

AI技术在人才甄选中的心理学效用研究综述

严玲, 吴然

四川轻化工大学教育与心理科学学院, 四川 自贡

收稿日期: 2026年9月7日; 录用日期: 2026年10月2日; 发布日期: 2026年10月13日

摘要

人工智能技术已经深度融入企业人才选拔的整个流程, 利用自然语言处理、机器学习以及音视频行为特征分析等方式, AI可以捕捉和判断候选人的心理特质, 人才选拔本质上是组织心理学的实践场景, AI在提高效率的同时, 也产生了复杂的心理效应: 既可以减少社会赞许性偏差, 做到大样本心理特征识别, 降低主观偏见, 也会带来算法黑箱、公平感知受损、信任危机以及测评效度不足等问题。本文从积极效用、消极风险、边界条件与研究不足四个方面整理现有文献, 并且从实践启示与未来研究方向两个维度给出优化建议, 希望能给AI甄选的应用提供来自心理学视角的理论参考。

关键词

人工智能, 人才甄选, 工业与组织心理学, 公平感知, 算法偏见

A Review of the Psychological Utility of AI Technology in Talent Selection

Ling Yan, Ran Wu

College of Education and Psychological Sciences, Sichuan University of Science & Engineering, Zigong Sichuan

Received: September 7, 2026; accepted: October 2, 2026; published: October 13, 2026

Abstract

Artificial intelligence (AI) technologies have become deeply integrated into the entire enterprise selection process. By leveraging natural language processing, machine learning, and audio-visual behavioral feature analysis, AI can capture and assess candidates' psychological traits. As talent selection is essentially a practical application of organizational psychology, the deployment of AI—while enhancing efficiency—has also triggered complex psychological effects. On the one hand, it can mitigate cognitive biases of interviewers, reduce social desirability responding, and enable large-scale psychological profiling. On the other hand, it introduces challenges such as the

“algorithmic black box,” impaired perceptions of fairness, trust crises, and insufficient assessment validity. This paper synthesizes existing literature across four dimensions: positive utilities, negative risks, boundary conditions, and research gaps. Furthermore, it provides optimization recommendations from the perspectives of practical implications and future research directions. The objective is to offer a theoretical framework rooted in psychology for the application of AI in talent selection.

Keywords

Artificial Intelligence, Talent Selection, Industrial and Organizational Psychology, Perceived Fairness, Algorithmic Bias

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人才甄选是人力资源管理的主要内容，它的底层逻辑来自工业与组织心理学，目的是识别候选人的人格、动机、情绪能力以及岗位匹配度，以此预测工作绩效与离职风险。传统选拔依赖面试官的主观经验，成本高、效率低，测评结果容易受社会赞许性作答干扰(朱伟, 2019; 贺伟等, 2024)。

近年来，随着大模型与感知智能技术普及，AI 甄选工具得到广泛应用。AI 不仅提高效率，还直接参与心理评估：智能面试系统分析语音、微表情和文本，推断外向性、尽责性等大五人格维度；大语言模型生成情境问题并解析回答，判断成就动机与冲突处理能力，算法挖掘简历文本推断职业价值观(张光磊, 2024)，技术的介入改变了甄选双方的心理交互模式。

现有研究大多关注 AI 甄选技术的落地，而从应用心理学视角出发，系统整合应聘者、HR 与组织三方心理利弊的专门综述还需要进一步深化，AI 评估影响应聘者、HR 与组织三方的心理认知，既有积极作用也有负面后果。理清这些效应的作用机制，对 AI 的合理应用、平衡测评效度和保护求职者权益都十分重要。本文整理国内现有文献成果、辨析其中分歧、提炼研究缺口。

2. AI 人才甄选的积极心理学效用

2.1. 抑制人类面试官的认知偏差，增强判断客观性

传统面试过程容易受相似吸引偏差、刻板印象与晕轮效应等认知偏差影响，会严重损害选拔的公平性和预测效度(朱伟, 2019; 贺伟等, 2024)。围绕预设指标，基于算法的 AI 评估可以进行统一打分，能避开主观情感带来的干扰。

2.2. 降低社会赞许性作答，获取更真实心理信号

部分质性研究指出，不少求职者在 AI 面试中感受到的社会评价压力有一定程度降低，但同时也出现了镜头恐惧这类新的焦虑来源(刘潇肖, 2023)。这一差异或与个体评价敏感度有关，敏感度低者减压、高者增焦。

2.3. 实现大规模心理特征识别，拓展评估边界

传统测评受人力和时间成本限制，没办法对大量求职者做精细化评估。AI 技术的应用打破了这个限制：

按照自然语言处理与语音分析技术,系统能批量解析文本与语音数据当中的心理线索,把精细评估从抽样式小样本拓展到全覆盖式的大规模初筛,还能捕捉到反应时间、语言风格、语调变化等传统测评里难以量化或者容易被忽略的细微线索,为人才筛选提供了更多维度的信息参考(张光磊, 2024; 贺伟等, 2024)。

2.4. 标准化流程对程序公平感知的潜在增益

程序公平理论认为,甄选程序保持一致有助于提高应聘者的公平感知(陈扬, 宋琪, 2024)。AI 对所有候选人都采用统一规则与题目,理论上强化了“一视同仁”的程序公平,陈扬、宋琪(2024)提到,算法透明度是影响应聘者公平感知的关键条件,在明确告知 AI 使用规则的前提下,标准化流程能增强应聘者对组织的正面印象,但这种作用高度依赖算法透明度,缺乏透明度时,这种积极效应会被明显削弱,甚至直接消失。

3. AI 人才甄选的消极心理学风险

3.1. 算法黑箱诱发程序不公平感知与算法厌恶

算法厌恶表现为:即使统计结果表明算法预测的实际情况比人类判断更准,人们还是偏向不信任算法,尤其是抵触算法给出的负面判定结果,如果 AI 直接淘汰求职者,却不给出具体评估理由和决策依据,也就是算法黑箱,就会严重破坏人们对程序公平的感受。唐贵瑶、孙倩(2024)提到, AI 招聘中算法决策的不透明性会让应聘者质疑组织公平性,被 AI 淘汰的人更容易产生愤怒、委屈的情绪,会怀疑评估结果不真实。贺伟等(2024)在整理文献时也提到, AI 人才甄选中算法决策不透明,会明显降低应聘者对程序公平的感受,进而对组织吸引力和雇主品牌造成损害。

3.2. 心理特质测评效度缺陷导致错误画像

正如朱伟(2019)质疑的那样,利用外在行为线索推断个人内在心理特质会存在明显的效度风险——紧张场景里语速变快或者眼神躲闪,可能只是说明当下存在社交焦虑,并不是候选人责任心或者抗压能力的真实体现。当前 AI 对人格特质的评估效度有着明显的维度差异,刘潇肖(2023)的质性研究发现,求职者普遍对 AI 能否准确识别自己的人格持怀疑态度。这与实证报告的良好信效度证据表面矛盾,或源于主观知觉与统计效度测量对象不同,怀疑或源于效度信息未被有效传达。贺伟等(2024)整理现有实证文献后也提到,虽然 AI 在人格评估的聚合效度、区分效度和重测信度这些方面已经初步表现出不错的水平,但是现有研究对信效度的检验还不够全面,尤其是内容效度和结构效度的检验还需要后续进一步探索。

3.3. 历史数据带来的算法偏见,制造新型心理歧视

AI 系统利用学习历史数据中的模式来做决策,当历史数据本身就嵌入了社会偏见,算法很可能重复甚至放大原有的不公平,张光磊(2024)提到,如果训练样本本身存在性别、种族这类少数群体特征偏差,训练出来的 AI 模型在实际使用中也会再次出现不公平的决策,唐贵瑶、孙倩(2024)提到,算法偏见主要来自训练数据的局限性(比如历史数据本身就带有的性别、种族偏向)、系统本身的技术偏差,还有算法设计者的无意识偏见。举个例子,当历史管理岗位数据呈现性别偏向时,模型可能会错误判断男性更具备领导特质。少数群体感受到系统性的不利对待后,会产生被技术歧视的感受,降低自我效能和求职意愿,最终损害组织形象(贺伟等, 2024)。这与 2.1 节抑制即时认知偏差并不矛盾,二者分别对应过程性认知偏差与结构性数据偏见,即算法公平取决于过程标准化与数据代表性并重。

3.4. 人机交互压力引发独特心理应激反应

智能视频面试里,摄像头容易让求职者觉得自己被监控,进而引发焦虑,还会催生刻意表演的行为。

刘潇肖(2023)的研究也提到,不同个体对 AI 面试的评价敏感度有所差别:敏感度高的人更容易因为被机器审视产生焦虑和伪装行为,敏感度低的人则会因为感受到的评价压力更低,表现得更真实。

3.5. HR 的算法依赖弱化人类判断能力

人力资源管理者做决策时更容易出现自动化偏差,会盲目相信 AI 给出的结论而放弃自身专业判断,进一步放大了算法出错带来的损失(张光磊, 2024)。

4. AI 人才甄选心理学效用的边界条件

AI 甄选带来的心理效用并非技术特性的直接投射,而是受多个调节变量交互影响的动态系统。

算法透明度是最首要的边界条件。陈扬和宋琪(2024)在评述中提到,算法透明度与可解释性是影响应聘者公平感知的关键边界变量——当组织提前向候选人讲清楚 AI 的评估指标和基本评分逻辑时,应聘者的程序公平感知可能得到提高;但在完全“黑箱”的情况下,AI 的标准化优势反而可能引发算法厌恶,同时也要注意,透明度过高也可能带来隐私问题,或是让应聘者进行策略性伪装,据此可假设,存在一个兼顾公平感知与作答真实性的最优透明度区间,其位置随岗位性质与候选人特征而变化。

人机协同方式是关键的结构性边界。贺伟等(2024)在展望未来研究时提出,可以把 AI 定位为初筛和信息标注工具,由经过训练的人类专家掌握最终决策权,这种人机协同方式,更有助于抑制应聘者的负面心理反应,这个思路在组织心理学实践中说明,AI 更适合扮演“辅助者”的角色,而不是完全取代人类决策。

个体差异构成微观层面的边界,贺伟等(2024)梳理的文献表明,对 AI 技术更熟悉或是持积极态度的候选人,他们对公平感知和组织吸引力的评价,明显高于对技术陌生的群体。年轻群体和计算机自我效能感较高的应聘者对 AI 甄选的接纳度更高,心理应激反应更弱;年长群体或者技术效能感较低的个体则更容易感受到威胁和不公。

组织制度约束是上述边界条件能够有效运行的根本保障。为了遏制算法歧视、保护应聘者权益,企业需要建立配套的算法审查与效度校验机制,定期进行偏差审计与公平性检验(唐贵瑶, 孙倩, 2024),换句话说,算法透明度、人机协同方式与个体差异调节作用的发挥,都离不开组织层面完善的制度框架。

5. 当前研究不足

虽然学界已经开始关注 AI 甄选相关的心理学议题,但是现有研究在深度和广度上都有明显不足,需要加以审视。

第一,国内研究范式重思辨轻实证,因果证据链薄弱。目前国内关于 AI 甄选的文献以理论推演和观点综述为主(张光磊, 2024; 唐贵瑶, 孙倩, 2024),虽然贺伟等(2024)已经系统梳理了国际上的 21 篇实证研究,但是基于中国本土样本的大样本、纵向追踪研究仍然非常少,多数研究只截取了应聘者的即时心理反应(比如一次性问卷),没法揭示算法厌恶、信任重建等心理状态的动态演化过程和长期职场影响。

第二,核心变量操作化定义模糊,结论可比性差。对于“算法透明度”、“公平感知”、“人机信任”等关键构念,不同研究的界定与测量工具各异,导致研究发现之间难以进行有意义的比较与整合。例如,部分研究仅关注过程信息透明,而另一些研究则强调过程解释透明,Robert 等(2020, 转引自陈扬, 宋琪, 2024)将透明度区分为“过程信息透明”与“过程解释透明”两个维度,二者的作用机制可能有所不同。此种操作化差异严重影响了研究结论的可比性与知识积累的有效性。

第三,研究角度比较单一,没有关注关键利益相关方的互动规律,现有大部分研究把重点放在应聘者心理上,对于真正使用 AI 工具的 HR 从业者的认知过程和决策行为,几乎没有相关研究。张光磊(2024)

从理论层面提到, AI 可能会让 HR 产生“自动化偏差”的风险, 但这个结论目前还没有经过实证检验, 未来研究可以通过认知实验或者眼动追踪这类方法, 找出 HR 阅读 AI 评估报告时的信息关注模式和判断权重。

第四, 没有关注本土文化和社会情境的调节作用, 现有理论框架大多来自西方个体主义文化背景, 贺伟等(2024)梳理的核心实证研究大多基于西方样本, 陈扬和宋琪(2024)也提到, 直接把西方情境下的研究结论用到中国文化情境里, 会有明显的适配风险。中国文化具备高权力距离、重视人际关系的特点, 在这种情境下, 应聘者对算法决策公平性的感受、对组织权威的信任机制, 以及接受“机器评判”的程度, 可能和西方样本存在本质区别, 直接套用现有理论解释中国实践, 就可能出现文化适配性问题。

第五, 研究粒度太粗, 没有关注测评目标和岗位的异质性。现有研究笼统地把“AI 甄选”当成单一同质化技术, 既没有区分 AI 是用来测评人格、动机还是认知能力, 也没有探讨它在不同类型岗位(比如操作型、技术型、管理型)应用时的心理效应差异, 这种“一刀切”的研究角度, 大幅降低了结论的外部效度和实践指导价值。

6. 实践启示与未来研究展望

6.1. 实践启示

(1) 职能定位: 严格遵循“人机协同”的原则, AI 只用来完成第一轮线索提取和风险标注工作, 最终面试、录用以及否决决定, 必须由接受过结构化反偏见训练的 HR 或者工业组织心理学家来做, 以此减少自动化偏差带来的影响。

(2) 透明度分级: 实行“结果 + 理由”的反馈机制

给进入终轮面试但没通过的候选人, 给出“优势维度”与“待发展维度”的结构化反馈, 既能避开黑箱操作带来的不满, 也能避免信息透露太多让之后的应聘者做针对性伪装。

(3) 效标持续校验: 搭建动态监测体系, 每季度进行 AI 评分和新员工试用期绩效(KPI 以及主管评价)的效标关联度分析, 对预测效度低于中等效应量(比如 $r < 0.20$)的测评模块做停用或者修正处理。

(4) 交互降噪: 在智能面试平台里增加“模拟练习”与“环境自检”环节(参考刘潇肖, 2023 年关于 AI 面试环境对求职者表现影响的研究结论), 减少因为机器监测带来的测评焦虑; 可以设置临界分数段的人工复核以及候选人申诉机制, 不过要说明的是, 这个建议是本文作者根据现有文献得出的推论, 它的有效性还需要后续实际研究来验证。

(5) 偏见审计: 定期按照人口学变量(性别、年龄、户籍)分组做公平性检验, 计算人口统计均等(不同群体录用比例相当)与机会均等(胜任者录用概率相当)两类主流指标, 后者更契合甄选胜任者的招聘目标。如果发现系统性偏差, 就要马上重新平衡训练集或者调整算法损失函数里公平性约束的权重。

6.2. 未来研究方向

未来研究可在巩固现有发现的基础上, 着力从“现象描述”往纵深推进到“机制解析”与“干预检验”, 在以下方向取得突破。

第一, 进行本土化的纵向与实验研究, 采用经验取样法或纵向追踪设计, 捕捉求职者在 AI 甄选不同阶段(测试前、测试中、结果反馈后)的心理动态变化及其对求职者职业决策与技能发展的长期影响。

还可设计情境实验, 调整算法透明度和反馈框架, 直接验证对中国求职者公平感知与组织吸引力的因果效应, 修正并构建契合中国文化语境的 AI 接纳理论。

第二, 深入探究 HR 的算法认知与决策行为。利用访谈、认知实验或眼动追踪技术, 明确 HR 阅读 AI 报告时的信息注意模式与判断权重, 验证“自动化偏差”的产生条件, 还可开发“算法素养”培训项

目, 提高从业者对 AI 结论的批判性评估与情境化校准能力。

第三, 建立 AI 心理测评的效度矩阵与伦理规范。学界与业界需要共同搭建 AI 测评工具的常模与信效度报告规范, 明确区分 AI 可有效预测的行为指标(如语言流畅性、情绪稳定性)与必须由专家深度研判的复杂构念(如领导力潜能、创造力)之间的界限。

第四, 进行岗位异质性的差异化研究。打破“泛 AI 化”的研究惯性, 对比分析 AI 在蓝领操作岗位(侧重行为合规性)、专业技术岗位(侧重逻辑推理与知识)与管理岗位(侧重冲突处理与愿景沟通)应用时, 产生的差异化心理效应与适用边界。

第五, 强化干预性实证研究。不再仅仅停留于描述问题, 而是利用准实验设计, 系统验证“透明度干预”、“人机权责宣贯”、“申诉通道设立”等管理措施在提高 AI 甄选公平感、信任度与效度方面的实际效能, 并检验 AI 反向赋能求职者以促进招聘市场信息对称的干预效果。给组织给出基于证据的解决方案。

7. 结论

从应用心理学视角来看, AI 人才甄选是一把双刃剑。它可以减少面试官的认知偏差, 降低社会赞许性带来的干扰, 还能做到大样本心理特征的高效提取, 在优化程序公平上有潜力; 它也存在算法黑箱、算法厌恶、测评效度有限、复刻历史偏见、增加求职者应激反应以及 HR 过度依赖自动化等深层风险, 本文提出, AI 应用带来的心理影响不是技术本身注定的, 而是由它的应用模式决定的, 算法透明度、人机权责划分、偏见审计和效度管控, 共同构成了调节 AI 应用效果的核心约束。实践中应当放弃技术万能论或者全盘否定 AI 的极端看法, 坚持以证据为导向的人机协同方式, 把工业与组织心理学的理论深度融入到 AI 应用的全流程中, 从而在甄选效率、测评效度和人文关怀之间找到动态平衡。

参考文献

- 陈扬, 宋琪(2024). 评《基于人工智能的人才甄选》: 关于概念范畴、边界条件及中国情境化的思考. *管理学季刊*, 9(3), 47-59.
- 贺伟, 李亚莉, 汪默(2024). 基于人工智能的人才甄选: 研究进展与未来展望. *管理学季刊*, 9(3), 1-18.
- 刘潇肖(2023). 求职者对 AI 面试的接受度及其影响因素: 基于公平视角的质性研究. *中国人力资源开发*, 40(3), 117-130.
- 唐贵瑶, 孙倩(2024). 跨越技术边界: AI 招聘的困境与应对策略. *管理学季刊*, 9(3), 37-46.
- 张光磊(2024). 基于人工智能的人才甄选的底层逻辑和伦理困境. *管理学季刊*, 9(3), 19-27.
- 朱伟(2019). 在招聘中应用 AI, 当心法律及伦理. *人力资源*, (13), 83-85.