人工智能赋能人力资源管理研究综述：理论框架、实践悖论与研究缺口

罗子敬

(新余青云联合会计师事务所，江西 新余 338000)

摘 要：人工智能已经越过单纯的信息化辅助阶段，开始进入招聘甄选、培训开发、绩效评价、薪酬服务、员工关系与人才规划等管理环节。已有研究较充分地呈现了算法在信息处理、预测分析和个性化服务方面的潜力，但效率提升并不会自动带来公平、信任或更好的组织结果。本文在规范化检索的基础上，对 2015—2026 年中英文研究进行整合性述评，重点考察概念边界、研究演进、理论解释、应用情境及治理争议。梳理表明，相关研究的关注点经历了由流程电子化到预测分析、再到价值与责任治理的移动；国内成果偏重技术落地和组织转型，国外文献则较早讨论歧视、劳动控制、员工权利与问责。在此基础上，本文把技术、组织、个体、价值和结果纳入同一分析链条，并从实际管理过程归纳出五类持续存在的实践悖论：效率与公平、数据利用与隐私、自动化与人本沟通、技术更新与组织承载、量化判断与具体情境。文章认为，人工智能在人力资源管理中的价值取决于制度安排和使用方式，而非模型性能本身；对高影响人事决策实行分级管理、保留有意义的人工审查、建立解释和申诉渠道，是推进负责任应用的基本条件。

关键词：人工智能；人力资源管理；智能人力资源管理；算法治理；人机协同；研究综述

DOI：doi.org/10.70693/202607121395

一、引言

（一）研究背景

在组织管理中，人工智能最先改变的是信息处理速度，随后才逐渐触及判断和决策。简历解析、考勤核验、薪资计算等任务具有较清晰的规则，容易被系统接管；潜力识别、绩效沟通、冲突调解和职业发展却依赖语境、关系与经验，难以压缩成稳定指标。正因为人力资源管理同时包含标准化事务与高度情境化活动，它成为观察技术如何进入组织权力和劳动关系的一个典型窗口。国内研究已从数字化转型视角讨论人工智能对组织结构、岗位能力和管理流程的影响[1]，国外研究则提醒，人力资源数据的样本、标签和因果关系具有特殊性，不能照搬一般商业分析的做法[2]。

算法带来的收益与代价往往同时出现。它可以缩短筛选时间，帮助管理者识别能力缺口、人员流动和用工需求，却也可能把过去形成的招聘偏好写入模型，把原本难以观察的员工行为转化为持续记录和比较的对象。相关综述已经指出，AI-HRM 能否产生正面结果，不只取决于模型精度，还受到组织能力、员工接受以及制度环境的共同影响[3][4]。因此，研究重点不宜停留在“技术有没有用”，而应进一步追问：哪些任务可以交给系统，最终决定由谁作出，受影响者能否了解依据并提出异议。

现有文献的议题覆盖面已经较广，但知识积累仍有明显断点。一些研究用应用清单代替机制分析，能够说明系统“做了什么”，却没有解释算法建议怎样经过流程、权责和管理者判断转化为实际后果。国内外比较也常以主题数量为主，对研究重心为何由提效转向治理说明不足。与此同时，算法偏见、隐私和员工抵触已被反复讨论，能够把风险来源、组织条件与具体治理工具连起来的框架仍然不多。

本文采用“实践张力”这一概念，是为了分析同一项技术安排为何会在带来增益的同时制造新的损耗，而不是简单罗列利弊。

（二）研究问题与研究意义

作者简介：罗子敬（1989-），男，硕士，研究方向:人力资源管理、数字化转型与人工智能治理、工商管理。
通讯作者：罗子敬

本文关注三个相互关联的问题：人工智能赋能人力资源管理的边界如何划定，它与电子化、数字化人力资源管理有何实质差别；国内外研究在演进过程中呈现出哪些共同趋势，又为何形成不同的问题意识；现实应用中的矛盾由哪些机制触发，研究与实践应如何回应。分析并不预设人工智能必然改善绩效，而是把技术后果看作数据、制度、人员行为和价值规范共同作用的结果。

理论上，文章把技术接受、技术—组织—环境、人机协同和算法治理放入同一条解释链，区分“系统得到使用”与“系统产生正当且有效的结果”这两个不同命题。通过回顾研究演进，可以看到 AI-HRM 的知识基础已从工具效率逐步伸向权利、责任和劳动关系。实践上，本文依据任务的标准化程度、决策影响和人际互动强度识别适用边界，尤其强调录用、晋升、薪酬、处分等高影响事项不能仅凭模型输出完成，以避免企业在转型中形成重设备采购、轻流程治理，或重预测结果、轻沟通说明的倾向。

（三）研究方法与分析思路

本文采取整合性述评。为使文献来源与取舍过程可以复核，检索范围限定为 2015 年至 2026 年 7 月 2 日。中文文献从 CNKI 和万方获取，英文文献由 Web of Science 与 Scopus 检索。检索时分别设置技术、人力资源职能和治理三组词项，再依据各数据库的字段规则，在题名、摘要、关键词或主题项中进行组合查询。本文旨在辨析概念边界及作用机制，不进行统计意义上的效应量合并。

表 1 文献检索与筛选设置

检索维度	设置内容
时间范围	2015—2026 年；检索截止时间为 2026 年 7 月 2 日
中文数据库	CNKI、万方数据库
英文数据库	Web of Science、Scopus 数据库
检索词组	技术词（人工智能、机器学习、生成式人工智能等）+人力资源职能词+治理词
纳入原则	主题与 AI-HRM 直接相关；来源可核验；论证信息较完整；兼顾中英文代表性成果
排除原则	重复文献、弱相关材料、一般技术介绍及出版信息不完整的材料

材料整理不采用机械的主题计数，而是沿着问题形成过程展开。分析先处理人工智能赋能、智能人力资源管理和实践张力的概念关系，再结合技术条件和研究问题的变化梳理国内外演进；随后比较技术采纳、组织适配、人本管理及算法治理的解释范围，并将不同职能场景中的任务特征与风险放在一起考察。最后，文章从理论、样本、方法和生成式人工智能治理四个方面归纳尚未解决的问题。

这样的处理有助于避免把国外规范性概念直接移植到国内组织，也能保留不同制度情境下结论的适用边界。

二、核心概念界定

（一）人工智能赋能人力资源管理

人工智能赋能人力资源管理，是组织把机器学习、自然语言处理、知识图谱、智能推荐、生成式模型及其他数据技术嵌入人才获取、配置、评价、培养和保留过程，使系统参与信息识别、模式发现、内容生成、风险提示或决策支持。这里的“赋能”指管理能力得到扩展，并不等同于用机器替换人员。技术是否真正形成赋能，至少要满足两项条件：应用目标能够兼顾组织需要和员工发展；算法输出进入正式管理流程后，仍有明确的解释、判断和责任承担主体。计算过程复杂不能成为责任悬空的理由。

从实际作用看，人工智能大致通过四条路径进入人力资源管理。流程层面，它可以处理录入、问答、核验等重复事务；认知层面，它从分散数据中寻找关联，为岗位匹配、能力诊断和人员规划提供参考；互动层面，推荐和即时反馈改变了以往统一、周期性的服务方式；战略层面，岗位、能力和组织绩效数据被连接起来，用于人才盘点、继任计划与组织设计。四条路径的效果并不同步：流程更快，未必意味着解释更充分；预测更精细，也可能意味着员工受到更密集的观察。

因此，对“赋能”的判断不能只依据系统上线、处理时长或准确率。更有意义的评价还应包括决策质量是否改善、员工是否理解并接受程序、组织是否从反馈中学习，以及错误和风险能否被发现、纠正和追溯。若技术只是把原有的不合理规则自动执行得更快，所谓升级并没有带来管理意义上的进步。

（二）智能人力资源管理

智能人力资源管理可以理解为电子化和数字化管理继续发展的结果。电子化主要解决信息存储与流程上网，数字化强调数据贯通、可视化和业务协同，智能化则加入预测、推荐、生成及自适应反馈。它不是一套孤立的软件，而是由数据标准、算法模型、业务规则、岗位分工、决策程序和伦理规范共同构成的社会技术系统。若只评估模型而不考察数据来源、使用者能力和组织制度，技术的独立作用往往会被高估。

这一管理形态通常呈现五项特征。其一，判断依据由个人经验扩展到多源数据，但数据仍带有采集条件和组织历史；其二，模型会随新数据和使用反馈发生变化，需要持续监测而非一次验收；其三，不同场景对准确性、可解释性和人际互动的要求并不相同；其四，系统负责计算和提示，人类仍需处理价值冲突、例外情况与沟通；其五，数据边界、决策权限、复核程序、申诉渠道和责任主体必须被制度化。

由此可见，智能化成熟度不应以产品新旧或功能多少衡量，而应看组织是否具备稳定、可解释、可纠错的使用能力。技术新颖而制度空缺，往往只会把旧问题转化为更难识别的新问题。

（三）实践悖论与研究缺口

本文所说的实践悖论，是指一项 AI-HRM 安排在实现某个管理目标时，同时形成与该目标相牵制的后果。统一模型可以压缩人为随意性，却可能更隐蔽地延续历史偏见；扩大数据采集有利于风险预警，也可能损害心理安全；自动生成反馈提高了响应速度，却可能使管理者减少对个体处境的理解。这些现象不是彼此独立的优点和缺点，而是相互依赖的目标长期并存，组织很难通过一次技术调整彻底消除。

研究缺口则体现为解释范围、证据类型或治理效果仍不充分。把研究缺口与实践悖论结合起来，可以从“发现存在风险”进一步走向“说明风险怎样形成、何种条件下加剧、哪些措施能够缓解”。当前真正稀缺的并不是更多应用案例，而是能够连接技术设计、组织制度、员工感知、权力配置和结果反馈的中层理论，以及对审计、复核、员工参与等治理工具效果的持续检验。

三、国内外研究演进脉络

如果以技术所承担的任务和研究者提出的问题为线索，AI-HRM 文献可以辨认出三种相继出现、又彼此重叠的取向：最初关注人事事务的信息化处理，随后转向预测分析和职能场景扩展，近年则把人机协同、算法权力和责任治理置于中心。这里的阶段划分并非严格按年份切开，而是用于说明研究问题如何发生变化；在不少组织中，三种取向仍然同时存在。

（一）萌芽阶段：事务处理与工具导入

国内早期讨论主要承接人力资源信息化和电子化管理，关注档案、考勤、工资、招聘发布和培训记录等事务如何在线完成。评价标准多是行政成本、处理速度和信息共享，技术被放在既有流程的后台。由于数据尚未跨模块连接，预测和自动决策并非核心议题，员工感受、权利边界与模型责任也较少进入分析。这个阶段的重要贡献在于形成基础数据和流程接口，但也留下一个惯性：容易把纸面流程搬到线上视作管理改进本身。

国外同期研究同样从人力资源信息系统和在线测评起步，却较早注意技术对 HR 角色及员工关系的改变。Stone 等指出，技术不仅调整招聘、培训和绩效管理的操作方式，还会改变人力资源从业者所需能力以及其与员工的互动[5]。这一观察为后续研究提供了关键前提：技术并非置身组织之外的中性工具，它会重新安排信息可见性、判断权和控制方式。

（二）发展阶段：预测分析与场景扩展

云计算、大数据和机器学习被企业采用后，人力资源系统开始从记录过去转向估计未来。国内研究的对象随之扩展到简历解析、岗位匹配、人才画像、学习推荐、绩效预警和离职预测，并把人工智能与流程再造、职能升级联系起来。张琪等对相关理论和实践的梳理，反映了研究重心由事务自动化向数据分析与决策支持的转变[6]。这一时期仍以效率、成本和匹配精度为主要指标，但数据质量、系统集成和组织适配已逐渐成为不可回避的问题。

国外研究在拓展应用场景的同时，更强调人力资源数据的限制。Tambe 等指出，样本规模小、标签含义不稳定、结果受情境影响以及实验难以开展，都会削弱模型结论的可靠性[2]。招聘和人才分析研究由此不再只比较准确率，而开始考察程序公平、解释透明和候选人接受：预测出错由谁负责，被评价者能否理解结果，组织是否提供纠正机会，逐渐与技术效果一起成为评价内容。

（三）深化阶段：价值治理、人机协同与责任边界

2021 年以来，大模型和生成式人工智能显著降低了内容生成与知识交互门槛，国内议题也从系统功能延伸到数字压力、算法信任、组织能力和数据合规。越来越多的实践说明，购买产品不能替代流程重构。数据口径不统一、岗位责任含混、管理者缺少数字素养，或员工在部署前没有得到充分沟通，都会使系统停留在展示层面，甚至增加重复录入和核对。于是，AI-HRM 开始被视为一项组织变革，而不只是信息部门的项目。

国外文献在这一阶段明显加强了跨学科解释。Pan 与 Froese 提出应连接管理学、信息系统、心理学、伦理学和法学[3]；Vrontis 等把组织能力与制度环境视为理解技术后果的重要条件[4]；Budhwar 等则从国际人力资源管理出发，说明文化、监管和劳动关系差异会改变 AI 的可接受程度与治理方式[7]。Kellogg 等把算法放回劳动过程，揭示规则设定、持续监测和自动奖惩如何形成新的控制场域[8]。这类研究把问题从“工具是否先进”推向“工具怎样改变管理权力”。

伦理与权利研究进一步讨论个人完整性、差别影响和救济。Leicht-Deobald 等认为，过度追求效率的算法决策可能压缩员工的自主判断和尊严[9]；Ajunwa 以工作中的“黑箱”说明劳动者常常难以获知自动决策依据[10]；

Barocas 和 Selbst 指出，即使不直接使用敏感属性，代理变量仍可能造成差别影响[11]；Raghavan 等对招聘算法供应商的公平声明进行评估，发现商业宣称与可检验效果之间存在距离[12]。由此，研究关注点逐渐由采用与否转向采用过程是否负责任。

（四）国内外研究的差异与联系

国内外研究都显示出由流程电子化、预测分析走向价值治理的总体趋势，但两类文献的问题起点并不相同。国内研究更直接回应企业数字化转型、管理提效和人才供给变化，重视系统怎样进入已有组织；国外研究较早受数据保护、反歧视和劳动过程传统影响，更关心算法对公平、权力和责任的重构。这种差别源于制度环境、研究传统和企业应用阶段，不宜用简单的先进或落后来评价。

两种路径可以相互补充。国外关于透明、审计和劳动者权利的成果，有助于国内企业划定高影响决策边界并建立问责；国内组织在复杂流程协同、本土管理文化和多层级落地方面的经验，也能修正把治理理解为统一模板的倾向。更有价值的后续比较，应在共同概念下区分所有制、行业、规模和用工关系，既不直接套用国外结论，也不以情境特殊为由回避普遍的公平与权利问题。

表 2 国内外 AI-HRM 研究取向的演进对照

研究取向	国内文献关注	国外文献关注	问题变化
事务处理与工具导入	人事信息化、线上流程、行政提效	技术对 HR 角色、沟通和员工关系的影响	从纸面处理转向数字记录
预测分析与场景扩展	岗位匹配、人才画像、绩效预警、离职预测	效果评估、程序公平、解释和候选人接受	从事后记录转向预测与建议
价值治理与责任边界	组织适配、数字压力、数据合规、人本管理	算法权力、员工权利、问责与跨学科治理	从效率问题转向权利与责任

四、理论基础与整合性分析框架

（一）技术采纳与组织适配

技术接受模型指出，感知有用性和感知易用性影响个体的使用意愿[13]。在人力资源场景中，系统能否减少实际负担、界面是否便于操作、输出能否理解，都会影响 HR 人员和员工是否持续使用。不过，AI-HRM 经常用于评价和资源分配，系统使用者与被评价者并非同一群体。管理者愿意使用，不代表候选人或员工认可；使用频率高，也不能证明结果正当。因此，技术接受需要与程序公平、信任和权力关系共同分析。

技术—组织—环境框架从技术条件、组织准备和外部环境解释创新采纳[14]。对应到 AI-HRM，技术条件包括数据质量、系统集成、稳定性和安全；组织条件包括高层支持、流程清晰度、数字素养、跨部门协同及文化氛围；外部环境则涉及监管、行业竞争、供应商和劳动力市场。该框架的意义在于把模型性能放回组织承载条件：技术可用只是起点，组织是否能够解释、复核和纠错，往往更直接地决定应用后果。

（二）人本管理与人机协同

人力资源管理面对的是具有尊严、情感和发展诉求的行动者，而不是可任意加工的数据集合。谢小云等从人与技术交互角度提出，数字化人力资源管理需要同时考察系统特征与员工的行为、心理和关系变化[15]。据此，智能化评价不能只看成本和速度，还要观察员工是否得到充分说明、是否感到被尊重、能否参与与自身利益有关的程序，以及系统究竟支持能力发展，还是主要用于单向监控。

人机协同理论提供了任务分工的进一步解释。Raisch 和 Krakowski 指出，人工智能既可能替代人的任务，也可能增强人的判断，组织设计需要同时处理这两种逻辑[16]。在人力资源管理中，数据清洗、信息检索、规则校验和异常提示适合由系统承担；价值冲突、例外情形、绩效沟通和职业辅导则需要人的语境理解。合理边界不是固定清单，而应随决策影响、错误代价和可逆程度调整。

（三）算法治理与伦理约束

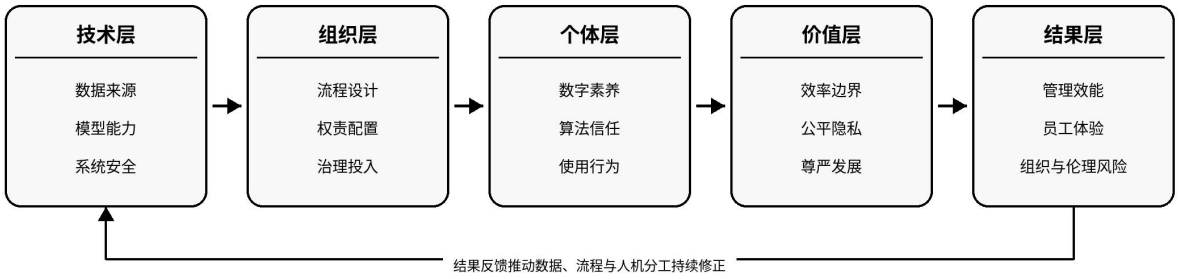
算法治理覆盖数据获取、模型设计、部署使用和结果反馈等环节，核心议题包括公平、透明、可解释性和问责。可解释人工智能研究显示，“解释”面对不同对象具有不同含义：开发者需要理解模型机制，管理者需要知道判断依据，员工则需要获得能够理解且与自身决定相关的说明[17]。受影响者不必掌握复杂算法，但应知道哪些数据被使用、哪些因素影响结果、如何补充信息，以及谁对最终决定负责。

制度规范为组织设定了最低边界。NIST 人工智能风险管理框架把治理、识别、测量和管理视为持续过程，要求兼顾可靠性、安全、隐私、公平和透明[18]；OECD 人工智能原则强调以人为本、稳健和问责[19]；我国《个人信息保护法》要求个人信息处理遵循合法、正当、必要和诚信，并对自动化决策的透明与公平作出规定[20]。这些规范虽然表述不同，却共同否定了一种做法：仅以商业效率证明高影响人事决策的正当性。

（四）整合性分析框架

综合上述理论，本文把 AI-HRM 置于技术、组织、个体、价值和结果五个层次中观察。技术层讨论数据来源、模型能力、系统稳定和安全；组织层考察流程设计、权责配置、治理结构与资源投入；个体层关注 HR 人员的数字素养、管理者使用建议的方式，以及员工的理解、信任和应对；价值层处理效率、公平、隐私、尊严与发展机会之间的边界；结果层同时观察管理效能、员工体验、组织绩效和伦理风险。五层并非平行清单，而是前后相接的转化链条。

图 1 人工智能赋能人力资源管理的五层转化链条



这一框架要说明的是，计算能力不会直接变成管理价值。技术层只能提供可能性，组织层决定算法被放入何种流程和由谁掌握决定权，个体层影响建议是被审慎采用、机械服从还是消极抵制，价值层为效率设置可接受边界，结果层则检验收益及副作用。当结果偏离目标时，组织需要回到数据、规则和使用方式查找原因，而不能把问题简单归结为员工“不适应”。从这个意义上说，AI-HRM 的效果来自技术能力、组织承载、人员信任和价值治理的共同作用。

五、应用场景与实践悖论

（一）主要应用场景

人工智能是否适合进入某项人力资源任务，首先取决于任务结构。规则稳定、数据相对完整、结果容易复核的工作，更适合自动化或半自动化；涉及重大权益、信息不充分或强情感互动的工作，则需要保持较高程度的人类参与。招聘甄选是目前最常见的应用场景，系统可以解析简历、计算岗位匹配、辅助测评并形成候选人排序。Bogen 与 Rieke 提醒，训练样本、成功标准和代理变量都可能把偏差带入筛选[21]。

因此，招聘系统不能只看处理量和召回率，还要比较不同群体的错误率，说明筛除依据，并允许候选人补充无法结构化呈现的信息。

培训开发、绩效、薪酬和员工关系的任务要求各不相同。学习推荐可以依据岗位要求、能力测评和学习记录提高资源匹配，但历史记录也可能把员工锁定在既有路径；过程数据能为绩效评价提供更及时的证据，却难以完整表示创造性、协作贡献和长期潜力；薪资核算、福利问答适合明确规则的自动处理，薪酬差异和例外仍需管理者解释；离职预警、情绪识别和服务机器人提高了响应速度，同时也最容易触碰隐私、监控感与信任边界。

对于公共部门以及程序约束较强的组织，成本节约不能替代公共责任。凡自动化结果会改变人员录用、岗位安排、收入待遇或处分结果，相关组织都应说明决定所依据的关键信息，并为当事人保留复查和申诉路径。效率固然需要评估，但只能与程序正当、可救济性等指标一并考察。

表 3 人工智能在人力资源管理中的主要应用场景

应用场景	常见功能	可能收益	需要警惕的问题
招聘甄选	简历解析、岗位匹配、测评辅助、候选人排序	提高处理速度，扩大候选人识别范围	历史偏见、代理歧视、说明不足
培训开发	能力诊断、课程推荐、学习路径安排	提高资源匹配和学习个性化	历史记录固化、路径依赖、过度服从推荐
绩效管理	过程数据分析、异常提示、反馈草拟	增强及时性，补充评价证据	忽视隐性贡献、指标操纵、沟通减少
薪酬福利	薪资计算、福利问答、成本分析	提升规则执行与服务响应	数据安全、规则不透明、差异解释不足

应用场景	常见功能	可能收益	需要警惕的问题
员工关系	员工咨询、流失预警、情绪与风险识别	提高服务可及性和组织预警能力	隐私侵犯、监控感、误判和信任下降

（二）核心实践悖论

其一，效率提高并不等于公平改善，两者甚至可能朝相反方向变化。统一模型确实能够压缩部分人工判断，并加快大规模筛选和分析。

问题在于，模型所学习的仍是既有数据以及其中沉淀的用人偏好。如果某些群体过去较少获得机会，算法便可能把历史结果误当作能力差异。学校、居住区域、职业中断或语言特征等变量，也可能间接对应敏感属性。因此，形式上对所有人采用同一规则，并不等同于结果公平。

其二，预测所依赖的数据扩张，常与隐私空间收缩相伴。为了判断流失、绩效或情绪状态，系统可能把履历、学习、沟通和日常行为记录汇入同一分析过程。数据一旦跨越原定用途，越权调用、长期留存和再次利用的风险便随之增加，尤其是由模型推断出的离职倾向、心理或情绪信息，更不能按普通业务数据处理。数据发生在工作场所或可从公开渠道取得，并不意味着企业可以不受限制地使用。若员工不知道采集理由、保存期限以及分析结果将被用于何种决定，工具就容易被视为监控装置；信任下降后，还可能出现回避、修饰数据等策略性反应。

其三，自动化可以释放事务时间，也可能削弱人本沟通。智能问答、审批和文本生成确实能减少重复工作，但节省出的时间是否用于更高质量的交流，取决于岗位和流程是否同步调整。如果系统只是替代了接触窗口，员工在绩效争议、职业困惑或冲突处理中得到的可能只是标准答案，管理者也可能借“系统结论”回避说明。技术减少了事务负担，却未必增加倾听、协商和关怀。

其四，产品迭代速度往往超过组织调整速度。模型和功能快速更新，企业的数据标准、审批流程、岗位能力与责任制度却需要较长时间才能改变。试点阶段依靠少数人员维护的问题，在跨部门推广后会集中暴露，例如口径不一、接口割裂、规则含混和重复核对。技术越复杂，组织越需要具备问题定义、供应商管理、模型评估和风险沟通能力；缺少这些能力，先进系统反而可能成为新的信息孤岛。

其五，量化看似精准，却有明确的情境边界。算法擅长在规则相对稳定的任务中寻找模式，但模式并不等于因果，也不等于理解个体处境。出勤、薪资和资格核验可以较充分地算法化，创造潜力、团队影响、伦理判断和员工困难则很难由少数指标完整表达。当组织把“容易测量”误当成“最重要”，员工会围绕指标调整行为，隐性贡献和长期价值反而被挤出视野。真正的精准应包括对不确定性和适用范围的诚实表达。

表 4 AI-HRM 实践悖论及其治理着力点

实践张力	典型表现	治理着力点
效率与公平	统一规则提高速度，却可能延续历史偏见或形成代理歧视	群体差异测试、独立审计、人工复核
数据利用与隐私	数据范围扩大有利于预测，也增加目的漂移和监控风险	目的限定、最小必要、权限控制、保存期限
自动化与人本沟通	事务响应更快，但解释、倾听和协商空间可能缩小	高影响环节保留人类判断和面对面沟通
技术更新与组织承载	产品迭代快于流程、能力和责任制度调整	同步推进流程重构、人员能力和责任配置
量化判断与具体情境	模型擅长可测任务，却难以表示隐性贡献和复杂处境	明确适用范围、不确定性和例外处理

（三）张力的形成机制

上述张力并非来自某个孤立错误，而是技术、组织、个体和制度共同作用的结果。训练数据、标签定义和优化目标把特定的管理偏好写入模型；企业在部署中往往重视上线时间和可视化成果，对流程重构、责任划分和员工参与投入不足；HR 人员若缺少评估模型和质疑输出的能力，容易形成自动化依赖，员工则会因决定不透明而不信任或采取策略性适应；与此同时，数据权利、劳动管理权和自动决策责任在具体场景中仍需细化。国内关于 AI-HRM 理论模型、数字化整合、机器学习应用和公共部门实践的研究，都提示技术后果必须放回组织情境与制度约束中理解[22][23][24][25]。

这些机制之间还会相互强化。例如，组织为了追求短期上线而降低数据治理要求，模型错误便更容易发生；错误发生后若缺少解释和复核，员工信任下降，随后提供的数据和使用行为也会发生变化，进一步影响模型效果。由此可见，张力不是部署完成后的附带问题，而是在技术进入组织的全过程中逐步生成。

（四）技术、组织与人本的联动治理

治理需要同时作用于三个层面。技术层面应建立数据质量检查、群体差异测试、性能监测、版本留痕与安全控制，并使高影响模型能够提供与决定相关的说明；组织层面要明确算法建议、管理者判断和复核部门的责任，将风险评估嵌入需求提出、采购、试点、上线和迭代，而不是在争议发生后临时补救；人本层面则应在必要范围内告知员工数据用途，为其提供补充说明、人工复核和申诉渠道，并让受影响群体参与关键规则的讨论。

组织管理中的人工智能决策和算法劳动研究都表明，治理对象不能只限于模型误差[26][27]。同样重要的是决定权如何分配、员工是否具有发声空间、系统建议能否被质疑，以及问题发生后责任是否可追溯。只有把这些安排写入日常流程，治理才不会停留在原则声明。

六、研究缺口与未来方向

AI-HRM 文献增长很快，但理论仍较分散，样本和方法也存在集中现象。个体对算法决定的评价并不完全由结果好坏决定，程序是否透明、本人能否表达意见，以及决定被归因于机器还是管理者，都会影响公平感[28]。接下来的研究不宜只寻找变量之间的相关关系，更需要解释作用过程，并比较不同治理安排是否真正改善决策质量和员工体验。

（一）理论研究缺口

理论推进的首要任务，不是继续叠加解释变量，而是说明技术、制度和员工体验如何在同一过程中相互作用。现有技术接受、组织变革与伦理治理研究各自揭示了一部分机制，却较少回答相同系统为何在不同单位呈现迥异后果。后续可围绕“建议—判断—反应—修正”的过程建立中层模型：观察管理者如何读取并使用算法建议，员工如何理解决定并调整行为，组织又怎样依据争议和反馈改动规则。在此基础上，再检验数据质量、权力距离、任务能否解释以及错误是否可逆等条件何时改变上述关系。

其次，技术性能、管理效能和规范正当性需要被分别测量。准确率高并不必然改善组织绩效，绩效改善也不能自动证明员工权利获得保障。后续研究可同时考察预测表现、决定一致性、员工信任、申诉成本、群体差异和长期组织学习，防止用一个效率指标代替多元价值判断。

再次，技术赋权并非对所有主体均等发生。人工智能可能增强管理者获取和处理信息的能力，也可能改善员工自助服务与职业发展；结果取决于谁能看到数据、谁拥有解释权和选择权。同一系统在平台企业、制造企业、中小企业或公共部门中的作用很可能不同。研究需要更具体地回答：技术赋权了谁，成本由谁承担，收益与风险在组织内部如何分配。

此外，算法治理不能被简化为对模型的技术修补。在线劳动平台研究显示，算法通过派单、评价、排名和奖惩重塑劳动控制[29]。在人力资源管理中，偏差既可能来自数据，也可能来自组织对“优秀”“高绩效”和“可用人才”的定义。只有把优化目标、组织文化和劳动关系纳入分析，才能判断公平指标究竟改变了权力结构，还是仅为既有标准增加了技术合法性。

本土化理论建构仍显不足。中国企业在层级结构、集体关系、用工形式和数据基础上差异显著，新一代人工智能又在加速岗位和能力重构[30]。可围绕员工数字权利、组织信任、算法申诉、人机责任和供应商治理开发本土量表与过程模型，再通过跨地区、跨制度比较辨认哪些机制具有普遍性，哪些结论依赖具体情境。

（二）实践研究缺口

现有案例和实证研究较多来自大型企业、互联网平台或数字化基础较好的组织，中小企业、传统制造业、劳动密集型服务业和公共机构受到的关注明显较少。已有综述主张把研究视线由技术功能转向组织情境、实施过程和实际绩效[31]。未来尤其需要记录失败案例和低成熟度组织，考察在预算有限、数据基础薄弱的情况下，如何确定最低数据标准、选择可解释工具、管理外部供应商，并避免系统闲置和重复建设。

治理工具是否有效，也缺少直接比较。算法审计、伦理委员会、员工参与和人工复核常被作为建议提出，但它们在不同场景中的成本、执行难度和实际作用尚不清楚。以情绪识别为例，技术效度、推断合理性和员工同意本身都存在争议[32]。现场实验、准实验和比较案例可以用来检验不同告知方式、解释类型、复核层级和申诉程序是否降低不公平感，同时观察治理安排是否出现形式化执行。

（三）研究方法缺口

AI-HRM 并非一次部署即可定型，系统和制度都会在使用中改变。然而，现有证据仍以横截面问卷和概念讨论为主，这类资料只能截取某一时点，难以看清员工从新奇、接受到依赖或抵触的变化，也无法还原试点扩围、规则调整和错误纠正的过程。今后的研究可把纵向案例、关键事件追踪、多轮访谈和数字行为记录结合起来，连续观察需求提出、供应商遴选、模型训练、试运行以及常态化使用等环节。

跨学科研究也需要从“同时引用不同领域”走向共同设计。算法偏差研究可以把计算检测与组织公平理论结合，隐私研究需要同时使用法律分析和员工感知测量，人机协同则可连接任务实验、认知研究和管理情境。关于算法与认知的成果提示，系统判断能否转化为管理创新，取决于使用者怎样理解、解释和修正输出[33]。建立同时包含模型指标、决策过程与员工结果的多层数据，是突破单端研究的重要方向。

（四）本土化治理与生成式人工智能

本土治理首先要正视程序公平与责任归因。中国组织普遍重视效率、执行和规模化，但员工对隐私、尊严和公平的期待也在提高。研究发现，算法作出不利判断时，员工可能因为它看似较少受人情影响而提高公平评价，也可能因缺少解释和互动而降低公平评价[34]。组织信任、领导沟通和申诉可及性如何改变这种双重反应，值得在不同用工关系中开展分级比较。

生成式人工智能扩大了岗位说明、面试题、培训材料、绩效反馈、政策问答和职业建议的自动生成能力，同时也带来事实错误、内容偏见、知识产权、敏感信息泄露与责任不清。相较传统预测模型，生成内容语言流畅，使用者更容易高估可靠性。研究和实践都应区分“内容辅助”与“权益决定”：前者经过审核可以提高效率，后者直接涉及录用、晋升、薪酬或处分，必须由有权限的人复核并保留依据。

还应把注意力放在 AI 建议进入管理决定的中间环节。算法权力研究指出，默认选项、排序和界面设计本身就会影响行为[35]。当生成式系统成为管理者的首份草稿或主要信息入口，它可能在不易察觉的情况下改变问题定义和沟通措辞。后续研究有必要记录提示词、模型版本、引用来源及人工修改，比较不同协同方式对判断质量、责任感和员工体验的影响，也要防止管理者借“系统生成”转移其应承担的解释与决定责任。

七、研究结论

人工智能在人力资源管理中的角色已经从事务辅助扩展到预测分析、内容生成和决策支持，研究焦点也随之由流程效率转向价值与责任。国内外文献在总体方向上具有一致性，但国内更加关注数字化转型和组织落地，国外较早进入公平、隐私、劳动控制和问责讨论。两类研究共同提示，AI-HRM 既不能被理解为单纯的提效工具，也不能脱离真实组织条件只作规范性批判。

本文提出的五层框架表明，技术价值需要经过制度和行为过程才能实现。数据与模型限定系统能够提供何种信息，流程和权责决定信息如何被使用，管理者与员工的理解和信任影响建议能否转化为行动，价值规范则划定哪些效率收益可以接受。管理结果必须反向用于修正数据、制度和人机分工，而不是把系统视作部署后即可独立运行的装置。

上述五类张力表明，智能化管理没有“一次建成”的终点，其风险控制应随决策影响程度而变化。日常查询、规则核验等低影响事项，可侧重自动化效率和服务感受；影响适中的应用应设置抽检、持续监测和异常复核；涉及录用、晋升、薪酬或处分的决定，则必须由有权限的人员作出最终判断，并同步提供说明与申诉途径。后续研究还需在本土组织中检验这些安排，增加对中小企业和公共部门的长期观察，结合纵向与混合方法，考察生成式人工智能使用中的协作方式、过程留痕和责任落实是否真正有效。

参考文献：

[1] 罗文豪, 霍伟伟, 赵宜萱, 等. 人工智能驱动的组织与人力资源管理变革: 实践洞察与研究方向[J]. 中国人力资源开发, 2022, 39(1): 4-16.
[2] TAMBE P, CAPPELLI P, YAKUBOVICH V. Artificial intelligence in human resources management: challenges and a path forward[J]. California Management Review, 2019, 61(4): 15-42.
[3] PAN Y, FROESE F J. An interdisciplinary review of AI and HRM: challenges and future directions[J]. Human Resource Management Review, 2023, 33(1): 100924.
[4] VRONTIS D, CHRISTOFI M, PEREIRA V, et al. Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review[J]. The International Journal of Human Resource Management, 2022, 33(6): 1237-1266.
[5] STONE D L, DEADRICK D L, LUKASZEWSKI K M, et al. The influence of technology on the future of human resource management[J]. Human Resource Management Review, 2015, 25(2): 216-231.
[6] 张琪, 林佳怡, 陈璐, 等. 人工智能技术驱动下的人力资源管理: 理论研究与实践应用[J]. 电子科技大学学报(社会科学版), 2023, 25(1): 77-84.
[7] BUDHWAR P, MALIK A, DE SILVA M T T, et al. Artificial intelligence—challenges and opportunities for international HRM: a review and research agenda[J]. The International Journal of Human Resource Management, 2022, 33(6): 1065-1097.
[8] KELLOGG K C, VALENTINE M A, CHRISTIN A. Algorithms at work: the new contested terrain of control[J]. Academy of Management Annals, 2020, 14(1): 366-410.
[9] LEICHT-DEOBALD U, BUSCH T, SCHANK C, et al. The challenges of algorithm-based HR decision-making for personal integrity[J]. Journal of Business Ethics, 2019, 160(2): 377-392.
[10] AJUNWA I. The black box at work[J]. Big Data & Society, 2020, 7(2): 2053951720966181.
[11] BAROCAS S, SELBST A D. Big data’s disparate impact[J]. California Law Review, 2016, 104(3): 671-732.
[12] RAGHAVAN M, BAROCAS S, KLEINBERG J, et al. Mitigating bias in algorithmic hiring: evaluating claims and practices[C]//Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. New York: Association for Computing Machinery, 2020: 469-481.
[13] DAVIS F D. Perceived usefulness, perceived ease of use, and user acceptance of information technology[J]. MIS Quarterly, 1989, 13(3): 319-340.
[14] TORNATZKY L G, FLEISCHER M, CHAKRABARTI A K. The processes of technological innovation[M]. Lexington: Lexington Books, 1990.

[15] 谢小云, 左玉涵, 胡琼晶. 数字化时代的人力资源管理: 基于人与技术交互的视角[J]. 管理世界, 2021, 37(1): 200-216.
[16] RAISCH S, KRAKOWSKI S. Artificial intelligence and management: the automation-augmentation paradox[J]. Academy of Management Review, 2021, 46(1): 192-210.
[17] BARREDO ARRIETA A, DÍAZ-RODRÍGUEZ N, DEL SER J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI[J]. Information Fusion, 2020, 58: 82-115.
[18] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. Artificial Intelligence Risk Management Framework (AI RMF 1.0)[R/OL]. Gaithersburg: NIST, 2023[2026-07-08]. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.
[19] OECD. OECD AI Principles[EB/OL]. Paris: OECD, 2019[2026-07-08]. https://oecd.ai/en/ai-principles.
[20] 全国人民代表大会常务委员会. 中华人民共和国个人信息保护法[EB/OL]. (2021-08-20)[2026-07-08]. https://www.gov.cn/xinwen/2021-08/20/content_5632486.htm.
[21] BOGEN M, RIEKE A. Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias[R/OL]. Washington, DC: Upturn, 2018[2026-07-08]. https://www.upturn.org/work/help-wanted/.
[22] 赵宜萱, 赵曙明, 栾佳锐. 基于人工智能的人力资源管理: 理论模型及研究展望[J]. 南京社会科学, 2020(2): 36-43.
[23] 李燕萍, 李乐, 胡翔. 数字化人力资源管理: 整合框架与研究展望[J]. 科技进步与对策, 2021, 38(23): 151-160.
[24] 张敏, 赵宜萱. 机器学习在人力资源管理领域中的应用研究[J]. 中国人力资源开发, 2022, 39(1): 71-83.
[25] 陈鼎祥, 刘帮成. 人工智能时代的公共部门人力资源管理: 实践应用与理论研究[J]. 公共管理与政策评论, 2022, 11(4): 38-51.
[26] 张亚莉, 李辽辽, 丁振斌. 组织管理中的人工智能决策: 述评与展望[J]. 外国经济与管理, 2024, 46(10): 18-38.
[27] 喻术红, 常春. 算法劳动管理的伦理诉求及其法制建设研究[J]. 华南理工大学学报(社会科学版), 2022, 24(5): 52-59.
[28] 蒋路远, 曹李梅, 秦昕, 等. 人工智能决策的公平感知[J]. 心理科学进展, 2022, 30(5): 1078-1092.
[29] 刘善仕, 裴嘉良, 葛淳棉, 等. 在线劳动平台算法管理: 理论探索与研究展望[J]. 管理世界, 2022, 38(2): 225-239.
[30] 彭剑锋. 新一代人工智能对组织与人力资源管理的影响与挑战[J]. 中国人力资源开发, 2023, 40(7): 8-14.
[31] 贺伟, 汪林, 吴小玥. 人工智能技术对人力资源管理研究的影响述评[J]. 中国科学基金, 2024, 38(5): 831-840.
[32] 胡心约, 张恬路, 李英武. 基于 AI 的情绪识别在组织中的实践: 现状、未来和挑战[J]. 中国人力资源开发, 2022, 39(1): 57-70.
[33] 付春香, 闫宇航. 算法与认知：人工智能驱动的人力资源管理创新[J]. 科技创业月刊, 2024, 37(8): 22-27.
[34] 高记, 冯婧雯. 不利结果下算法决策对员工公平感的双刃剑效应: 基于归因理论的视角[J]. 中国人力资源开发, 2024, 41(5): 36-53.
[35] 包艳, 马伟博, 廖建桥, 等. 算法权力在管理领域的研究回顾、探索与展望[J]. 外国经济与管理, 2024, 46(5): 120-135.

Artificial Intelligence-Enabled Human Resource Management: Theoretical Perspectives, Practical Paradoxes and Open Research Questions

ZOU Zijing

(Xinyu Qingyun Certified Public Accountants Co., Ltd., Xinyu, Jiangxi, China)

Abstract: AI is no longer used in human resource management only to automate clerical work. Organizations also apply it to screening applicants, recommending training, drafting performance feedback, providing compensation and employee services, and planning the workforce. The literature reports faster information handling and stronger predictive support, but these improvements do not in themselves secure procedural fairness, employee trust, or better organizational results. This article reviews Chinese and international studies published between 2015 and 2026. Rather than counting topics mechanically, it compares how the field has defined AI-enabled HRM, how its questions have changed, which theories have been used, and what disputes have emerged in practice. The review finds a movement from digitizing processes, to prediction and recommendation, and then to governance of values and responsibility. Chinese research has concentrated on implementation and organizational transformation; international research addressed discrimination, labor control, rights, and accountability earlier. A five-level framework links technological conditions, organizational arrangements, individual responses, value constraints, and management outcomes. Five recurring tensions are then examined: efficiency and fairness, data use and privacy, automation and human communication, rapid technical change and organizational capacity, and quantitative indicators and contextual judgment. The article concludes that model performance alone cannot determine whether AI-enabled HRM is responsible. What matters is how outputs enter decisions, who may question them, and who remains accountable. High-impact decisions should therefore be subject to differentiated risk controls, genuine human review, understandable reasons, appeal routes, employee participation, and auditable responsibility.

Keywords: artificial intelligence; human resource management; intelligent human resource management; algorithmic governance; human–AI collaboration; literature review