arXiv:2303.03548v2 [cs.RO] 2 Oct 2024

大语言模型作为人机交互的零样本人类模型

Bowen Zhang[1] 和 Harold Soh[1,2]

摘要—人类模型在人机交互（HRI）中发挥着关键作用，使机器人能够考虑自身动作对人的影响，并据此规划其行为。然而，构建优秀的人类模型颇具挑战；要捕捉依赖上下文的人类行为，需要大量先验知识和／或海量交互数据，而这两者都难以获得。在本研究中，我们探索大语言模型（LLM）作为人机交互零样本人类模型的潜力——这些模型已吸收了海量由人类生成的文本数据。我们在三个社交数据集上进行的实验取得了令人鼓舞的结果；大语言模型的表现能够媲美专门构建的模型。尽管如此，我们也讨论了当前的局限性，例如对提示的敏感性，以及空间／数值推理失误。基于研究发现，我们展示了如何将基于大语言模型的人类模型整合到社交机器人的规划过程中，并应用于聚焦信任这一重要要素的人机交互场景。具体而言，我们在一个模拟的基于信任的清桌任务上进行了案例研究，并复现了依赖定制模型的既有结果。接下来，我们开展了一项新的机器人传递餐具实验（$n = 65$），初步结果表明，使用基于大语言模型的人类模型进行规划，能够优于基本的短视规划。总之，我们的结果表明，大语言模型为人机交互中的人类建模提供了一种前景可观但尚不完善的方法。

I. 引言

人类模型赋予机器人推理人类行为的能力，从而实现以人为中心且个性化的机器人行为 [1]。尽管具有这些优势，精确的人类建模仍是一项重大挑战。手工构建的人类模型通常编码了强假设（例如 [2]–[5]），这可能会限制灵活性，也难以扩展到现实世界场景。另一种方法是使用非参数数据驱动模型（例如 [6], [7]）。近年来，由于深度学习在计算机视觉到游戏对弈等众多领域取得成功，这些模型日益流行。然而，这些方法通常需要大量难以收集且成本高昂的人类交互数据。

在这项工作中，我们尝试通过使用大型预训练模型（或称“基础模型”）来规避这些问题。具体而言，
我们考察了大语言模型（LLM），这些模型凭借生成可信且听起来像人类的文本的出色能力，以及在未经明确训练的下游任务中的表现，引发了广泛关注。

本研究由新加坡国家研究基金会和DSO国家实验室通过新加坡人工智能计划（AISG奖项编号：AISG2-RP-2020-017）资助。[1]新加坡国立大学计算机科学系 {bowenzhang, harold}@comp.nus.edu.sg。 [2]新加坡国立大学智能系统研究所（SSI）。

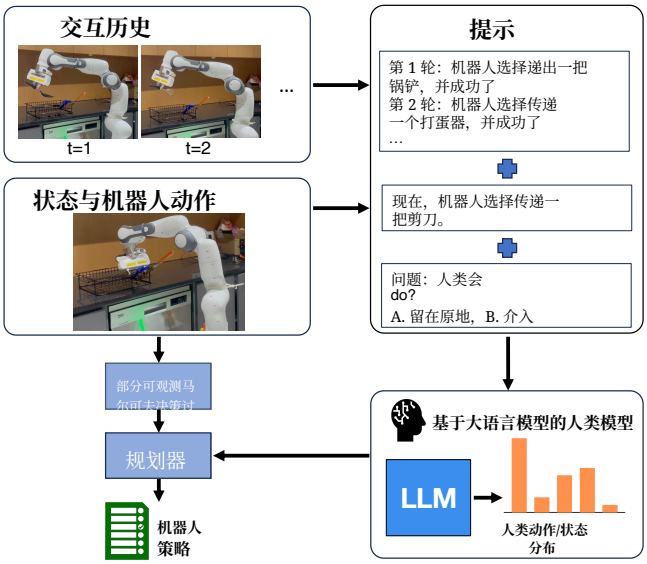

图1：在本工作中，我们探索如何将大语言模型（LLMs）用作人机交互中的零样本人类模型。我们首先使用基准数据集评估此类模型的有效性。随后，我们展示了如何将基于大语言模型的模型用于两个与信任相关的人机交互场景中的规划。

在此，我们提出问题：**大语言模型能否有效地充当人机交互中的人类模型?**

由于大语言模型经过海量人类生成文本的训练，因此可以设想，它们已经捕捉到了人类行为的某些要素。如果我们的这一问题最终得到肯定回答，那么大语言模型将造福人机交互领域，尤其是我们这些开发或使用计算模型来开展人机协作与交互的人 [8], [9]。我们尤其关注大语言模型的零样本 能力，因为这不需要额外的人类数据。另一方面，我们也有充分理由认为，大语言模型并不能很好地替代现有的人类模型——人机交互任务同时涉及社会和物理层面，而近期证据表明，在需要物理推理等超出语言能力范围的技能的任务中，大语言模型可能会遭遇灾难性的失败（例如 [10]–[13]）。

我们首先介绍一项实证研究，结果表明，大语言模型确实能够很好地捕捉人类的潜在状态和行为。通过使用三个不同的数据集，我们发现，两种最先进的大语言模型（FLAN-T5 和 GPT-3.5 的一个变体）的预测性能可与专用机器学习模型相媲美。这是通过零样本方法实现的，无需使用任何额外的训练数据。

虽然很容易得出大语言模型是有效的人类模型这一结论，但实际情况更为复杂。对结果的深入分析表明，大语言模型在需要空间推理和数值推理的任务上表现欠佳，并且对提示语法较为敏感。这可能使其难以应用于真实世界的人机交互场景，也表明大语言模型或许最适合用作“任务级”人类模型。

为进一步推进研究，我们探索了如何将基于大语言模型的人类模型用于人机交互（HRI）的规划设置中（图 1）。我们首先在先前人机信任研究中使用的模拟清理桌面设置中进行了实验 [5] ，并取得了相当的结果。随后，我们介绍了一项新的餐具传递实验的初步结果，该实验有 $n = 65$ 位参与者。在这一更复杂的设置中，配备基于大语言模型的人类模型的规划器生成了合理的计划，从而缓解了过度信任。我们的实验还发现，一些参与者的判断依据是物体的物理特征，而我们的模型未能对此加以考虑。这表明，人机交互任务的复杂性可能要求人类模型将大语言模型与考虑几何形状、运动及其他物理特征的“低层级”模型相结合；这将是一个值得探索的未来研究方向。

II. 预备知识与相关工作

大语言模型。 语言模型（LM）是词元序列上的概率分布。自回归语言模型根据先前输入词元组成的序列 $w_{<k}$，对 k^{th} 位置的下一个词元 w_k的条件分布 $p(w_k|w_{<k})$进行建模。最先进的大语言模型拥有数百亿甚至数千亿个参数，例如 OpenAI 的 GPT-3 [14] 和 Google 的 PaLM [15]）这些大语言模型通过自监督方式在海量文本数据上进行训练，不仅能够生成类似人类的文本，还在下游任务中展现出令人瞩目的性能。

在本研究中，我们关注大语言模型生成合理人类文本的能力，即生成与作为一系列词元输入的先验信息一致的文本，这些词元序列被称为提示。我们的工作与近期使用大语言模型在经典心理学/经济学实验中复现人类行为的研究 [16] 以及关于大语言模型中可能出现心智理论 [17], 的研究相关，但后者仍存在高度争议 [18]。关键区别在于，我们关注大语言模型在人机交互中的适用性，而人机交互涉及包含机器人以及物理与社会交互的场景。另一个相关研究方向是使用语言模型进行规划。已有研究使用大语言模型预测物体可供性 [19]或直接生成行动计划 [12]。然而，近期研究也表明，大语言模型可能并不擅长规划 [11]。我们与上述工作的不同之处在于，我们提出使用大语言模型来近似人类行为，并利用基于大语言模型的人类模型进行规划，以获得合适的机器人策略。

HRI人类模型。总的来说，对人类进行建模是一项涉及多个学科的大型工作；本文聚焦于用于人机交互的计算模型。我们简要回顾近期研究，并建议希望了解更全面内容的读者参阅优秀的综述文章 [1], [8]。

现有的人机交互中人类行为建模方法大致可分为两种范式：心智理论（ToM）模型和“黑箱”数据驱动模型。心智理论模型泛指纳入一系列关于人类心理加工和行为假设的方法。例如，贝叶斯心智理论（BToM）假设人类会理性行事，并以贝叶斯方式更新信念 [4], [20], [21]。相比之下，“黑箱”模型几乎不作假设，而是以数据驱动的方式对人类行为建模（例如 [7]）。这类方法通常需要大量真实人类数据，而这些数据可能难以获取。近期有研究提出以多智能体方式训练智能体，例如 Fictitious Co-Play [22], ，从而免去对人类数据的需求。然而，这些基于强化学习的方法计算开销巨大，训练成本高，并且需要大量交互数据；在现实世界中，获取这些数据可能代价高昂，甚至不可能。近期的一种混合方法 MIRROR [6],，基于机器人的内部自我模型（通过强化学习训练）对人类进行建模，然后利用少量人类数据，使模型适应特定个体。然而，目前尚不清楚 MIRROR 或上述任何方法能否充分捕捉人类心理状态，例如情绪，而这在许多现实场景中至关重要。本文将探究基于大语言模型的人类模型能否克服上述不足。

III. 大语言模型作为零样本人类模型

A. 问题陈述

考虑一个由 N 个体$a^{(i)} \in \mathcal{A}$组成的环境，其中 $i = 1, \ldots, N$。个体可以是人类，也可以是机器人。给定时刻 t 的环境状态 $s_t \in \mathcal{S}$ 以及交互历史 $h_t = \{(s_k, \{u_k^{(i)}\}_i)\}_{k=1}^{t-1} \in \mathcal{H}$ ，其中 s_k 表示时刻 k 的环境状态， $u_k^{(i)}$ 表示个体 i 在时刻 k的动作，我们旨在对其中一个人类个体的特定属性 $z_t^H \in \mathcal{Z}^H$ $a^H \in \mathcal{A}$进行建模。更准确地说， z_t^H是一个抽象的、特定于场景的随机变量（例如，人类动作或心理状态），我们希望对其分布 $p(z_t^H|s_t, h_t)$进行建模。例如， z_t^H 可以表示人类对机器人执行特定任务的信任。

我们假设可以通过函数 f 将环境状态、个体和交互历史文本化为提示， $x_t = f(s_t, h_t, \mathcal{A})$。我们将基于大语言模型的人类模型表示为$p_l(z_t^H|x_t)$；为简单起见，我们假设大语言模型能够输出关于 z_t^H的分布，或者存在合适的映射。我们的研究旨在评估 $p_l(z_t^H|x_t)$是否能够很好地近似 $p(z_t^H|s_t, h_t)$。

B. 模型、数据集和提示

模型。 我们使用两种不同的大语言模型：（i）text-davinci-003（175B） [27]和（ii）FLAN-T5-XXL（11B） [28], ，分别将其称为 DAVINCI 和 T5。DAVINCI 是一个经过人类反馈微调并实现最先进性能的大型模型。T5 是一个较小的开源大语言模型，经过指令微调优化。在我们的实验中，我们在本地工作站上运行 T5，并通过

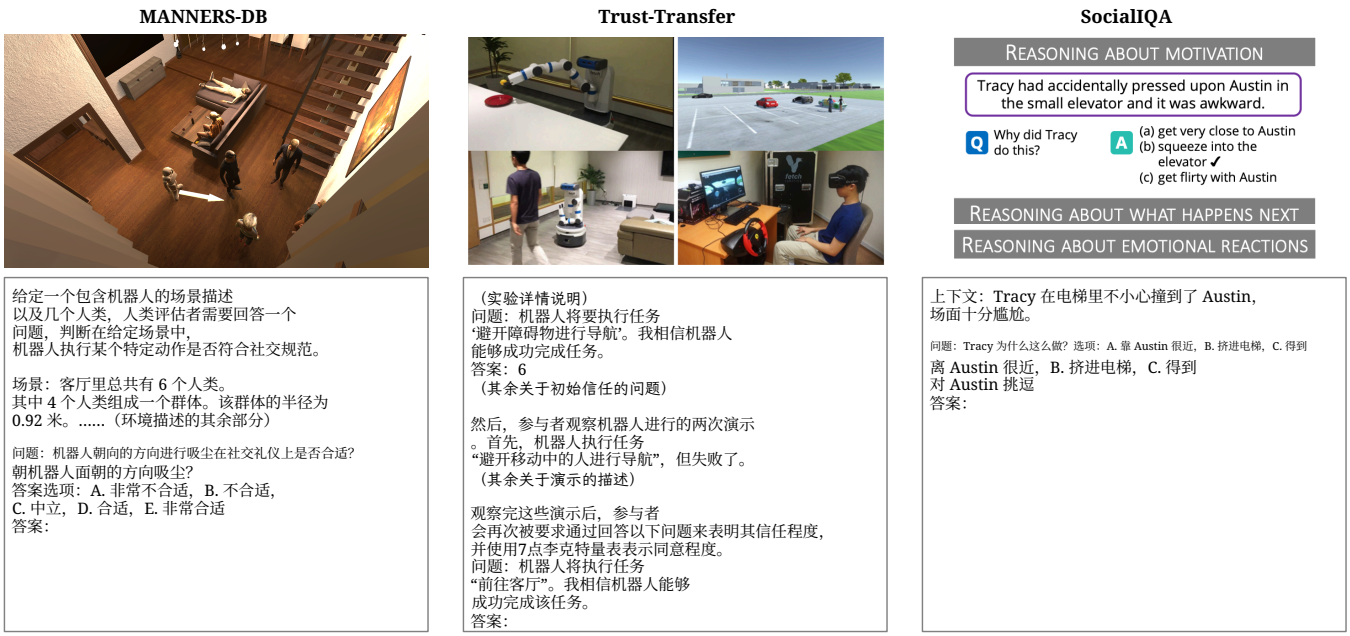

图 2：预测实验中的数据集和示例提示。我们使用两个人机交互数据集：MANNERS-DB [23] 和 Trust-Transfer [24], [25], , 并纳入 SocialIQA [26], 这一用于人际互动的通用社会推理基准测试。对于每个数据集，我们展示一张说明性图片（分别转载自 [23], [25], [26], ）和一个示例提示。

OpenAI API 访问 DAVINCI。这两种大语言模型都可以提供样本或下一个词元的分布。

数据集。我们使用两个与人机交互相关的数据集：

- **MANNERS-DB** [23], 一个近期数据集，用于评估机器人在特定社会情境中动作的适切性。测试集包含 800 个实例，每个实例都是一个模拟的家庭场景以及对机器人动作的描述。这些场景分为两组，具体取决于机器人动作是在其周围的圆圈内执行，还是沿箭头所指的方向执行。每个实例均由 15 位人工标注者标注；他们会看到模拟场景的图像，并对机器人动作的社会适切程度进行评分，评分范围为 1（非常不适当）至 5（非常适当）。该数据集使我们能够评估基于大语言模型的人类模型能否捕捉人机交互中物理和空间特征的影响。

- **信任迁移** [24] 是一个数据集，记录了人类在观察机器人执行任务后，对其能力的信任如何变化。该数据集涵盖家庭和驾驶两个不同领域，包含 189 个实例。每个实例包括参与者对三项测试任务的初始信任程度、观察到的两段演示，以及参与者在观察演示后对相同三项测试任务的最终信任程度。参与者使用 7点李克特量表评定对机器人的信任程度。我们纳入信任迁移数据集，(i) 以评估大语言模型能否捕捉交互历史对人类心理状态的影响；(ii) 因为该数据集与我们在第 IV 节中基于信任的案例研究密切相关。

以及一个通用的人际社会互动数据集：

- **SocialIQA** [26], 这是一个旨在衡量社会常识推理能力的基准测试。测试划分包含 1954 个实例，每个实例由上下文、一个问题和三个答案选项组成，其中一个是真实标签。在 SocialIQA 上表现出色，表明大语言模型具备推理人类社会互动（包括意图）的能力。例如，机器人可以观察两个人或多个人之间的互动，并推断他们的心理状态。

提示设计。 不同数据集所使用的具体提示各不相同（示例见图 2），但总体而言，我们遵循 [16] 提出的做法，设计清晰的模板化提示，以尽量减少无效补全，例如在给出多项选择题时，模型生成所列选项以外内容的可能性。提示的设计旨在包含：

- 向原始人机交互／社会实验中的人类参与者提供的实验信息，例如任务规则和机器人的特征；
- 交互历史、当前环境，以及对机器人 a^R动作的描述（在人机交互数据集中）；
- 用于探测大语言模型响应的查询。为提取关于 z_t^H 的概率分布，我们设计提示，使有效补全的数量仅限于 Z^H 中的元素；我们为

每个 z^H 通过多项选择题或李克特量表题分配一个单词元标签 [1]。因此，每个有效补全的长度都是一个词元，在比较词元概率时，我们可以避免偏向较短的回答。

由于篇幅限制，我们在在线补充材料 [29] 中提供了每个数据集的提示具体细节。

C. 基线模型、评估标准与局限性

基线模型。 我们将大语言模型与针对每个数据集开发的专用已发表模型进行了比较：

- **BNN-16CL**[2] [23], 一种在 MANNERS-DB 数据集上训练、用于预测机器人动作社会适切性的贝叶斯持续学习模型。
- **GPNN**，一种混合高斯过程与神经网络模型，用于预测 Trust-Transfer，并应用于 Trust-Transfer 数据集。
- **BERT-large** [30],一种拥有 3.4 亿参数、针对 SocialIQA 数据集微调为分类器的语言模型。

报告结果时，我们使用统称“基线模型”来指代各数据集对应的模型。

评估指标。我们主要使用误差分数和与众数的一致性（CwM）分数来评估上述模型。误差分数用于衡量给定模型的输出与数据集目标值的平均吻合程度。该分数取决于数据集，指各数据集作者所使用的指标；这使我们能够直接与先前报告的数值（如有）进行比较。MANNERS-DB、Trust-Transfer 和 SocialIQA 的误差分数分别为均方根误差（RMSE）、平均绝对误差（MAE）和错误率。为简便起见，我们将这些不同的指标统称为“误差分数”。

CwM 分数衡量大语言模型与多数人类标签之间的一致性。为使不同数据集的分数保持一致，我们首先将目标值二值化；对于 MANNERS-DB，如果多数标注者将某项机器人动作的适宜性等级评为 ≥ 3（中性），我们就将其标记为社会可接受（否则标记为社会不可接受）。对于分数为 ≥ 4 的 Trust-Transfer 数据集，我们也进行了类似处理。SocialIQA 数据集的每个实例只有一个标签，我们将其作为众数。

为了了解分布 $p_l(z_t^H|x_t)$ 与 $p(z_t^H|s_t, h_{t-1})$ 的匹配程度，我们还纳入了两个额外的度量：(i) 熵相对相似度分数 $S(q, p) = \frac{\sum_{z^H} q(z^H) \log p(z^H)}{\sum_{z^H} q(z^H) \log p(z^H)}$，以及 (ii) 2-瓦瑟斯坦度量 $W_2(q, p)$。我们将模型输出设为 q，并将目标经验分布设为 p。直观而言，分布的熵反映了该分布的“宽广”程度，或者其固有的不确定性。 $S(q, p)$ 分数揭示了模型的输出分布 q 是否反映了目标 p 的离散程度，或是过度／欠缺自信——数值为 1 表示熵相同，而 $S(q, p) < 1$ 表示分布范围更窄（反之亦然）。2-瓦瑟斯坦度量反映了两个分布的相似程度；它衡量需要移动多少概率质量，才能将 q 转换为 p。请注意，这两个与分布相关的度量仅能针对 MANNERS-DB 计算（因为 Trust-Transfer 和 SocialIQA 每个实例仅包含一个标签）。

局限性。对于 MANNERS-DB 和 SocialIQA，标签均由第三方人员提供，他们实际上并非环境的一部分。根据先前的研究 [31],[32],，我们假设人工标注者能够对他人的心理状态或行为作出可靠评估。其次，依赖手工设计的提示是一项重要局限；采用其他提示结构可能会取得更好的表现。最后，评估大语言模型时的一个问题是存在“数据泄漏”的风险，即大语言模型在训练过程中见过某个数据集。由于 DAVINCI 和 T5 的训练数据集并未公开，我们无法确认是否存在这种情况。尽管如此，SocialIQA 发生数据泄漏的可能性不大 [33]；而且，人机交互数据集（提示和答案）是为本研究创建的，因此大语言模型此前并未直接见过这些数据。不过，我们所用提示的原始数据集均已公开，大语言模型可能见过原始数据或相关论文。我们通过扩展信任任务，加入原始数据集中没有的条目，尝试缓解这一问题 [25]，详见第 IV-C 节。

D. 结果

误差和 CwM 分数汇总于表 I 中。在所有三个数据集中，DAVINCI 的零样本性能出人意料地出色，有时误差分数甚至低于专用模型。T5 在 MANNERS-DB 和 SocialIQA 数据集上也取得了不错的分数，但在 Trust-Transfer 数据集上的表现不佳；我们将在后续分析中深入探讨其表现欠佳的原因。CwM 分数表明，在 MANNERS-DB 和 SocialIQA 数据集中，大语言模型与多数标签的一致性达到了人类水平。我们观察到 SocialIQA 数据集中的人工标注者意见不一，因为平均而言，单个标注者与多数标签的一致率仅为 68.4%。

表 I 还展示了 MANNERS-DB 上的分布度量。总体而言，基线（贝叶斯模型）对目标分布的捕捉效果最佳。DAVINCI 往往会低估分布的离散程度，这一点可从其较低的熵相似度 $S(q, p)$ 分数看出。相比之下，T5 的分数为 > 1，表明其分布过于宽泛，但与经验性人工标签分布更为吻合；在所有上下文中，T5 的瓦瑟斯坦度量均低于 DAVINCI。

综上所述，上述结果表明，大语言模型的表现可与强大的基线模型相媲美。我们强调，这一成果是在零样本 方式下取得的，

[1] 大语言模型偶尔会生成无效补全（在我们的实验中占实例的 < 3 %）。我们将这些情况归入一个单独的“兜底”类别，并标记为错误。2

[2] 16CL 使用作者在 https:// github.com/jonastjoms/MANNERS-DB 提供的代码。我们报告的数值与 [23] 不同，可能是因为测试数据集划分有所不同。

表 I：误差和 CwM 分数及分布度量。 $S(q,\ p)$ 越接近 1，表示越能捕捉人类分布的离散程度。

		误差分数(↓)			CwM（↑）			$S(q,\ p)$			$W_2(q,\ p)$（↓）		
数据集	**上下文**	DAVINCI	T5	基线	DAVINCI	T5	人类	DAVINCI	T5	基线	DAVINCI	T5	基线
礼貌	圆圈	0.787	**0.708**	0.876	0.730	0.750	**0.760**	0.467	1.354	**1.127**	0.168	**0.164**	0.173
	箭头	**0.695**	0.762	0.664	**0.734**	0.632	0.717	0.586	1.226	**1.199**	0.206	0.141	**0.124**
	总体	**0.737**	0.739	0.764	0.733	0.684	**0.736**	0.645	1.281	**1.159**	0.189	0.151	**0.145**
Trust-Transfer	家庭	0.162	0.211	**0.156**	**0.896**	0.760	-	-	-	-	-	-	-
	驾驶	**0.154**	0.211	0.163	**0.806**	0.720	-	-	-	-	-	-	-
	总体	**0.158**	0.211	0.159	**0.852**	0.741	-	-	-	-	-	-	-
SocialIQA	总体	0.278	**0.181**	0.340	0.722	0.819	0.684[1]	-	-	-	-	-	-
							0.869[2]	-	-	-	-	-	-

[1] 平均人类表现，报告见 https://github.com/google/BIG-bench。[2]最佳人类表现，报告见 [26]。

表 II：DAVINCI 在 MANNERS-DB 中各个动作类型上获得的 CwM 分数。

机器人动作	**CwM**（↑）	**精确率**（↑）	**召回率**（↑）
吸尘	0.560	0.506	0.933
拖地	0.520	0.478	1.000
搬运热食	0.740	0.875	0.814
搬运冷食	0.670	0.944	0.698
搬运饮料	0.940	0.959	0.979
搬运小物件	0.960	0.960	1.000
搬运大型物体	0.710	0.583	0.600
清洁（拾取物品）	0.886	0.886	1.000
开启对话	0.661	0.755	0.841

这表明**大语言模型无需进一步训练或微调，就能作为有效的零样本人类模型应用于人机交互**。

E. 人机交互数据集的讨论与分析

上述结果令人鼓舞，但得分也表明，大语言模型并非完美模型，例如，T5 在信任迁移数据集上的表现不佳。接下来，我们将进一步分析大语言模型所犯的错误，以指出其中的陷阱和潜在的改进方向。

大语言模型在需要空间、物理或数值推理的人机交互任务上表现可能不佳。通过分析大语言模型在 MANNERS-DB 上的表现，可以看到这一问题。表 II 展示了 DAVINCI 在数据集中各项具体动作上的 CwM 分数，以及精确率和召回率。T5 的分数与之相近，因此我们将 T5 的 CwM 分数（以及比较各模型均方根误差的表格）放在在线附录中。「吸尘」和「拖地」的分数相对较低；这两项是数据集中最具侵入性的动作，需要理解个人空间 [23],，而根据提示，这需要物理、空间和数值推理——这些任务已知对大语言模型而言颇具难度 [11], [13]。我们还观察到，「搬运大型物体」和「开始对话」的表现也较低；在这些动作中，社会适切性也被发现与智能体间距离高度相关 [23]。

为了进一步探究这一问题，我们检查了“吸尘”和“拖地”的89个失败案例。通过人工检查场景图像和标注者的评论，我们确定了56种失败情境，其中机器人的动作明显侵入了人类的个人空间，例如，人类正站在机器人预定的工作区域内，但是

...*(*实验细节描述*)*参与者对“拿起并放置一个塑料罐”这项任务的信任评分为5分（满分7分）。... *(*其余初始信任描述以及演示描述*)*根据这些演示和初始信任，现在参与者将对“拿起并放置一个塑料罐”这项任务的信任评分为

图 3：在家居领域的信任迁移实验中使用的修改后的提示词示例。

表 III：使用修改后的提示词进行 Trust-Transfer 的结果。

	误差分数(↓)			CwM(↑)	
领域	DAVINCI	T5	基线	DAVINCI	T5
家庭	0.163	0.170	0.156	0.875	0.796
驾驶	0.155	0.160	0.163	0.806	0.802
总体	0.159	0.165	0.159	0.841	0.799

大语言模型预测该动作在社交上是恰当的。对于这些实例，我们直接向大语言模型提出了一个“是—否”问题：“机器人的指定工作区域是否侵入了任何人的个人空间？”。DAVINCI 和 T5 对全部 56 个问题都给出了错误答案。我们还尝试在 DAVINCI 上使用零样本思维链提示，该方法已被证明能够诱导出更合理的回答 [10]。性能有所提升，但仍不尽如人意（56 题中答对 15 题，26.7%）。以下是一个具有代表性的失败回答：

> “首先，距离机器人最近的人类与机器人相距 0.53 米。这意味着机器人的指定工作区域为 2.00 米，不会侵入最近的人类的个人空间。”

我们观察到，该回答在语言上是正确的，也利用了提示中的信息，但得出的结论不正确，因为人类出现在清洁区域内。

大语言模型对提示结构很敏感。回想一下，T5 在 Trust-Transfer 上表现不佳。检查 T5 的输出后发现，它非常倾向于将参与者观察后的信任预测为与其初始信任相同。我们推测，这是因为我们的提示结构类似于少样本提示，导致模型倾向于复制初始答案，因为问题完全相同（见

图 2）。为了验证这一猜想，我们设计了另一种提示，改变了初始信任和查询的呈现方式（图 3）。表 III 显示，这种新的提示风格显著提升了 T5 的表现。

摘要。 我们的研究表明，大语言模型出人意料地擅长充当零样本人类模型。尽管如此，我们的分析也表明：（i）大语言模型在需要空间与数值理解的人机交互任务中可能表现不佳；（ii）大语言模型对提示设计很敏感。这些发现进一步丰富了有关大语言模型优势与局限的日益增多的研究文献 [11], [13], [16]–[18]。在人机交互的上下文中，应用大语言模型时应谨慎；人机交互通常立足于现实世界，因此涉及空间与物理推理。总体而言，我们的结果表明，**大语言模型更适合作为任务层面（符号化）的人类模型，而对于连续空间中的几何与运动，可能需要采用其他"低层级"模型**。

IV. 案例研究：使用基于大语言模型的人类模型进行规划

本节中，我们将从预测转向人机交互规划。借助研究中获得的见解，我们将展示如何在用于人机交互的规划器中使用基于大语言模型的人类模型。我们首先复现一项关于清理桌面任务中信任的先前研究，随后介绍一项新的餐具传递实验。

A. 问题陈述

根据我们此前的研究结果，我们将人机交互场景设定在任务层面，并将其形式化为马尔可夫决策过程。在每个时间步t，人类 H 和机器人 R都能完全观测到世界状态 s_t。机器人执行动作 u_t^R，随后人类作出响应并执行动作 u_t^H。之后，世界状态按照转移函数 $p(s_{t+1}|s_t,\ u_t^H,\ u_t^R)$发生变化，并获得奖励 $r(s_t,\ u_t^H,\ u_t^R,\ s_{t+1})$。因此，机器人的最优策略 π_*^R为，

$$\pi_*^R = \arg\max_{\pi_R} \mathop{\mathbb{E}}_{u_t^H\sim\pi^H, u_t^R\sim\pi^R} \sum_{t=0}^{\infty} \gamma^t r(s_t, u_t^H, u_t^R, s_{t+1})$$

其中 γ 为折扣率。我们使用 $p^l(z_t^H|x_t)$近似未知的人类策略 π^H，其中 $z_t^H = u_t^H$且$x_t = f(s_t,\ h_t,\ u_t^R,\ \{H,R\})$，并假设转移动态已知。因此，机器人的决策过程成为一个马尔可夫决策过程，可以使用现成规划器。在以下实验中，由于马尔可夫决策过程规模相对较小且规划时域较短，我们采用简单的离线值迭代。未来的工作可以使用更复杂的规划器和基于大语言模型的模型来研究更复杂的场景。

B. 清理桌面实验

实验设置。 我们首先在模拟的清理桌面实验中测试基于大语言模型的规划器 [5], , 其中人类与机器人协作，将物体从桌上清理掉。这些物体包括三个水瓶、一个鱼罐头和一个酒杯。在每个时间步，机器人会选择一个物体来移除。随后，人类选择是否介入并拿起物体，或原地不动，让机器人移除物体。如果人类原地不动且机器人成功，他们将根据物体获得奖励：塑料瓶得1分，鱼罐头得2分，酒杯得3分。然而，如果他们原地不动而机器人失败，则会受到惩罚：塑料瓶不受惩罚，鱼罐头罚4分，酒杯罚9分。如果他们选择介入，则不会获得奖励或受到惩罚。这里假设机器人绝不会失败，但不会向人类参与者透露这一信息。

```
... （实验设置和规则说明）
第1回合：机器人选择：塑料瓶；人类选择：原地不动；
结果：机器人成功移除了塑料瓶。
（在以下情况下加入"人类对机器人的信任增加了。"
TC）
... （其余交互历史）

问题：现在机器人选择移走酒杯，人类会怎么做？答案选项：A.
干预，B. 留在原地。或者在 是/否：问题：人类现在会信任机器人
移走酒杯吗？答案选项：A. 会，B. 不会。
```

图4：清理桌面实验中使用的提示示例。

提示设计。 如第 III-E 节前文所述，提示结构对于引导大语言模型给出正确回应十分重要。这里，我们考虑一种带有若干变体的标准提示（示例见图4）：

- TC：我们使用一道多项选择题，在每个回合中明确纳入信任变化（大语言模型根据观察结果预测的最可能的信任变化），选项为 {增加、减少、不变}
- YN：机器人选择移走某个物体时，提示并不询问人类会采取什么动作，而是询问一个是非问题：人类是否会信任机器人执行这一操作。我们假设信任与人类动作之间存在确定性关系，即如果人类不信任机器人完成任务，就会进行干预；否则就会保持不动。

结果。我们使用作者提供的仿真代码，将结果与 TRUST-POMDP 规划器 [5] 进行比较。10^4 次仿真运行的结果如表 IV 所示。所有基于 DAVINCI 的规划器始终优于其 T5 对应版本，且表现与 TRUST-POMDP 规划器相当。采用 TC 和 YN 提示都提升了 T5 和基于 DAVINCI 的规划器的性能。

C. 餐具传递实验

实验设置。在这个场景中，人类正在厨房里清洗餐具，机器人（Franka Emika Panda机械臂）则帮忙把脏餐具递给人类。物体包括锅铲、打蛋器、一把剪刀和一把刀（图5）。在每个时间步，机器人会选择一个物体递出。随后，人类会在两种

表 IV：模拟清理桌面实验结果。

	平均回报 (↑)	Glass 上的干预概率 (↓)
DAVINCI-TC-YN	**6.17 (0.034)**	**0.352**
DAVINCI-YN	6.13 (0.034)	0.366
DAVINCI-TC	6.15 (0.034)	0.357
DAVINCI	6.14 (0.034)	0.360
T5-TC-YN	6.10 (0.034)	0.368
T5-YN	6.01 (0.035)	0.398
T5-TC	5.94 (0.034)	0.395
T5	5.95 (0.035)	0.405
TRUST-POMDP	**6.17 (0.034)**	**0.352**



图5：（左）实验中使用的餐具：锅铲、打蛋器、剪刀和刀。（右）实验环境模拟了一个厨房。为便于观察，餐具托盘以高亮显示。

动作中选择一种：（A）介入并自行取回物体，或（B）待在原地，让机器人把物体递过来。

参与者遵循与清理餐桌实验类似的规则：如果人类留在原地且机器人成功，他们将获得1点奖励。由于餐具是脏的，只有当机器人以便于人类握住干净手柄的方式递交物品时，才视为递交成功。如果人类留在原地而机器人失败，他们将受到 -1的惩罚。如果人类选择介入，则不会获得奖励或受到惩罚。

以下信息不会 透露给参与者：机器人总能成功递出锅铲、打蛋器和剪刀，但总是无法正确递出刀，并会将其掉落。发生这种情况时，实验将终止，并受到 -10 的惩罚。因此，机器人必须主动校准人类信任，以便

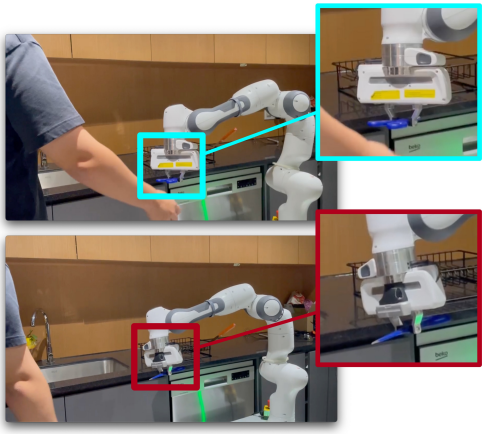

图6：递交剪刀时，通过递交安全且干净的手柄而成功（上图），以及通过递交错误的一端而故意失败（下图）。

机器人尽可能多地提供帮助（并在此过程中累积奖励），同时让人类**不** 信任它递出刀具。此设置旨在模拟人类可能过度信任 机器人的能力，因而需要机器人降低人类信任，以改善长期任务表现 [5]的情形。在此，机器人通过故意失败来降低信任，但未来的工作可以探究其他策略，例如口头沟通。为便于进行此校准过程 [34],，我们允许机器人故意在递送所有餐具（刀除外）时失败，方法是将错误的部位递给人类（使人类只能抓住脏污的一端），即将这些失败动作加入机器人的动作空间。图6）。这种故意失败会受到 -1的惩罚。

假设与规划器。 我们考虑两种不同的计划：

- LLM-PLAN 是一种确定性规划方案，由基于 DAVINCI-TC-YN 的人类模型通过规划生成，其提示与清理桌面设置中的提示相似。；
- BASIC-PLAN 是一种短视规划，不考虑信任，并始终按锅铲、打蛋器、剪刀和刀的顺序传递物品（因为相应的奖励相同），且从不故意失败；它与 [5] 中使用的短视规划相似。

我们的假设是，与遵循 BASIC-PLAN 的机器人相比，遵循 LLM-PLAN 的机器人能够减少过度信任并带来更高的回报。

参与者招募与分配。 我们从大学校园招募了 65 名参与者（年龄为 22 至 54 岁）。参与者被随机分为两组，每组分别与遵循 LLM-PLAN 或 BASIC-PLAN 的机器人配对。出于安全考虑，参与者不会与机器人进行身体接触。相反，他们完成了一项交互式视频调查，在每个回合观看预先录制的机器人行为视频。

结果。 BASIC-PLAN组中有33名参与者中的21名（63.6%）允许机器人递刀，而LLM-PLAN组中有32名参与者中的9名（28.1%）允许机器人递刀。因此，LLM-PLAN机器人的平均回报（-1.88）高于BASIC-PLAN机器人（-4.24）。单因素方差分析显示，这一差异在 α =5%水平上具有统计显著性（$F(1, 63) = [4.302], p = 0.042$）。这些结果支持了我们的假设。

讨论。BASIC-PLAN和LLM-PLAN机器人都选择在前两轮递上锅铲和打蛋器。之后，两种计划会根据人类行为而有所不同。与BASIC-PLAN不同，如果人类决定不对前两件物品进行干预，LLM-PLAN机器人会选择故意在剪刀上失败，以防止过度信任。目睹这一故意失败的大多数参与者随后会选择干预递刀（20名参与者中有18名，90%）。正如预期，体验过BASIC-PLAN机器人成功递上剪刀的参与者不会干预递刀（15名参与者中有1名进行了干预，占6.67%）。

然而，并非所有参与者都经历了上述事件序列。LLM-PLAN组中有五名参与者选择在第二轮干预打蛋器，

这随后促使机器人在下一轮递出刀（因为它预测人类不会信任它递刀）。然而，这些参与者随后选择了不 干预。当被问及为何这样决定时，他们解释说，与手柄较短、被认为对机器人而言更难处理的打蛋器相比，锅铲和刀的手柄形状更为相似（图5）。在观察到机器人未能正确递出剪刀后，仍选择信任机器人递刀的两名参与者也给出了类似的理由：他们认为剪刀的形状与其他物品不同。这些评论表明，对物体形状的感知会影响人机信任，也凸显了将此类信息纳入大语言模型所面临的挑战。

五、结论

在本研究中，我们首次探讨了人机交互中的基于大语言模型的零样本人类模型。我们的主要发现是，大语言模型可以作为有效的任务级人类模型——它们能够对高层次的人类状态和行为进行建模。我们进一步证明，在两个基于信任的社交人机交互场景中，融入基于大语言模型的人类模型可以生成合理的规划。然而，我们的研究表明，当前的大语言模型不太可能单独成为准确的人类模型；由于空间／数值推理能力的局限，它们难以处理低层次的几何／形状特征。尽管如此，它们能够捕捉人类行为的某些方面，将大语言模型与其他模型相结合将成为一个有趣的未来研究方向。

参考文献

[1] R. Choudhury、G. Swamy、D. Hadfield-Menell 和 A. D. Dragan，《论人机交互中模型学习的效用》，载于*2019*年第*14*届*ACM/IEEE*国际人机交互会议（*HRI*）。IEEE，2019年，第317–325页。[2] B. D. Ziebart、N. Ratliff、G. Gallagher、C. Mertz、K. Peterson、J. A. Bagnell、M. Hebert、A. K. Dey 和 S. Srinivasa，《行人的基于规划的预测》，载于*2009*年*IEEE/RSJ*国际智能机器人与系统会议。IEEE，2009年，第3931–3936页。[3] D. Sadigh、S. Sastry、S. A. Seshia 和 A. D. Dragan，《利用对人类行为影响的自动驾驶汽车规划》，载于机器人学：科学与系统，第2卷。美国密歇根州安娜堡，2016年，第1–9页。[4] C. L. Baker、J. Jara-Ettinger、R. Saxe 和 J. B. Tenenbaum，《人类心理推理中对信念、欲望和知觉的理性定量归因》，《自然·人类行为》，第1卷，第4期，第0064页，2017年。[5] M. Chen、S. Nikolaidis、H. Soh、D. Hsu 和 S. Srinivasa，《面向人机协作的基于信任的规划》，载于*2018*年*ACM/IEEE*国际人机交互会议论文集，2018年，第307–315页。[6] K. Chen、J. Fong 和 H. Soh，《MIRROR：用于辅助性人机交流的可微深度社会投影》，*arXiv* 预印本 *arXiv:2203.02877*，2022年。[7] E. Schmerling、K. Leung、W. Vollprecht 和 M. Pavone，《面向人机交互的多模态概率模型规划》，载于*2018*年*IEEE*国际机器人与自动化会议（*ICRA*）。IEEE，2018年，第3399–3406页。[8] A. Thomaz、G. Hoffman、M. Cakmak 等，《计算人机交互》，《机器人学基础与趋势》，第4卷，第2–3期，第105–223页，2016年。[9] O. Nocentini、L. Fiorini、G. Acerbi、A. Sorrentino、G. Mancioppi 和 F. Cavallo，《社交机器人行为模型综述》，《机器人学》，第8卷，第3期，第54页，2019年。[10] T. Kojima、S. S. Gu、M. Reid、Y. Matsuo 和 Y. Iwasawa，《大语言模型是零样本推理器》，*arXiv* 预印本 *arXiv:2205.11916*，2022年。[11] K. Valmeekam、A. Olmo、S. Sreedharan 和 S. Kambhampati，《大语言模型仍然不会规划（用于评测大语言模型规划与变化推理能力的基准测试）》，arXiv 预印本 arXiv:2206.10498，2022年。[12] W. Huang、P. Abbeel、D. Pathak 和 I. Mordatch，《语言模型作为零样本规划器：为具身智能体提取可执行知识》，载于国际机器学习会议。PMLR，2022年，第9118–9147页。[13] Y. Xie、C. Yu、T. Zhu、J. Bai、Z. Gong 和 H. Soh，《利用大语言模型将自然语言转换为规划目标》，

2023 [ ] 2302 05128 。在线。可访问：[14] https://arxiv.org/abs/。T. Brown、B. Mann、N. Ryder、M. Subbiah、J. D. Kaplan、P. Dhariwal、A. Neelakantan、P. Shyam、G. Sastry、A. Askell 等，《语言模型是少样本学习器》，《神经信息处理系统进展》，第33卷，第1877–1901页，2020年。[15] A. Chowdhery、S. Narang、J. Devlin、M. Bosma、G. Mishra、A. Roberts、P. Barham、H. W. Chung、C. Sutton、S. Gehrmann 等，《PaLM：利用 Pathways 扩展语言建模》，arXiv 预印本 arXiv:2204.02311，2022年。[16] G. Aher、R. I. Arriaga 和 A. T. Kalai，《利用大语言模型模拟多人类》，arXiv 预印本 arXiv:2208.10264，2022年。[17] M. Kosinski，《心智理论或已在大语言模型中自发涌现》，arXiv 预印本 arXiv:2302.02083，2023年。[18] T. Ullman，《大语言模型在心智理论任务的细微改动下表现不佳》，arXiv 预印本 arXiv:2302.08399，2023年。[19] M. Ahn、A. Brohan、N. Brown、Y. Chebotar、O. Cortes、B. David、C. Finn、K. Gopalakrishnan、K. Hausman、A. Herzog 等，《照我能做的做，而非照我说的做：将语言建立在机器人可供性之上》，arXiv 预印本 arXiv:2204.01691，2022年。[20] J. J. Lee、F. Sha 和 C. Breazeal，《基于贝叶斯心智理论的非语言交流方法》，载于2019年第14届ACM/IEEE国际人机交互会议（HRI）。IEEE，2019年，第487–496页。[21] C. L. Baker 和 J. B. Tenenbaum，《利用贝叶斯心智理论对人类计划识别建模》，《计划、活动与意图识别：理论与实践》，第7卷，第177–204页，2014年。[22] D. Strouse、K. McKee、M. Botvinick、E. Hughes 和 R. Everett，《在没有人类数据的情况下与人类协作》，《神经信息处理系统进展》，第34卷，第14 502–14 515页，2021年。[23] J. Tjomsland、S. Kalkan 和 H. Gunes，《注意礼貌！用于评估机器人动作社会适切性的数据集与持续学习方法》，《机器人学与人工智能前沿》，第9卷，第4页，2022年。[24] H. Soh、P. Shu、M. Chen 和 D. Hsu，《人类对机器人能力的信任在不同任务间的迁移》，载于 R:SS，2018年。[25] H. Soh、Y. Xie、M. Chen 和 D. Hsu，《人机交互中的多任务信任迁移》，《国际机器人研究期刊》，第39卷，第2–3期，第233–249页，2020年。[26] M. Sap、H. Rashkin、D. Chen、R. LeBras 和 Y. Choi，《SocialIQA：关于社会互动的常识推理》，arXiv 预印本 arXiv:1904.09728，2019年。[27] L. Ouyang、J. Wu、X. Jiang、D. Almeida、C. L. Wainwright、P. Mishkin、C. Zhang、S. Agarwal、K. Slama、A. Ray 等，《使用人类反馈训练语言模型遵循指令》，arXiv 预印本 arXiv:2203.02155，2022年。[28] H. W. Chung、L. Hou、S. Longpre、B. Zoph、Y. Tay、W. Fedus、E. Li、X. Wang、M. Dehghani、S. Brahma 等，《扩展指令微调语言模型》，arXiv 预印本 arXiv:2210.11416，2022年。[29] B. Zhang 和 H. Soh，《大语言模型作为人机交互零样本人类模型的补充材料》，https://github.com/clear-nus/llm-human-model，2023年。[30] J. Devlin、M.-W. Chang、K. Lee 和 K. Toutanova，《BERT：面向语言理解的深度双向变换器预训练》，arXiv 预印本 arXiv:1810.04805，2018年。[31] C. M. Heyes 和 C. D. Frith，《读心的文化演化》，《科学》，第344卷，第6190期，第1243091页，2014年。[32] J. Zhang、T. Hedden 和 A. Chia，《序贯行动博弈中的观点采择与心智理论推理深度》，《认知科学》，第36卷，第3期，第560–573页，2012年。[33] A. Srivastava、A. Rastogi、A. Rao、A. A. M. Shoeb、A. Abid、A. Fisch、A. R. Brown、A. Santoro、A. Gupta、A. Garriga-Alonso 等，《超越模仿游戏：量化并推断语言模型的能力》，arXiv 预印本 arXiv:2206.04615，2022年。[34] J. Lee、J. Fong、B. C. Kok 和 H. Soh，《相互了解：为人机协作校准意图、能力与信任》，IEEE国际智能机器人与系统会议，第6296–6303页，2020年。