Scheller 和 Ville Hautamäki. 2020. 深度强化学习中的动作空间塑形. 载于2020 IEEE游戏会议 (CoG). IEEE, 479–486. [38] Igor V Karpov, Jacob Schrum 和 Risto Miikkulainen. 2013. 通过回放人类轨迹实现可信机器人导航. 载于《可信的机器人》. 施普林格 , 151–170. [39] Man-Je Kim, Kyung-Joong Kim, Seungjun Kim 和 Anind K. Dey. 2018. 即时战略游戏中人类玩家与人工智能玩家之间的性能评估差距. 《 IEEE Access 》 6 (2018), 13575–13586. https: //doi.org/10 1109 2018 2800016 . /ACCESS. . Diederik P Kingma 和 [40] 2014 Jimmy Ba. . Adam：一种随机优化方法. arXiv 预印本 arXiv:1412.6980 (2014). [41] Raph Koster. 2013. 游戏设计的乐趣理论. " O'Reilly Media, Inc.". [42] Victoria Krakovna. 2018. 人工智能中的规范博弈示例. https://vkrakovna.wordpress.com/2018/04/02/specification-gaming-examples-in-ai/ [43] Guillaume Lample 和 Devendra Singh Chaplot. 2017. 使用深度强化学习玩第一人称射击游戏. 载于第三十一届 AAAI 人工智能会议. [44] Joni Lämsä, Raija Hämäläinen, Mikko Aro, Raine Koskimaa 和 Sanna-Mari Äyrämö. 2018. 用于提升基础阅读和数学技能的游戏：支持学习障碍者学习的教育游戏设计系统综述. 《英国教育技术杂志》49, 4 (2018), 596–607. [45] P Lavega 等. 2004. 用于教育以学习建立纽带的传统游戏. 建立纽带以学习. 《体育文化与旅游研究》11, 1 (2004), 9–32.[46] Kuang-Huei Lee, Ofir Nachum, Mengjiao Yang, Lisa Lee, Daniel Freeman, Winnie Xu, Sergio Guadarrama, Ian Fischer, Eric Jang, Henryk Michalewski, 等. 2022.

多游戏决策变换器. arXiv 预印本 arXiv:2205.15241 (2022). [47] Joel Lehman, Jeff Clune 和 Dusan Misevic. 2018. 数字进化惊人的创造力. 载于人工生命会议论文集. 麻省理工学院出版社, 55–56. [48] Daniel Livingstone. 2006. 图灵测试与游戏中的可信 AI. 《娱乐计算机》 (CIE) 4, 1 (2006), 6–es. [49] Rossella Lorenzi. 2013. 青铜时代土耳其墓葬中出土已知最古老的游戏筹码. NBC 新闻 (2013). https://www.nbcnews.com/sciencemain/oldest-known-gaming-tokens-dug-bronze-age-turkish-graves-6 10920354 c [50] 2004 Brian Mac Namee. . 主动式持久型智能体——使用情境智能在角色中心型电脑游戏中创建辅助角色. (2004). [51] Colt McAnlis 和 James Stewart. 2008. 导航网格生成中的内在细节. 《AI 游戏编程智慧》4 (2008), 95–112. [52] Matheus RF Mendonça, Heder S Bernardino 和 Raul F Neto. 2015. 使用强化学习和人工神经网络模拟格斗游戏中的人类行为. 载于2015 年第14届巴西电脑游戏与数字娱乐研讨会 (SBGames). IEEE, 152–159. [53] Lazaros Michailidis, Emili Balaguer-Ballester 和 Xun He. 2018. 电子游戏中的心流与沉浸感：一项概念挑战的余波. 《心理学前沿》9 (2018), 1682. [54] Maximiliano Miranda, Antonio A Sánchez-Ruiz 和 Federico Peinado. 2016. 一种模仿《吃豆小姐》电子游戏中类人玩法的神经进化方法.. 载于 CoSECivi. 113–124.[55] Maxim Mozgovoy 和 Iskander Umarov. 2011. 行为捕捉：为电子游戏构建可信且有效的人工智能智能体. 《国际艺术与科学杂志》4, 20 (2011), 243. [56] Kara Murias, Kathy Kwok, Adrian Gil Castillejo, Irene Liu 和 Giuseppe Iaria. 2016. 电子游戏使用对虚拟导航任务表现的影响. 《人类行为中的计算机》58 (2016), 398–406.[57] Mohamed Musbah, Matthew Mitchell Dixon, Geoffrey Jacoby Gordon, Mahmoud Adada, Soroush Mehri, Andrew James McNamara 和 Jonathan David Morrison. 美国专利 11213746，2019 年 10 月. 在中断期间向应用程序提供自动化用户输入. [58] Andrew Y Ng, Daishi Harada 和 Stuart Russell. 1999. 奖励变换下的策略不变性：奖励塑形的理论与应用. 载于 Icml，第 99 卷. 278–287.[59] John O'Keefe 和 Lynn Nadel. 1978. 作为认知地图的海马体. 牛津：克拉伦登出版社，纽约市，纽约. [60] Gabriele Paolacci, Jesse Chandler 和 Panagiotis G Ipeirotis. 2010. 在亚马逊土耳其机器人上开展实验. 《判断与决策》5, 5 (2010), 411–419. [61] Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexan- der Novikov, Gabriel Barth-Maron, Mai Gimenez, Yury Sulsky, Jackie Kay, Jost To- bias Springenberg, 等. 2022. 一个通用型智能体. arXiv 预印本 arXiv:2205.06175 (2022). [62] Anthony E Richardson, Morgan E Powers 和 Lauren G Bousquet. 2011. 电子游戏经验可预测虚拟导航表现，但不能预测现实导航表现. 《人类行为中的计算机》27, 1 (2011), 552–560. [63] Ariel Rosenfeld, Moshe Cohen, Matthew E Taylor 和 Sarit Kraus. 2018. 在表格型强化学习中利用人类知识：一项人类受试者研究. 《知识工程评论》33 (2018), 1–26. [64] Matthias Scheutz, Paul Schermerhorn, James Kramer 和 David Anderson. 2007. 迈向自然类人人机交互的第一步. 《自主机器人》22, 4 (2007), 411–423.[65] Douglas Schuler 和 Aki Namioka. 1993. 参与式设计：原则与实践. CRC 出版社. [66] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford 和 Oleg Klimov. 2017. 近端策略优化算法. arXiv 预印本 arXiv:1707.06347 1 (2017), 12 页. [67] Peter Seiler, Bongsob Song 和 J Karl Hedrick. 1998. 防碰撞系统的开发 . 《SAE 汇刊》 (1998), 1334–1340. [68] Alessandro Sestini, Linus Gisslén, Joakim Bergdahl, Konrad Tollmar 和 An- drew D Bagdanov. 2022. 利用好奇心条件化近端轨迹进行自动化游戏测试与验证. 《IEEE 游戏汇刊》 (2022).[69] Rohin Shah, Cody Wild, Steven H Wang, Neel Alex, Brandon Houghton, William Guss, Sharada Mohanty, Anssi Kanervisto, Stephanie Milani, Nicholay Topin, 等. 2021. 从人类反馈中学习的 MineRL BASALT 竞赛. arXiv 预印本 arXiv:2107.01969 (2021).[70] Noor Shaker, Julian Togelius, Georgios N Yannakakis, Likith Poovanna, Vinay S Ethiraj, Stefan J Johansson, Robert G Reynolds, Leonard K Heether, Tom Schu- mann 和 Marcus Gallagher. 2013. 2012 年马里奥 AI 锦标赛的图灵测试赛道：参赛作品与评估. 载于2013 IEEE计算智能与游戏会议 (CIG). IEEE, 1–8. [71] Aaron Sloman. 1999. 类人智能体需要何种架构？载于《理性主体基础》. 施普林格, 35–52. [72] Bhuman Soni 和 Philip Hingston. 2008. 经过训练、能像人类一样游戏的机器人更有趣. (2008). [73] Richard S Sutton 和 Andrew G Barto. 2018. 强化学习：导论. 麻省理工学院出版社，马萨诸塞州剑桥百老汇大道 1 号. [74] Jonathan Sykes 和 Melissa Federoff。2006。以玩家为中心的游戏设计。载于 *CHI'06* 计算机系统中的人为因素扩展摘要。1731–1734。[75] István Szita。2012。游戏中的强化学习。载于强化学习。施普林格，539–577。[76] Fabien Tencé 和 Cédric Buche。2010。面向电子游戏行为可信度的可自动化评估方法。*arXiv*预印本 *arXiv:1009.0501* （2010）。[77] Christian Thurau、Christian Bauckhage 和 Gerhard Sagerer。2004。为电脑游戏学习类人运动行为。载于自适应行为模拟国际会议论文集。315–323。[78] Julian Togelius、Georgios N Yannakakis、Sergey Karakovskiy 和 Noor Shaker。2013。评估可信度。载于可信的机器人。施普林格，215–230。[79] Edward B Tylor。1879。游戏史。《*Fortnightly*》25，149（1879），735–747。[80] Teija Vainio。2010。2000年代导航人机交互研究综述。《国际交互式移动技术期刊》（*iJIM*）4，3（2010）。[81] Oriol Vinyals、Igor Babuschkin、Wojciech M Czarnecki、Michaël Mathieu、An- drew Dudzik、Junyoung Chung、David H Choi、Richard Powell、Timo Ewalds、Petko Georgiev 等。2019。利用多智能体强化学习在《星际争霸II》中达到特级大师水平。《自然》575，7782（2019），350–354。[82] Luis Von Ahn 和 Laura Dabbish。2004。利用电脑游戏标注图像。载于*SIGCHI*计算机系统中的人为因素会议论文集。319–326。[83] Marcel Walch、Julian Frommel、Katja Rogers、Felix Schüssel、Philipp Hock、David Dobbelstein 和 Michael Weber。2017。从玩家体验的角度评估虚拟现实驾驶模拟。载于*2017*年*CHI*计算机系统中的人为因素会议扩展摘要集。2982–2989。[84] Dakuo Wang、Elizabeth Churchill、Pattie Maes、Xiangmin Fan、Ben Shneiderman、Yuanchun Shi 和 Qianying Wang。2020。从人际协作到人机协作：设计能够与人协同工作的人工智能系统。载于*2020*年*CHI*计算机系统中的人为因素会议扩展摘要集。1–6。[85] Henrik Warpefelt。2013。注意差距：探索非玩家角色的社交能力。博士学位论文。[86] Eric Wiewiora。2010。奖励塑形。施普林格美国，马萨诸塞州波士顿，863–865。https://doi.org/10.1007/978-0-387-30164-8_731[87] Erik Wijmans、Manolis Savva、Irfan Essa、Stefan Lee、Ari S Morcos 和 Dhruv Batra。2023。盲导航智能体的记忆中地图的涌现。*arXiv*预印本 *arXiv:2301.13261* （2023）。[88] Anthony Zador、Blake Richards、Bence Ölveczky、Sean Escola、Yoshua Bengio、Kwabena Boahen、Matthew Botvinick、Dmitri Chklovskii、Anne Churchland、Claudia Clopath 等。2022。迈向新一代人工智能：推动神经人工智能革命。*arXiv*预印本 *arXiv:2210.08340* （2022）。[89] Yunqi Zhao、Igor Borovikov、Jason Rupert、Caedmon Somers 和 Ahmad Beirami。2019。关于团队体育游戏中的多智能体学习。*arXiv*预印本 *arXiv:1906.10124* （2019）。[90] Brian D Ziebart、J Andrew Bagnell 和 Anind K Dey。2010。基于最大因果熵原理的交互建模。（2010）。[91] Evelyn Zuniga、Stephanie Milani、Guy Leroy、Jaroslaw Rzepecki、Raluca Georgescu、Ida Momennejad、Dave Bignell、Mingfei Sun、Alison Shaw、Gavin Costello 等。2022。人类如何感知电子游戏导航中的类人行为。载于*CHI*计算机系统中的人为因素会议扩展摘要。1–11。

超参数	值
批次大小	2048
丢弃率	0.1
学习率	2.5e-4
优化器	Adam [40]
伽马	0.99
拉姆达	0.95
裁剪范围	0.2
梯度范数裁剪系数	0.5
熵系数 c_2	0.0
价值函数系数 c_1	0.5
每次更新的小批量数	4
每次更新的训练轮数	4
回放缓冲区大小	5 × 批次大小

表6：训练符号型、混合型和奖励塑形智能体的超参数。我们使用近端策略优化算法 [66]训练所有智能体。关于这些超参数具体对应什么，感兴趣的读者可参阅原始近端策略优化论文，以了解更多细节。我们提供这些超参数值，以便复现实验。

A 强化学习智能体详情

本节首先介绍本研究中使用的基线智能体，然后说明所有智能体的训练详情。

A.1 基线智能体

混合型 和奖励塑形 智能体会额外接收一个 32×32 裁剪后的深度缓冲区视觉输入，而符号型 智能体则不会。视觉画面展示的是环境中智能体的第三人称视角。为处理这一额外的视觉通道，混合型和奖励塑形 智能体配备了一个卷积神经网络，用于学习提取高层视觉特征，并将这些特征与符号输入的表示拼接起来。

A.2 训练详情

我们在此提供有关训练设置的重要详情。

我们使用 PPO [66], 这一广泛应用的深度强化学习算法训练全部三种智能体。选择 PPO 训练智能体主要有几个原因。该算法之所以得到广泛使用，是因为实证研究发现，它在各种任务中都具有稳健性和高效性 [19]。我们对三种智能体架构分别训练 15 小时，相当于 1000 万个训练时间步，并至少使用 3 个不同的随机种子。我们使用 TensorFlow 2.3 [1] 以及带有分布式采样器的 OpenAI Baselines PPO2 实现 [17] 训练所有智能体。有关所有智能体所使用的训练超参数的完整列表版本，请参阅表 6。我们发现，这组参数在初步实验中的表现最佳。

为了在复杂的电子游戏环境中有效训练智能体，我们采用分布式方法，利用内部开发的样本收集框架和 Azure 云资源。训练样本由一组规模为 20 台的低优先级 GPU 虚拟机（Azure NV6）收集，每台运行 3 个电子游戏实例。随后，样本会发送至一个训练主节点，即仅配备 CPU 的 Azure E32s 内存优化型虚拟机。

B 行为研究详情

本节将介绍 MTurk 行为研究的更多详情。

B.1 完整说明

我们在此详细说明 MTurk 研究中包含的完整说明。

> 我们正在开展一项关于电子游戏导航的研究问卷。请阅读**说明**和**要求**，然后选择下方的链接完成问卷。问卷结束时，您会收到一个代码，请将其粘贴到下方的箱子中，以获得参与问卷的报酬。
>
> **说明**:
> • **概述**：问卷为匿名问卷，包含一份必填的知情同意书、理解力检查、一些背景信息，以及 6 个视频部分，每部分有 3 道题。所有问题均为必答题。• **所需时间**：约**30 分钟**。
> • **报酬**：完成任务后，您将获得固定报酬**$6.50**；如果回答质量较高，还有机会获得**$1 的奖金**。例如，复制粘贴答案，或回答内容与各页面上的视频无关，将无法获得奖金。
>
> • 该 MTurk HIT 的时限为 1 小时。超过 1 小时后，您**将无法**提交（请记得在 *1* 小时内提交或退回 *HIT*，以免超时！）
>
> • 如果您开始任务后改变主意，可以随时终止参与，并在 1 小时内**退还 HIT**，但**不会** 因退还的 HIT 或未完成的部分获得报酬。
>
> **要求**:
> • 您必须完成所有问题。
> • 您之前不得完成过名为"导航图灵测试（NTT）"的 HIT。重复参与者不符合资格，也不会获得报酬。• 您不能使用平板电脑或手机参与。**完成问卷时，请确保此窗口保持打开状态**。 完成后，您将返回此页面，将代码粘贴到箱子中。

B.2 收集的数据

除了收集同意、熟悉度和 HNTT问题外，我们还收集了以下数据：按页面划分的时间数据，以及向每位参与者呈现试次的顺序。

C 编码详情

我们提供了两位标注者在标注回答时共同商定并使用的编码指南。图 6 展示了该指南的屏幕截图。

编码的目标

- 独立性
- 区分

通用流程

- 理想情况下，我们希望不同编码之间有明确区分，也不希望这些编码高度依赖。例如，我们不希望每个目标导向的实例也都被编码为碰撞规避。因此，如果某个内容可以用两种方式编码，最好将其编码为更具体的 实例。
- 如果我们认为这是碰撞规避，就只将其编码为碰撞规避，而不是环境响应性。
- 如果我们认为是流畅度，就只编码为流畅度，而不编码为机械性。即使回应指出了刻板的机器人行为，也仍然如此。

编码

1. **动作平滑度：**指智能体导航或摄像机运动的质量。此编码同时考量较为即时的顿挫动作，以及持续时间较长的之字形运动行为。

   a. 关键词：i. 流畅（+）ii. 卡顿（-）iii. 顿挫（-）iv. 直线（+）v. 急转（-）vi. 摄像机移动（-）vii. 摄像机平稳（+）viii. 流畅（+）ix. 僵硬（-）x. 之字形（-）xi. 笨重（-）

2. **目标导向**关注智能体的行为看起来有多么有意图。参与者描述与其感知到的目标相关的行为时，即使该目标并非视频中的主要目标，也会将其纳入此代码。例如，如果他们提到有意探索，我们就将其编码为目标导向。

   a. 关键词：
      i. "知道自己要去哪里" (+) ii. 带着意图移动 (+) iii. 有目的地导航 (+) iv. 专注地移动 (+) v. 做出决策 (+)

3. **碰撞规避**可视为目标导向行为的一种体现，但我们单独设立了一个编码，因为我们认为它本身就是一项重要特征。为保持编码的独立性，被编码为碰撞规避的行为不应再被编码为其他类别——除非参与者在同一回应中指出了另一种目标导向行为，等等。碰撞规避的一个例子是智能体试图避免"撞上"其他物体。若智能体试图收集一个强化道具，则这种回应不应被视为碰撞规避。

   a. 关键词：i. 碰撞（-）ii. 避开（+）iii. 撞毁（-）iv. 撞上障碍物（-）

4. **环境响应性**旨在捕捉智能体与游戏环境之间的关系。在现实场景中，这可能表现为一个人走在人行道上而非草地上，或对环境作出响应（例如行人信号灯）。动态。这里专注于行为的意图。这段代码还旨在捕捉遵守规范或（可能未明说的）惯例的情况，而人们可能只是隐约有所感知。

   a. Key关键词：
      i. 探索（+）ii. 沿着路径行走（+）iii. 收集强化道具（+）iv. 已经了解环境中的一切（-）v. 走捷径（+）vi. 看穿墙壁（-）vii. 走了捷径（+）

5. **非机械性**是一种更为模糊的编码，试图捕捉对人类不完美的先入之见。机械性的一个例子行为被认定为过于完美。此代码也涵盖任何提及犯错或错误纠正的内容。

   a。关键词：i. 过于完美（-）ii. 犯错（+）iii. 精确（-）iv. 微小修正（-）v. 随机（+）vi. 过度纠正（-）

6. **直觉**指的是人们感觉某种行为更像人类，却没有充分且具体的理由。我们纳入此代码，以涵盖这类情况即参与者能够辨认出自己认为更像人类的行为，却难以将其表达出来。
   a. 关键词：
      i. 自然
      ii. 感觉
      iii. 看起来是
      iv. 正常
      v. 逼真

7. **自我指涉** 是一种编码，用于记录评判者将智能体行为与自己的游戏方式联系起来的情况。例如，参与者可能会提到，如果自己这样玩会感到不适，或者提到自己玩游戏时通常会收集所有强化道具。a. 关键词：i. "和我的游戏方式一样"

弃用

1. 如果参与者似乎是随机判断其是否具有人类特征，便归入此类。这可能表现为他们说"我不知道"或"我无法判断"。相较之下，如果他们说判断是基于感觉的，我们则将其归类为直觉。2. 如果自由形式响应与实际响应相矛盾。

图 6：标注员使用的编码指南。该指南包含对编码的说明、通用流程和弃用指南。