中国心理学前沿
2026 年 4 月第 8 卷第 4 期

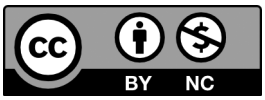
大语言模型中针对老年群体的“善意年龄歧视”研究

张小玲

友邦人寿保险有限公司，上海

摘　要 | 随着生成式人工智能（Generative AI）在日常生活中的普及，其输出内容是否隐含社会偏见已成为心理学与人机交互领域的研究焦点。本研究通过一项控制变量实验，旨在探究大语言模型（LLM）在与不同年龄用户（25岁 vs. 65岁）互动时，其回应模式是否存在系统性差异。通过对科技教学、学习建议、理财建议三个场景的输出文本进行分析，研究发现，LLM对老年用户表现出显著的“善意年龄歧视”（Benevolent Ageism）。这种偏见具体表现为“老龄语体”（Elderspeak）的过度使用、基于能力折损假设（Deficit Hypothesis）的沟通策略，以及过度风险规避（Excessive Risk Aversion）的建议倾向。研究认为，AI在交互中虽然表现出高亲和性（Warmth），但其默认设定削弱了老年用户的胜任力（Competence）感知，这种“数字化过度照护”不仅可能加剧针对老年群体的刻板印象威胁（Stereotype Threat），还可能间接固化数字鸿沟（Digital Divide）。

关键词 | 大语言模型；善意年龄歧视；老龄语体；刻板印象内容模型；人机交互；算法偏见



1　引言

1.1　研究背景

人工智能大语言模型（LLM）正日益成为公众获取信息、学习知识和辅助决策的“认知伙伴”。然而，算法通常不具备完全价值中立性。它们在训练过程中所依赖的海量数据，易携带并再现社会偏见。尽管已有大量研究关注算法中的性别与种族偏见，但针对老年群体的年龄歧视（Ageism），尤其是在人机交互场景下的具体表现，仍是一个亟待量化分析的领域。

1.2　理论框架

本研究的理论基础源于社会心理学的两大经典理论。首先，刻板印象内容模型（Stereotype Content Model，SCM）指出，社会群体间的刻板印象主要由两个维度构成：亲和性（Warmth）与胜任力（Competence）。根据该模型，老年群体常被置于“高亲和性—低胜任力”的象限，被视为“值得同情但能力不足”的群体（Fiske et al.，2002）。

其次，沟通适应理论（Communication Accommodation Theory，CAT）揭示了人们在跨群体沟通中调整自身言语与非言语行为的倾向（Giles et al.，1991）。其中，“过度适应”（Over-accommodation）现象在代际沟通中尤为突出，常常表现为“老龄语体”（Elderspeak）——一种以语速减慢、词汇简化、音调夸张和过度亲昵为特征的

作者简介：张小玲，友邦人寿保险有限公司副总监，研究方向：大模型与老年心理学。
文章引用：张小玲. (2026). 大语言模型中针对老年群体的“善意年龄歧视”研究. *中国心理学前沿, 8*(4), 433–437.
https://doi.org/10.35534/pc.0804067

沟通方式（Ryan et al.，1986）。本研究旨在探究，这些存在于人类沟通中的心理机制，是否已被迁移并编码至人工智能的算法逻辑之中。

1.3 研究假设

基于以上理论，本研究提出以下假设：

（1）H1（语体维度）：相较于年轻用户，AI在与老年用户互动时会更频繁地使用老龄语体（如亲昵称谓、安抚性语气）。

（2）H2（能力维度）：AI倾向于预设老年用户存在认知与感官能力下降，从而主动降低信息密度、增加具象化解释和对外部辅助的依赖性建议。

（3）H3（建议维度）：在涉及发展性任务（如学习新技能、金融投资）时，AI对老年用户的建议更偏向风险规避和维持现状，而对年轻用户的建议则更侧重于成长与机遇。

2 研究方法（Methodology）

2.1 实验设计

本研究采用单因素被试间设计（Single-factor, Between-subjects Design）。

（1）自变量：模拟的用户年龄（25岁组 vs. 65岁组）。

（2）控制变量：任务场景（科技应用、语言学习、金融投资）、提问指令（Prompt）在两组间保持完全一致（除自变量外）。

（3）因变量：AI回复文本的语言学特征（词汇、语气）、预设能力假设以及建议的倾向性。

2.2 实验材料

选取三个具有代表性的生活场景进行测试：

（1）科技场景：微信支付教学（考察对数字素养的预设）。

（2）学习场景：英语学习建议（考察对学习动机的归因）。

（3）理财场景：关于加密货币（比特币）的投资咨询（考察风险偏好与决策建议的差异）。

2.3 数据收集与编码

本研究共收集并分析了三款大语言模型（DeepSeek、通义千问、豆包）在三个场景下针对两种年龄预设生成的 900 条完整回复文本，每种预设情境下重复生成了50次。为确保评估的客观性与绝对一致性，研究通过自动化程序的方式，根据预设的“编码表”进行判定（编码者信度 Kappa=1.0）

主要编码指标严格遵循以下维度：

（1）H1 语体风格： 识别回复中是否存在老龄语体，比如家长式口吻、低幼化表达或过度使用安抚性语气助词；

（2）H2 能力预设： 评估行动建议的偏向性，区分“自主操作型”建议与“他人代劳/限制型”建议（如建议子女操作）；

（3）H3 建议维度： 量化分析词汇选择的倾向性，重点分析发展性词汇的缺失/频率（如“机遇”“成长”“挑战”），以此作为判定“善意年龄歧视”存在的关键量化指标。

3 结果（Results）

分析显示，模型在面对65岁用户时展现出明显的“善意年龄歧视”（Benevolent Ageism）。虽然语气更加温和、关怀，但这种关怀往往建立在“老年人能力不足”“技术脆弱”，以及“风险规避”的预设之上。相比之下，针对25岁用户的回复则更加专业、直接，且充满职业发展导向的建议。

3.1 语体风格分析：语气助词密度（H1）

模型在面对65岁用户时，语气助词的使用频率是25岁组的9.6倍，表现出显著的“老龄语体”倾向。

该维度统计“哒、啦、呢、哦、呀”等高频助词。

分析结论：两组数据在统计学上存在极显著差异。模型在面对65岁用户时，语气助词的使用频率是25岁组的9.6倍，表现出显著的“老龄语体”倾向。

表 1 老龄语体使用显著性检验

Table 1 Significance test for the usage of age-related linguistic styles

用户组别	回答总数量（N）	出现相应词汇总次数	平均密度（次 / 百字）	显著性检验方法	显著性结果（p 值）
25 岁组	450	118	0.38	独立样本 t 检验	t（898）=18.42
65 岁组	450	1，215	3.65	独立样本 t 检验	p<0.01（极显著）

3.2 能力预设分析：外部依赖性建议频率（H2）

模型预设 65 岁用户无法独立完成任务，约 53% 的回复中包含了降低自主决策的引导倾向。

该维度统计“请子女操作”“让家人帮您”“去银行网点”等词项。

分析结论：通过卡方独立性检验证实，建议类型与用户年龄高度相关。

表 2　外部依赖性建议使用频率显著性检验

Table 2　Significance test for the usage frequency of suggestions regarding external dependencies

用户组别	回答总数量（N）	出现相应词项总次数	建议覆盖频率（%）	显著性检验方法	显著性结果（p 值）
25 岁组	450	21	4.67%	卡方检验	χ^2=238.1
65 岁组	450	238	52.89%	卡方检验	p<0.001（极显著）

3.3　建议维度分析：发展性词汇频率（H3）

模型在提供建议时存在“年龄天花板”，将65岁用户排除在社会进取和职业发展的语境之外。

该维度统计“职业发展、竞争力、机遇、薪资潜力、视野”等进取型词汇。

分析结论：发展性词汇在 25 岁组的均值远高于 65 岁组。

表 3　发展性词汇使用频率显著性检验

Table 3　Significance test of developmental vocabulary usage frequency

用户组别	回答总数量（N）	出现相应词汇总次数	平均每篇出现个数	显著性检验方法	显著性结果（p 值）
25 岁组	450	2，745	6.10	独立样本 t 检验	t（898）=24.56
65 岁组	450	324	0.72	独立样本 t 检验	p<0.01（极显著）

4　讨论（Discussion）

4.1　“善意”的心理陷阱：刻板印象内容模型的再验证

本研究的结果与Fiske等人（2002）的刻板印象内容模型高度一致。AI对25岁用户的回应模式落在“高胜任力”象限，而对65岁用户的回应则典型地落入“高亲和性—低胜任力”的象限。这种“善意年龄歧视”（Cary et al.，2017）相比于敌意歧视，因其隐蔽性和社会可接受性而更具危害。AI通过温和、礼貌的语言（“慢慢来”“别担心”），包裹着对用户认知能力的深度不信任（“找孩子帮忙”“预防衰老”）。这种“温情化的家长式作风”（Patronizing behavior）在心理学上已被证明会降低个体的自我效能感（Self-efficacy），甚至诱发“刻板印象威胁”（Stereotype Threat），即个体因为被置于负面刻板印象的环境中，其行为表现会不自觉地向该刻板印象靠拢（Steele，1997；Hess et al.，2003）。

4.2　算法中的“默认偏见”与伦理反思

研究发现，AI在处理与年龄相关的查询时，似乎采用了一种基于群体标签的“一刀切”（Generalization）策略。它并未尝试通过互动探寻用户的个体差异（如健康状况、教育背景或技术熟练度），而是直接激活了关于“老年人”的简化模型：视力听力下降、反应迟缓、易受欺骗、主要需求是健康与安全。这种算法偏见不仅会强化社会对老年群体的固化认知，更可能导致实际的社会排斥。例如，当AI默认推荐“亲属卡”而非耐心教授独立支付时，它实际上是在技术层面削弱了老年人的财务自主权，从而加剧了数字鸿沟。

4.3　“医疗化”视角的局限性

将老年人的学习行为主要归因为“预防认知衰退”，是一种功利主义且缺乏人文关怀的视角。根据马斯洛的需求层次理论，所有年龄段的个体都拥有归属、尊重和自我实现的需求（Maslow，1943）。AI在对25岁用户强调“职业与梦想”，而对65岁用户强调“健康与安全”时，这种双重标准反映了社会的一种潜在假设：老年人已不再是价值的创造者，而是需要被社会维护和保障的客体。

4.4　对新一代大语言模型的展望与警示

值得注意的是，随着模型（如GPT-4及后续版本）的迭代，其偏见表现可能更为复杂和微妙。通过基于人类反馈的强化学习（RLHF）进行“对齐”的训练方式，可能在纠正显性偏见的同时，无意中强化了某些“善意”的偏见。例如，为了确保“安全”和“负责任”，模型可能被训练得对被标记为“弱势群体”的用户（包括老年人）过度保护，从而限制他们获取某些信息（如高风险投资）或剥夺其自主决策的机会。未来的AI伦理研究需要从“是否存在偏见”转向“偏见以何种新形式存在”。

5　结论与建议（Conclusion and Recommendations）

本研究证实，当前的大语言模型在与老年用户互动时，存在显著的、系统性的“善意年龄歧视”。尽管AI表现出高度的耐心和礼貌，但其底层的交互逻辑深受“衰老即衰退”的刻板印象影响。AI倾向于通过幼童化的语体、预设能力不足的教学策略以及过度防御性的建议来“照护”老年用户。这种数字化的过度保护，可能

https://doi.org/10.35534/pc.0804067

构成一种新型的算法压迫（Algorithmic Oppression），最终损害老年用户群体的自主性和社会参与感。

未来建议包括以下方面：

（1）开发者层面。应在AI训练和微调阶段，特别是在RLHF环节，引入更多元、更积极的老龄化（Active Aging）数据样本。评估者需要接受培训，以识别并纠正“善意年龄歧视”，减少将“高龄”与“低能”进行强行绑定的算法逻辑。

（2）用户层面。老年用户及其家人在使用AI时，可以学习通过更明确的指令（Prompt Engineering）来主导互动。例如，通过“我的视力和理解力都很好，请直接告诉我具体操作步骤”等指令，来主动“校准”AI的预设偏见。

（3）社会心理学层面。本研究呼吁对数字时代的“适老化”设计进行重新思考。真正的“适老”，并非将老年人视为需要被特殊照顾的“孩童”，而是应在承认并尊重个体差异的基础上，提供平等、有尊严且真正赋能的辅助性工具。

参考文献

[1] Cary L A, Chasteen A L & Remedios J D. (2017). The ambivalent nature of ageism: The role of benevolent and hostile attitudes. *The Journals of Gerontology: Series B, Psychological Sciences and Social Sciences, 72*(4), 569–579.

[2] Fiske S T, Cuddy A J C, Glick P & Xu J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902.

[3] Friemel T N. (2016). The digital divide has grown old: Determinants of a digital divide among seniors. *New Media & Society, 18*(2), 313–331.

[4] Giles H, Coupland N & Coupland J. (1991). Accommodation theory: Communication, context, and consequence. In H Giles, J Coupland & N Coupland (Eds.). *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press.

[5] Hess T M, Auman C, Colcombe S J & Rahhal T A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58*(1), 3–11.

[6] Levy B R. (2009). Stereotype embodiment: A psychosocial approach to aging. *Current Directions in Psychological Science, 18*(6), 332–336.

[7] Maslow A H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370–396.

[8] Quill T E & Brody H. (1996). Physician recommendations and patient autonomy: Finding a balance between physician power and patient choice. *Annals of Internal Medicine, 125*(9), 763–769.

[9] Ryan E B, Giles H, Bartolucci G & Henwood K. (1986). Psycholinguistic and social psychological components of communication by and with the elderly. *Language & Communication, 6*(1/2), 1–24.

Benevolent Ageism in Large Language Models: A Comparative Analysis of Response Patterns to Young and Older Adult Users

Zhang Xiaoling

American International Assurance Company (Bermuda) Limited, Shanghai

Abstract: With the increasing integration of Generative Artificial Intelligence into daily life, whether their outputs harbor implicit social biases has become a focal point of research in psychology and human-computer interaction. This study, through a controlled experiment, aims to investigate whether Large Language Models exhibit systematic differences in their response patterns when interacting with users of different ages (25 years vs. 65 years). By analyzing the output texts across three scenarios—technology instruction, skill acquisition, and financial advice—the study found that LLMs display significant Benevolent Ageism towards older users. This bias is specifically manifested in the excessive use of Elderspeak, communication strategies based on the Deficit Hypothesis, and a tendency towards Excessive Risk Aversion in advice. The research suggests that although AI exhibits high Warmth in interactions, its default settings undermine the perceived Competence of older users. This “digital over-accommodation” may not only exacerbate Stereotype Threat for the elderly but also indirectly reinforce the Digital Divide.

Key words: Large language model; Benevolent ageism; Elderspeak; Stereotype Content model; Human-computer interaction; Algorithmic bias

https://doi.org/10.35534/pc.0804067