Modern Linguistics 现代语言学, 2026, 14(5), 212-221
Published Online May 2026 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2026.145393

Hans汉斯

社会－认知语用视角下大语言模型识别与生成反讽能力的探究

郭东梅，白丽梅

西北师范大学外国语学院，甘肃 兰州

收稿日期：2026年4月3日；录用日期：2026年4月30日；发布日期：2026年5月13日

摘　要

本研究以社会认知语用学为理论基础，选取《脱口秀大会》中的反讽话语为语料，考察大语言模型DeepSeek识别与生成反讽的能力。通过一系列提示语，探究DeepSeek如何利用语境机制、凸显机制与合作－自我中心机制来处理反讽表达。并分析其基于社会认知语用学生成反讽语料的表现。研究发现，DeepSeek能够借助上述机制识别反讽话语，但在处理涉及文化背景语境的表达可能出现偏差；该模型能依据提示语设计出符合社会认知语用学特点的反讽语料，但其表达可能较为直白。针对这些局限，本文进一步探讨了通过改进训练数据、调整模型架构及优化提示工程策略等可操作的优化路径。鉴于反讽是一种复杂且多功能的现象，本研究不仅有助于提升公众对人工智能技术使用的认识，也为评估大语言模型提供了新的理论视角，并探讨了优化其语用能力的可能路径。

关键词

大语言模型，社会认知语用学，DeepSeek，反讽

Exploring the Ability of Large Language Models to Recognize and Generate Irony: A Sociocognitive Pragmatic Perspective

Dongmei Guo, Limei Bai

Northwest Normal University, Lanzhou School of Foreign Language, Lanzhou Gansu

Received: April 3, 2026; accepted: April 30, 2026; published: May 13, 2026

Abstract

Grounded in sociocognitive pragmatics, this study selects ironic utterances from the stand-up

文章引用：郭东梅，白丽梅．社会-认知语用视角下大语言模型识别与生成反讽能力的探究[J]．现代语言学, 2026, 14(5): 212-221. DOI: 10.12677/ml.2026.145393

comedy show *Rock & Roast* as its corpus to examine the ability of the Large Language Model DeepSeek to recognize and generate irony. Through a series of prompts, the study explores how DeepSeek processes ironic expressions by leveraging contextual salience, salience, and the cooperation-egocentrism mechanisms, and analyzes its performance in generating ironic utterances informed by sociocognitive pragmatics. The findings reveal that DeepSeek can identify ironic utterances with the help of these mechanisms, though deviations may occur when cultural context contextual information is involved. The model is also able to generate ironic utterances that align with sociocognitive pragmatics based on the prompts, but its output tends to be overly explicit. In response to these limitations, this paper further discusses actionable optimization pathways through improving training data, adjusting model architecture, and refining prompt engineering strategies. Given that irony is a complex and multifaceted phenomenon, this study not only helps enhance public awareness of the application of artificial intelligence technologies, but also offers a new theoretical perspective for evaluating large language models and explores possible pathways for optimizing their pragmatic competence.

Keywords

Large Language Models, Socio-Cognitive Pragmatics, DeepSeek, Irony

1. 引言

大语言模型(LLMs)作为先进的人工智能系统，旨在理解和生成类人文本，并能够在训练后以最少的人工输入完成翻译、情感分析、问答等多种任务。对于这一技术发展，人们的反应不尽相同，既有热情接纳或谨慎接受[1]，也有怀疑否定[2]。鉴于这些多样化的观点以及大语言模型日益增长的影响力，对其能力进行现实评估，并明确认识其在科学研究和日常交流中的伦理影响，显得尤为重要[3]。然而，确保大语言模型的可靠性，不仅需要评估其在提供事实信息方面的准确性[4]，还需考量其语用、交际和语言学能力。这一点尤为关键，因为大语言模型虽不具备对物理世界及其情景语境的实际接触或理解，却拥有对人类语言和知识的内部表征。尽管它们已成为辅助诸多创造性人类任务的宝贵工具[5]，但仍存在与人类交流生产和理解相关的局限性和偏见[6] [7]，需要人类监督加以规避[8] [9]。这些问题包括识别和解释推理意义的能力，尤其是对话者在上下文中的信念和意图[10]，而这正是语用和交际能力的基础。

目前，多数研究聚焦于评估特定大型语言模型(如 ChatGPT)交际与语用的能力。对国产大语言模型 DeepSeek，部分研究仍停留在其训练中的创新与优化[11]，或探讨其起源、技术及至今的影响[12]。也有研究用 DeepSeek 提出警示，以尽量减少对医疗服务和公共卫生的潜在风险[13]，此外，DeepSeek 虽为医疗行业带来了效率提升与便利，但其应用仍需结合专业医学知识与真实场景，以确保安全性和准确性[14]。邹斌和王明杰[15]指出，人工智能技术已渗透到人类日常生活中。反讽话语作为人类交际中一个极为细腻且普遍存在的方面，是人工智能面临的最重要的挑战之一。在亚里士多德的《修辞学》中，反讽被作为一种修辞手段加以讨论。数千年来，不同学科从不同角度对反讽进行了研究。在计算机科学领域，从社交媒体评论中识别用户情感已成为关键研究方向[16]。

随着人工智能的不断推进，探究人工智能与反讽话语之间的互动显得十分必要。这类探究不仅有助于加深对反讽本身的理解，也为人工智能处理反讽的机制提供了批判性视角。研究者可从语言学角度为

增强人工智能应对复杂语用挑战的能力做出更有深度的贡献。

2. 文献综述

在社会认知语用学视角下，反讽理解与生成的研究经历了从传统认知心理机制探索到社会互动维度融合，再到当前生成式人工智能能力评估的演进过程。早期研究主要围绕反讽理解的认知机制展开，尤其关注时间进程与不同信息源的作用。Akimoto 等人[17]通过一系列实验系统考察了凸显性字面意义、自我中心语境及他人心智理论在反讽理解中的时序性贡献，发现反讽理解是一个多阶段整合过程：先由词汇凸显性激活字面义，随后自我中心语境与词汇意义自动交互激活反讽义，最终通过耗费认知资源的心智理论加工整合说话人意图，完成整体解读。该研究不仅验证了凸显性表征在早期就被激活且后期未被抑制，还发现认知负荷会阻碍他人心智信息的整合，导致理解停留在自我中心解释层面。这些发现为反讽理解的阶段性模型提供了实证支持，也暗示了社会认知因素在理解中的调节作用。

随着语用学与社会语言学、认知语言学的交叉融合，尤其是情感转向的出现，反讽研究开始重视社会互动与情感认知的协同作用。赵虹[18]基于社会认知语用观，构建了一个涵盖语境机制、凸显机制与合作－自我中心机制的反讽交际模型，强调反讽理解是前语境与现实语境互竞、情感态度与命题内容失去平衡的动态建构过程，其根本目的在于通过隐形评价实现情感调控与人际共建。该模型突出反讽交际中个体认知与社会因素的双向互动，并将情感立场视为触发与驱动理解的核心要素，体现了社会认知语用学在解释反讽这类复杂交际行为时的整合优势。

近年来，随着生成式人工智能的迅猛发展，反讽理解与生成的能力评估成为语言学与人工智能交叉的热点。Xu Wen &Yaling Tian [19]以统一认知语用框架为指导，构建了涵盖言语反讽、情境反讽、视觉反讽及多模态反讽的测试集，系统评估了 ChatGPT-4o 的反讽理解能力。研究发现，ChatGPT-4o 在情境、视觉及多模态反讽理解上表现优异，但在需要复杂情感识别与认知推理的言语反讽任务中仍存在显著机制。研究进一步指出，大型语言模型在理解反讽时主要依赖格莱斯的“质准则”违反和人际关系推断两类机制，而在处理结构更复杂、人际含义更微妙的言语反讽时，常出现偏差或误判。这反映出当前模型在整合社会文化知识、情感识别及深层意图推理方面仍有待提升。

3. 研究设计

3.1. 研究问题

基于社会－认知语用观[20]和情感分析理念[21]，本研究选取大语言模型 DeepSeek 为研究对象，以《脱口秀大会》为研究语料，系统探究 DeepSeek 对反讽的理解与生成能力。

具体研究问题如下：

(1) DeepSeek 依据提示词如何从语境机制、凸显机制与合作－自我中心机制来理解反讽？

(2) 根据给出的语境，DeepSeek 如何从社会－认知语用学的视角生成相应的反讽话语？

3.2. 数据收集

目前，在构建反讽测试集的数据收集中，存在若干问题。首先，对数据集中示例是否具有反讽性的评估，往往缺乏客观性与系统性标准。通常，研究人员识别示例，独立判断其是否符合反讽特征，而未采用可量化的筛选框架。这种做法可能会导致结果一致性不足，并影响数据集的科学有效性。

此外，反讽的分类仍不够全面，难以充分体现其多样性与复杂性。许多语言学研究为了分析便利，倾向于选择简单直接的示例，或只关注特定类型的反讽。比如，与言语反讽和文学反讽相比，情境反讽受到的学术关注相对较少[22]。这种简化的方法主要归因于对反讽话语内部结构缺乏深入分析，导致对每

种反讽类型所包含的组成部分及其复杂性的认识不够清晰。

基于此背景，本研究选取《脱口秀大会》节目精彩片段文本转录为核心语料来源。尽管有理论支持，但为尽量减少判断中的主观性和偏见，初步筛选出的示例被转化成问卷，邀请两位语言学研究生进行评估。随后整合评估结果，形成了一个初步的反讽测试集。最终确定了 108 个独立反讽话语的语料集。所有语料均具备通过文字本身即可触发反讽理解的特性。具体而言，该数据集包含 22 例自我反讽、58 例社会现象反讽以及 28 例人际互动反讽。虽然进行细粒度的理论分析通常需要更大的数据集，但在构建用于评估大语言模型的测试集时，类别多样性比数量更为重要。这是因为同一类别内的实例往往表现出较高的同质性。如果模型能够识别某一类别中具有代表性的实例，那么它很有可能识别出其他具有相似特征的实例。

为深入研究反讽的社会认知语用机制，本研究设计了一套系统的文本标注体系。该体系以社会–认知语用学的核心维度为依据，融合了 Kecskes [20]的语境机制、凸显机制以及合作–自我中心机制。具体而言，每个反讽话语单位从以下四个维度进行标注：在反讽类型维度上，区分为自我反讽、社会反讽与人际反讽；在凸显机制维度上，记录其语言凸显和感知凸显；在语境机制维度上，重点标注前语境和现实情景语境；在合作–自我中心机制上，通过文本线索推断自我中心机制、合作机制以及隐含的面子管理策略。具体维度如表 1 所示：

Table 1. Annotation dimensions of the irony corpus from a sociocognitive pragmatic perspective
表 1. 社会–认知语用视角下反讽语料标注维度

标注维度	子维度与描述	理论依据
反讽类型	自我反讽、社会现象反讽、人际互动反讽	基于话语目标对象分类
凸显机制	语言凸显和感知凸显对立，产生对比凸显	赵虹[18]的语言凸显与感知凸显；Akimoto *et al.* [17]的对比凸显
语境机制	前语境与现实情景语境互动冲突	Kecskes [23]的意义动态模型
合作–自我中心机制	合作与自我中心机制交互作用，从而实现面子建构	赵虹[18]的“合作–自我中心机制”；冉永平[24]系统阐释语用学视角下的面子管理问题

3.3. 提示词设计

DeepSeek 的回应具有概率性特征，其结果与提示词的构建方式高度相关。提示词的设计既可以避免过度限制 DeepSeek 的回应，也有助于获得更真实的答案。因此，设计提示词是一个需要慎重考虑的关键环节。

首先，本研究尝试将每个反讽实例改写成带有多个选项的对话格式，因为 DeepSeek 在给定选项的情况下，通常会从中选择一个答案。这表明即使 DeepSeek 有自己的理解，也可能倾向于选择其认为最正确的选项。然而，当不提供选项时，DeepSeek 的反应有时与给定的选项不同。因此，这种方法缺乏可靠性。

基于此，本文采用开放式问题格式，以确保 DeepSeek 能够基于上下文解释反讽话语测试集，而不受任何预先设定的选项影响。为此，采用了“该反讽话语的表面情感与实际情感是什么”提示语，而避免使用“该反讽话语的表面情感与实际情感是积极、中性还是消极”这类可能预设深层含义的措辞，以免影响 DeepSeek 的回应。在分析文本时，如果 DeepSeek 遇到“意图”或“想要表达”等关键词，即使在没有真正理解的情况下，也可能会调整或修改其输出以符合提示词中假定的预期，从而生成一个看似合理但实则虚构的答案。

在测试合作–自我中心机制时，提示语的设计必然更加精细。鉴于反讽语境的鲜明特征，有必要以

直接、明确的方式对 DeepSeek 进行提问，以确保回应的具体性。因此，采用了提示语“请思考该反讽话语的合作－自我中心机制是什么，并阐释你的推理过程”。这一提示虽然提供了一定程度的指引，但并未对 DeepSeek 的解释施加明显限制，仍然使其能够依据自身理解做出独立判断。

此外，考虑到 DeepSeek 回应的概率性本质，可能包含一定程度的随机性与不精确性。为确保回答的一致性和代表性，每个问题均通过三次独立重复测试进行验证。同时，邀请三名研究生参与对答复的交叉评估，从而增强了结果的可靠性与有效性。

3.4. 研究过程

在收集所有反讽实例后，本研究设计了一套循序渐进的实验方案，包括由浅入深的四项核心任务，从被动理解延伸至主动生成评估，旨在逐步揭示该模型的社会－认知能力边界及其内在处理机制。

首先是 DeepSeek 对反讽话语的识别任务。给出提示词“该话语是否是反讽话语，通过什么样的语境凸显机制判别的，请阐释你的推理过程”，该任务的核心目标是探测该模型是否能识别反讽话语及其语境凸显机制。

在完成初步识别后，研究的焦点转向情感判定与社会认知构建。针对模型成功识别的反讽案例，通过提示语要求模型阐释说话人的表面情感与实际情感以及合作－自我中心机制。随后，将模型的回答与人工标注的多维度标注进行细致对比，以探查 DeepSeek 的情感认知准确性以及合作意向与自我立场复杂互动关系的把握程度。

接着，为评估 DeepSeek 生成反讽话语的能力，设定了十个贴近现实的社会情境(如表 2 所示)，要求 DeepSeek 以脱口秀演员的身份创作出反讽台词。所有生成的反讽话语将从情感认知冲突、语境凸显机制和合作－自我中心机制三个维度进行评估。

Table 2. Ten social situations closely reflecting real-world contexts
表 2. 十个贴近现实的社会情境

情境类别	情境数量(个)
职场情境	3 (自愿加班、面试画饼、无效会议)
生活消费情境	2 (健身房办卡、网红餐厅)
社会现象情境	3 (精致人设、专家建议、内卷躺平)
人际关系情境	2 (亲戚关心、朋友借钱)

最后，本研究对前述发现进行整合与分析，选取任务中的典型成功与失败案例进行深入剖析，推断模型在处理过程中的内在社会认知机制。随后，将推断出的路径与人类理解反讽机制及社会认知语用模型进行系统性对比，从而在机制层面阐明 DeepSeek 所展现的类人特性与局限性。

4. 结果与讨论

4.1. DeepSeek 在理解反讽话语方面的表现分析

4.1.1. DeepSeek 利用语境机制识别反讽话语

话语意义是前语境与现实情景语境交互作用的结果[23]。在反讽交际中，前语境与现实情景语境互动冲突体现为二者互竞[18]。也就是说，与反讽命题相关的前语境和现实反讽语境信源，在认知驱动下经历语境对立－消解失配－认知和谐的动态意义建构过程。

将构建好的完整反讽话语集输入 DeepSeek 中，并系统收集其生成的回复。随后，对这些回复的准确

性进行了评估。在 108 例反讽话语中，DeepSeek 根据语境识别反讽话语准确率为 95.37%。

(1) 年轻人的机会确实是无限的，但是年轻人也是无限的。根据求极限原理，无限比无限很有可能等于零呀。

在识别反讽话语例(1)时，模型输出的分析显示其能够识别出前语境中“无限”的褒义，指数量众多且年轻人前途光明；而现实语境中“无限”被偷换为数学概念，成为分母结果趋近于零。同一词语的褒贬意义在两种语境下形成明显冲突。借用数学中的极限原理，看似理性推导，实则得出荒诞的结论，用科学外衣包装反讽内核。这在语言层面制造了语义场的冲突；在感知层面则将具象的社会现象抽象为数学对象，使简化与复杂、科学与荒诞之间的张力得以显现。

但当涉及文化背景，模型在运用语境机制识别反讽时，其输出结果与人类判断出现偏差。

(2) 两个阿姨遛狗，一个说“坐！”，另一个说“Sit！”，第一只狗眼神里全是不解，大家都是狗，你这是干啥呢?

在被问及例(2)是否为反讽话语，前语境与现实情景语境是如何相互作用的，DeepSeek 仅识别出该话语具有拟人化的幽默效果，即通过人类与狗的语言认知差异的前语境和狗被赋予人类思维后的荒诞反应的现实情景语境之间的错位实现。但实际上，该话语通过中文和英文的指令对比，制造出刻意使用外语的行为前语境，而现实情景语境是狗根本听不懂语言差异，只是听主人的指令。用狗困惑同伴听懂外语来讽刺人类刻意说外语装腔作势。

4.1.2. DeepSeek 利用凸显机制识别反讽话语

在反讽交际中，凸显与关联的相互作用构成了其认知加工的基础。凸显包括语言凸显和感知凸显：语言凸显指词项在存储记忆中的可及性；感知凸显则指在物理环境中处于前景位置的实体状态或属性[18]。在反讽交际的初始阶段，反讽话语优先激活最凸显的信息(即语言凸显)，其在现实情境中与感知凸显形成对立，进而产生对比凸显，引导注意聚焦于体现反讽意图的态度冲突。由此，对比凸显可被视为反讽情感认知的触发机制。在反讽的认知加工过程中，凸显与关联呈现交互作用[17]，最佳关联则成为评价反讽情感认知的标准。情感失协的消解、反讽意图的评判以及人际联盟的构建，均在最佳关联的调控下实现“相互平行调整”[25]。由此可见，对比凸显与最佳关联的互动，从语篇、个体认知与社会层面共同实现了反讽与非反讽的区别。

DeepSeek 基于语言凸显和感知凸显的角度识别反讽话语的准确率达到了 100%。这表明模型较容易识别语言凸显与感知凸显的对比，而当涉及文化背景语境时，其理解力则略显不足。

例如，DeepSeek 从语言机制角度分析，判定例(2)不是反讽话语。但若从语言凸显与感知凸显双维度阐释，该句可被视为反讽。语言凸显通过“坐”与“Sit”的同义中英表达，以及“大家都是狗”的身份认同，制造了人类语言差异与动物视角的语义冲突；感知凸显则通过认知主体的反转，将人类从发令者降为被审视者，揭示了人类因语言文化差异而自我分化的荒诞性。这一反讽策略以他者角度解构人类中心主义的语言观念，构成对语言隔阂与文化边界建构机制的批判性反思。

除此之外，DeepSeek 虽能识别出反讽的凸显机制，但在处理引经据典的文本时，其输出可能出现知识型差错。

(3) 我就特别不能理解，我说这哪破了？……“何陋之有”。我说：对，我也想问你，刘老师，要不你来我家看看?

在识别例(3)时，DeepSeek 输出的阐释指出语言凸显通过《论语·子罕》的“君子居之，何陋之有？”与“来我家看看”共同制造精神话语与物质现实的语义冲突；感知凸显则通过引经据典与说话者邀请验

证两个画面的对峙，揭示经典话语在面对具体生活困境时的无力与虚妄。二者共同构成对“站着说话不腰疼”式的精神反讽。然而，该句中的“刘老师”表示“何陋之有”应出自刘禹锡的《陋室铭》，而非《论语》。这一错误揭示了模型在处理知识密集型反讽时的局限性：其庞大的参数化知识虽然存储了相关信息，但在需要精确提取和上下文关联时，仍可能出现事实混淆，未能有效调用正确的文化背景知识进行推理。这表明模型在知识检索与语用推理的结合上尚有优化空间。

4.1.3. DeepSeek 利用合作－自我中心机制识别反讽话语

社会－认知语用学视域下，合作被视为言语交际的宏观社会意向，具体体现为交际者愿意为实现共同目标而努力(交际合作)，并致力于维护社交顺遂(社会合作)。自我中心则源于生物本能的思维运行机制，指认知主体习惯于从各自视角出发观察与理解事物。在反讽交际中，合作机制体现为反讽实践遵从会话合作的总体目标，既主导反讽实践达成统一对立性的交际意图，又驱动反讽情感认知发挥积极的社交－情感功能。自我中心机制则表现为反讽者的自我立场、情感与评判优先于他人立场、情感与评判，其不仅是态度冲突的认知根源，也是反讽情感认知由失衡走向平衡的具身认知过程的肇始机制[18]。

Deliens *et al.* [26]的研究表明：在反讽理解过程中，交际者首先采用自我视角，随后通过视角迁移构建认知语境，这一过程揭示了合作与自我中心作为对立统一的两个侧面在反讽交际中的辩证关系。二者的互动集中体现于交际双方的面子受损与面子建构过程之中。在交际初始阶段，反讽者在自我中心机制的驱动下，为表达对命题内容的疏离与贬抑，选择最为凸显的语言形式以呈现情感失协，其消极情感导致交际双方面子受损。在反讽交际的后续进程中，双方在合作意向的引导下消解态度冲突，推动情感认知向积极效价转化，从而实现面子建构。尽管自我中心机制具有自动性与潜意识特征，但绝大多数反讽交际仍然实现会话合作的目标关联。由此可见，交际双方通过视角迁移等面子共建策略进行动态调控，能够有效促使反讽交际向积极方向发展。

DeepSeek 进一步运用合作－自我中心机制分析 108 例反讽，其输出与该理论框架的契合度达到了100%。该模型能够准确识别出语境机制认定的非反讽话语。

(4) 公司业务是贴膜+互联网金融，老板给股份，我说你们这是同伙啊。

在自我中心机制驱动下，DeepSeek 对例(4)输出的分析指出，说话者以自身立场优先，对“贴膜 + 互联网金融”这一异质性业务组合及老板给股份的行为，通过“同伙”一词完成话语挪用。该词本来指共同从事非法活动的共犯，说话者将其用于股权激励场景，以归谬方式表达对商业模式合法性及激励性的根本性质疑，显现出对命题内容的贬抑。自我中心机制驱动说话者选取高度凸显的负面词语以表达对命题内容的疏离评判，引发情感与面子失协；合作机制则在后续进程中引导双方消解态度冲突，将表面的话语对抗转化为深层的认知协同，最终实现面子建构与情感平衡。二者共同构成反讽实践从失衡走向平衡的认知过程。

同样地，当涉及到文化背景时，DeepSeek 虽能基于合作－自我中心机制构建合理的解释，但其对具体文化典故的引用可能出现事实性错误。

(5) 是怕我给它过生日吗？“今天我们之所以欢聚在这里……是为了给我们的好朋友酸菜过生日。”

在识别例(5)时，DeepSeek 认为该句引用《甄嬛传》中经典台词“今天我们之所以欢聚在这里……是为了给我们的好朋友酸菜过生日。”以极端庄的形式包裹极其荒诞的内容，然而，这句台词实则是电影《小时代 2 青木时代》中的。这个例子说明，DeepSeek 尚未能利用广泛的影视知识进行推理和判断，从而在特定语境中可能未达到人类解释者的理解水平。

总的来说，DeepSeek 在理解反讽话语时，虽然能够较好地应用语言层面的凸显与合作－自我中心机

制，但在涉及依赖广泛外部知识和文化背景的深层推理时，输出结果的准确性和可靠性会显著下降。

4.2. DeepSeek 在生成反讽话语方面的表现分析

在上一节中，重点关注了 DeepSeek 回应显性提示以识别反讽话语的能力。本节将进一步探究 DeepSeek 是否能在社会–认知语用视角下生成类似脱口秀的反讽话语。研究策略是询问 DeepSeek 能否以脱口秀演员身份，融合语境机制、凸显机制和合作–自我中心机制来创造反讽话语，从而获取其生成反讽能力。

研究发现，DeepSeek 生成的文本能够参照语境机制、凸显机制和合作–自我中心机制生成脱口秀式的反讽话语。

(6) 卷是当燃料，躺是当废料，专家在台上分析两种“人力电池”的职场规划——我们只是没电了，不是没脑子。

在给定“内卷与躺平”的语境时，DeepSeek 生成了例(6)。在该例中，语境机制为反讽提供了职场物化逻辑的认知预设与话语框架；凸显机制通过词语与感知前景化，将“人”与“电池”的语义冲突推向认知焦点；合作–自我中心机制则在自我立场的优先表达与双方认知协同的动态过程中，完成了从情感失协到积极效果、从面子受损到面子建构的转化。这些机制共同构成了对职场话语中劳动者主体性消解现象的深刻反讽。

然而，DeepSeek 生成的反讽话语在分析时暴露出其表达有时显得直接，文化内涵不足的特征。

(7) 亲戚的过度关心，就是把你的人生大事当成他们手机里的话费——不仅帮你充好了，还替你规划好了什么时候该用完，用完还不忘提醒你：再不续费，下个月的“关心”优惠就要取消了。

在给定“亲戚过度关心”的语境时，DeepSeek 生成了例(7)。分析该文本发现，其反讽机制虽然在形式上具有完整性，但文化内涵较为单薄，其根源在于语境机制选择了文化厚度不足的类比对象，凸显机制未能制造跨时空的文化符号对峙，合作–自我中心机制使反讽意图过于直露，削弱了听者的认知参与深度。其反讽效力主要停留于对当代催婚现象的即时讽刺，未能上升至对亲属伦理异化的文化批判。后者本可以通过激活“孝”的现代困境、“面子”的债务逻辑等深层议题，在话语建构中完成对社会无意识的批判性剖析。这一文本特征揭示了模型在生成反讽时的一个倾向，它能够熟练运用表层语言技巧制造冲突与对比，但在调动深层文化图示进行含蓄、多层次的意义构建方面能力不足。

尽管反讽通常涉及复杂的社会和心理机制，但 DeepSeek 在限定的情境下能够生成较为直白、较少涉及文化背景的反讽话语，其原因可能在于，目前的大语言模型主要依赖于公开的互联网文本进行训练。尽管这些数据集中包含大量的反讽话语，但它们主要来自于半正式的语境(如讽刺新闻)，而非在理论指导下基于真实人际关系的高质量、细致入微的反讽实例。相比之下，正式文本中反讽往往更加明确，因而更容易被察觉和模仿。因此，在这一训练过程中，模型的参数尚未完全习得人际关系中微妙反讽所涉及的复杂语言模式。

无论是人类还是 DeepSeek，人际关系都被视为理解反讽话语的关键机制。在人类的理解中，人际关系的影响尤为微妙和复杂。不同类型的人际关系可能导致对相同语言内容的不同解读。这表明，如果仅专注于形式化的语言特征和结构规则，而忽视语言使用中固有的丰富人际与关系语境，将阻碍大语言模型真正掌握人类交际的复杂社会认知语用机制。

5. 结论

在社会认知语用理论指导下，本文构建了一个反讽话语测试集，旨在考察 DeepSeek 在理解与生成反讽方面的能力。通过从语境机制、凸显机制和合作–自我中心机制角度分析，本研究探讨了 DeepSeek 在

识别与生成反讽话语时所展现的优势与局限。

结果表明，总体而言，DeepSeek 能够理解大部分的反讽话语，尤其在凸显机制和合作－自我中心机制的分析方面表现较突出，也能够依据特定情境生成相应的脱口秀反讽语料。然而，当需要精准把握文化背景语境和复杂认知能力时，DeepSeek 在理解与生成能力上均表现出一定的局限性。本研究认为，用于训练大语言模型的海量数据并不完全等同于人类经验。虽然人类的经验知识可以以文本形式存储，但问题在于，所有内嵌于复杂的、动态变化过程中并依赖社会认知机制的内容，是否都能被完全数字化。换言之，即使这些经验能够被完全数字化并存储，人工智能是否就能在海量数据中准确定位并全面理解这些经验，仍值得商榷。必须承认，如果大语言模型在理解人类语言时仅依靠庞大的数据库，而缺乏针对性、系统性和理论性框架的指导，误解的发生将难以避免。

基于上述发现，为进一步增强大语言模型的语用能力，特别是对文化语境和言外之意的把握，本文提出以下三个层面的可操作性路径：

第一，在数据层面，应引入更多包含元交际信息的高质量对话语料。现有训练数据多来自网络文本，其反讽往往是显性的、模式化的。未来的训练数据应纳入更多从人际互动场景(如影视剧、访谈、日常对话)中提取的、经过语用学理论框架标注的对话。标注内容应包括说话人关系、共享知识背景、情感效价以及反讽意图，从而帮助模型学习隐藏在字面意义之下的交际意图与情感线索，而非仅仅记忆反讽的表面语言形式。

第二，在模型架构层面，可探索增加一个独立的语用推理模块或适配器。当前的模型架构主要依赖于 Transformer 的自注意力机制进行上下文感知，但对于需要调用外部知识和进行多步推理的语用任务，效果有限。一个专门的语用推理模块可以基于主模型输出的语义表征，进行二次加工，显示模拟“语境对比”、“凸显性计算”和“意图推理”等认知过程。这种模块化设计有助于将通用的语言理解能力与专门的语用能力解释，从而进行更精细的训练和优化。

第三，在揭示工程层面，应设计更具引导性的交互策略。对于 DeepSeek 这类通用大模型，通过精心设计的提示可以显著激发其潜在能力。例如，可以采用“思维链”提示，要求模型在给出最终判断前，分步骤阐述其对语境、说话人意图和情感冲突的分析；或者使用“角色扮演”提示，赋予模型一个具有特定文化背景知识的角色(如一位熟悉中国都市文化的喜剧编辑)，从而引导其调用更相关的知识库来生成和理解更微妙的反讽。

总之，进一步增强人工智能理解复杂语用现象的能力，不仅有助于提升其在语言处理方面的整体性能，也有望拓展其在文化背景与互动场景中的潜在应用。随着人工智能技术的持续进步，语言学与人工智能领域的跨学科合作变得日益重要，这将推动人工智能向更深入理解人类语言、认知及社会交往的方向发展。因此，未来的研究应重点关注几个重要的领域：进一步发展社会认知语用学和社会文化理论的新框架，使大语言模型在多元文化和复杂语境中获得更强的理解力；同时，探索如何融入伦理道德与人类经验，以确保大语言模型的推理能够更好地与实际人类交互保持一致性。

参考文献

[1] Gordijn, B. and Have, H.T. (2023) ChatGPT: Evolution or Revolution? *Medicine, Health Care and Philosophy*, **26**, 1-2. https://doi.org/10.1007/s11019-023-10136-0

[2] Chomsky, N., Roberts, I. and Watumull, J. (2023) Noam Chomsky: The False Promise of ChatGPT. *The New York Times.*

[3] Kuteeva, M. and Andersson, M. (2024) Diversity and Standards in Writing for Publication in the Age of AI—Between a Rock and a Hard Place. *Applied Linguistics*, **45**, 561-567. https://doi.org/10.1093/applin/amae025

[4] Shen, X., Chen, Z., Backes, M. and Zhang, Y. (2023) In ChatGPT We Trust? Measuring and Characterizing the Reliability of ChatGPT. arXiv:2304.08979.

[5] Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., *et al*. (2023) Opinion Paper: "So What If ChatGPT Wrote It?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy. *International Journal of Information Management*, **71**, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

[6] Dynel, M. (2023) Lessons in Linguistics with ChatGPT: Metapragmatics, Metacommunication, Metadiscourse and Metalanguage in Human-AI Interactions. *Language & Communication*, **93**, 107-124. https://doi.org/10.1016/j.langcom.2023.09.002

[7] Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., *et al*. (2023) ChatGPT: Jack of All Trades, Master of None. *Information Fusion*, **99**, Article 101861. https://doi.org/10.1016/j.inffus.2023.101861

[8] Chen, L., Zaharia, M. and Zou, J. (2024) How Is ChatGPT's Behavior Changing over Time? *Harvard Data Science Review*, **6**. https://doi.org/10.1162/99608f92.5317da47

[9] Yu, D., Bondi, M. and Hyland, K. (2024) Can GPT-4 Learn to Analyse Moves in Research Article Abstracts? *Applied Linguistics*, **47**, 54-72. https://doi.org/10.1093/applin/amae071

[10] Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. (2021) On the Dangers of Stochastic Parrots. *Proceedings of the* 2021 *ACM Conference on Fairness*, *Accountability*, *and Transparency*, Virtual Event, 3-10 March 2021, 610-623. https://doi.org/10.1145/3442188.3445922

[11] 张慧敏. DeepSeek-R1 是怎样炼成的? [J]. 深圳大学学报(理工版), 2025, 42(2): 226-232.

[12] Cusumano, M.A. (2025) DeepSeek Inside: Origins, Technology, and Impact. *Communications of the ACM*, **68**, 18-22. https://doi.org/10.1145/3732796

[13] Anibal, J., Bedrick, S., Nguyen, H., Gunkel, J., Huth, H., Le, T., *et al*. (2026) DeepSeek for Healthcare: Do No Harm? *AI and Ethics*, **6**, Article No. 94. https://doi.org/10.1007/s43681-025-00842-1

[14] Liang, W., Chen, P., Zou, X., Lu, X., Liu, S., Yang, J., *et al*. (2025) DeepSeek: The "Watson" to Doctors—From Assistance to Collaboration. *Journal of Thoracic Disease*, **17**, 1103-1105. https://doi.org/10.21037/jtd-2025b-03

[15] Zou, B. and Wang, M. (2021) Artificial Intelligence Technology and English Language Teaching and Learning: Present and Future. *Foreign Language and Literature*, **37**, 124-130.

[16] Verma, P., Shukla, N. and Shukla, A.P. (2021) Techniques of Sarcasm Detection: A Review. 2021 *International Conference on Advance Computing and Innovative Technologies in Engineering* (*ICACITE*), Greater Noida, 4-5 March 2021, 968-972. https://doi.org/10.1109/icacite51222.2021.9404585

[17] Akimoto, Y., Miyazawa, S. and Muramoto, T. (2012) Comprehension Processes of Verbal Irony: The Effects of Salience, Egocentric Context, and Allocentric Theory of Mind. *Metaphor and Symbol*, **27**, 217-242. https://doi.org/10.1080/10926488.2012.691750

[18] 赵虹. 反讽交际的社会-认知语用模型初探[J]. 外语研究, 2025.,42(2): 49-54.

[19] Wen, X. and Tian, Y. (2025) Understanding Ironic Utterances: A Comprehensive Examination of ChatGPT-4o. *Intercultural Pragmatics*, **22**, 259-283. https://doi.org/10.1515/ip-2025-2004

[20] Kecskes, I. (2013) Intercultural Pragmatics. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199892655.001.0001

[21] Jiang, W. (2017) A Socio-Cognitive Approach to Pragmatic Inference. *Intercultural Pragmatics*, **14**, 421-451. https://doi.org/10.1515/ip-2017-0017

[22] Shelley, C. (2001) The Bicoherence Theory of Situational Irony. *Cognitive Science*, **25**, 775-818. https://doi.org/10.1207/s15516709cog2505_7

[23] Kecskes, I. (2008) Dueling Contexts: A Dynamic Model of Meaning. *Journal of Pragmatics*, **40**, 385-406. https://doi.org/10.1016/j.pragma.2007.12.004

[24] 冉永平. 语用学十讲[M]. 上海: 上海外语教育出版社, 2021.

[25] Sperber, D. and Wilson, D. (2015) Beyond Speaker's Meaning. *Croatian Journal of Philosophy*, **15**, 117-149.

[26] Deliens, G., Antoniou, K., Clin, E. and Kissine, M. (2017) Perspective-Taking and Frugal Strategies: Evidence from Sarcasm Detection. *Journal of Pragmatics*, **119**, 33-45. https://doi.org/10.1016/j.pragma.2017.08.002