DOI:10.19571/j.cnki.1000-2995.2026.03.012 CSTR:32148.14.kygl.2026.03.012

可解释性对用户接受AI用于知识创造的影响研究

胡保亮,王嘉雯,闫 帅

(杭州电子科技大学 管理学院,浙江 杭州 310018)

摘要:人工智能(AI)黑箱问题正困扰着用户接受AI用于知识创造。可解释AI被认为是解决AI黑箱问题的重要方案之一。然而,现有研究鲜有探讨可解释性如何影响用户接受AI用于知识创造的机制。为此,本文着重研究了这一机制,这包括AI可解释性影响用户接受AI用于知识创造的路径机制以及用户特征对此路径的调节机制。本文提出了理论假设并对425份问卷数据进行了结构方程模型分析与层次回归分析,检验了相关假设。研究发现:可解释性的完整性、格式化与现时性维度对用户接受AI用于知识创造具有正向的影响;可解释性对用户接受AI用于知识创造的影响是间接的,需要感知有用性与感知易用性的中介。研究也发现:可解释性对用户接受AI用于知识创造的一些影响受学历、使用经验与职位等用户特征的调节。本文的结论由于为AI知识创造研究提供了一个基于可解释性的用户接受模型而对AI知识创造理论与可解释AI理论有贡献,也为企业正确发挥AI可解释性的作用、推进AI知识创造提供了启示。

关键词:人工智能;可解释性;知识创造;用户接受

中图分类号:F270.7 **文献标识码**:A **文章编号**:1000-2995(2026)03-011-0117

0 引言

人工智能(Artificial intelligence, AI)是企业打造竞争优势的新力量[1]。在以往,AI主要用于自动化与提高效率[2]。近来,随着算法、算力、大数据等的融合,AI用于知识创造成为重要趋势[3]。在医疗、金融、零售、制造等多样的产业中,领先企业已经开始将AI嵌入到组织学习系统中并应用AI创造的知识来增强人类的知识进行决策、运营、创新等活动[4]。然而,黑箱问题(The black-box problem)正挑战AI用于知识创造。在AI领域,"黑箱"指的是一种数据驱动的算法,它产生有用的输出(例如知识)、但不会显现内部工作的信息[5]。尽管AI在知识创造方面具有很多优势,但由于模型黑箱,人们甚至是模型开发者都难以理解AI是如何将知识创造出来的[6]。如果不能理解AI如何进行知识创造,那就难以评估AI创造的知识。人们因而担心使用AI创造的知识[7],这严重制约了用户对AI用于知识创造的接受。

可解释AI被提出作为解决AI黑箱问题的方案,帮助实现更加透明的AI,从而推动用户接受AI[6]。学者主张将可解释AI理解为技术,重点研究了如何开发可解释的工具和方法来帮助理解AI的行为与输出[8]。例如,Dong等[9]提出了一种可解释的深度学习模型来缓解与深度神经网络相关的低可解释性问题;Kim等[10]建议将人机交互引入到模型开发中以完善可解释AI系统。进一步地,一些学者发现可解释性方法之间存在一些重叠,但在大多数情况下,每种方法似乎都在解决不同的问题,因而主张不应该单独使用不同的方法,而应该组合应用它们或者将它们作为组件开发新的技术[11]。例如,Love等[5]研究了不同透明方法与不透明方法的性能权衡与组合应用问题。当前研究为减少黑箱问题对用户接受AI用于知识创造的制约起到了积极作用。然而,当前研究也存在一些不足:

首先,当前研究缺乏从用户视角探讨可解释性对用户接受AI用于知识创造的影响。当前大多数研究主要围绕

收稿日期:2024-06-07;**修回日期**:2024-12-17.

基金项目:国家社会科学基金重点项目:"人工智能推动大中小企业融通创新的新模式、新困境与政策优化研究"(23AGL009,2023.09—2026.08);浙江省自然科学基金项目:"人机知识编排视角下的大数据分析能力形成机制研究"(LY24G020005,2024.01—2026.12)。

作者简介:胡保亮(1980—),男(汉),河南淮滨人,杭州电子科技大学管理学院教授,博士生导师,研究方向:人工智能与创新。
王嘉雯(1998—),女(汉),浙江嵊州人,杭州电子科技大学管理学院硕士研究生,研究方向:人工智能与创新。
闫 帅(1983—),女(汉),山东淄博人,杭州电子科技大学管理学院讲师,研究方向:商业模式创新。

通信作者:胡保亮,E-mail:bhu@hdu.edu.cn

技术视角展开[12],结果是对 AI 模型开发人员呈现出了什么是“好的解释”。这很大可能会导致解释的失败,因为最了解 AI 模型的技术专家并不能判断解释对非专业用户的有用性。可解释 AI 研究不能仅考虑专家也要考虑非专业的用户,因为可解释 AI 的核心问题是针对非专业用户进行解释[7]。

其次,当前研究十分缺乏探讨可解释性如何影响用户接受 AI 用于知识创造的机制。当前研究仅仅指出可解释性能够促进用户接受 AI[13],然而鲜有考虑用户对解释的多样性需求(也即可解释性的多元维度),鲜有探讨可解释性影响 AI 接受的中间路径,更鲜有探讨面向不同用户的可解释性影响 AI 接受的差异。缺乏这些研究也会导致解释的失败,因为无法指导不同的用户根据他们的不同需求对解释做出有效反应。

基于以上,本文旨在从用户视角探讨可解释性影响用户接受 AI 用于知识创造的机制。具体来说,本文采用 Haque 等[6]提出的用户视角的可解释性维度(完整性、准确性、格式化与现时性),将它们作为外部刺激因素并与感知有用性、感知易用性这两种驱动用户接受 AI 的心理动机与心理认知进行关联,探讨可解释性影响用户接受 AI 用于知识创造的路径机制。在此基础上,本文引入用户学历、经验与职位作为调节变量,研究可解释性影响用户接受 AI 用于知识创造的权变机制。

1 研究设计

1.1 理论基础与研究假设

1.1.1 AI 用于知识创造

AI 可以被描述为建立在算法基础上的技术、系统或机器,能够模仿人类智能[5]。AI 为人类带来了巨大的机会。在以往,知识创造由人来进行,起始于人的经验,并经过社会化、外部化、组合化和内部化产生新的知识[14]。如今,AI 也能用于知识创造[3]。AI 进行知识创造起始于数据,通过机器学习从海量数据中挖掘和提取知识[1]。AI 进行知识创造与人类进行知识创造并不矛盾,二者是协同关系。一方面,二者都是知识创造的方式,都能为人类带来新知识[15]。另一方面,二者又是交互的,表现在 AI 进行知识创造离不开人类将自身的知识与经验转移给 AI,而人类又可以在吸收与整合 AI 创造的知识的基础上创造新的知识[16]。

1.1.2 可解释性对用户接受 AI 用于知识创造的影响

AI 在过去十年中持续取得重大的进步,越来越多地被用于解决许多问题甚至是过去难以解决的问题。然而,这些杰出的成就伴随着利用缺乏透明度的黑箱模型[11]。机器学习模型,如深度学习,就是这样的黑箱,它们的基本结构是复杂的、非线性的,很难向外行人解释和说明,即便是专家也经常无法解释它们的内部机制[5]。因此,对于用户而言,将任务委托给一个无法自我解释的系统自然是有风险的、最起码是不放心的[9]。考虑到这一点,可解释 AI 被提出[11,17]。可解释 AI 从非符号的、统计的机器学习模型中自动构建一个符号的、人类可以理解的模型,旨在使人类能够理解 AI 系统如何达成特定的决定或结果,并为人类推理提供见解和使用系统提供逻辑[18]。

最近的研究趋势主张可解释 AI 并不是一个技术概念[6]。他们指出可解释 AI 不是“更多人工智能”,而是人机交互的一个问题,相关利益者不仅包括技术专家也包括最终用户,因为可解释 AI 的主要目标是提升 AI 的可解释性进而通过 AI 的可解释性满足用户的需求[11]。从用户角度来看,可解释 AI 展现的可解释性包括四个维度,分别为完整性(Completeness)、准确性(Accuracy)、格式化(Format)与现时性(Currency)[6]。完整性是指为用户提供的所有必要的解释,包括文本解释中的对决策过程和算法原理的详细描述、视觉解释中的不同属性的指标与图像以及情境信息与参考等,完整的信息可以增加用户的信任与可靠性感知[6]。准确性是指为用户提供准确的解释,可以帮助验证 AI 的输出并激励用户使用 AI,它一般基于用户对提供的预测确定性评估信息、算法决策程序、主张和证据以及领域专家参与开发过程信息等解释的感知[6]。格式化是指为用户提供格式化的解释,例如文本解释、视觉解释、听觉解释、混合解释等;用户更喜欢混合解释,这种解释能够增强用户信任感知[6]。用户经常会要求 AI 提供随需应变信息,这就要求 AI 提供的解释要具有现时性,它是指为用户提供的解释要包括最新的信息[6]。

从用户的角度来看,当 AI 具有可解释性时,用户能够获得 AI 提供的解释[6],这将帮助他们理解 AI 的行为与输出[5],进而帮助他们洞察 AI 的优势与风险[12]。相应地,用户将会减少将 AI 用于知识创造的疑惑与抵制,以及增加将 AI 用于知识创造的动力与意愿。可解释性也能满足法律和监管要求[11],这有利于提升 AI 应用的合法性,进而有利于用户接受 AI 用于知识创造。具体到可解释性的四个维度,它们是用户对 AI 提出的四个解释需求[6],因而,从用户的角度来看,当 AI 具有可解释性时,意味着用户解释需求可以得到满足,相应地,用户倾向于接受 AI 用于知识创造。进一步地,从前文四个维度介绍可知,它们能够增强用户对 AI 的信任感知,这也有助于用户接受 AI 用于知识创造。综上,提出如下假设:

H1:可解释性对用户使用 AI 进行知识创造的意向具有显著的正向影响。

H1a:完整性对用户使用 AI 进行知识创造的意向具有显著的正向影响。

H1b:准确性对用户使用 AI 进行知识创造的意向具有显著的正向影响。

H1c:格式化对用户使用 AI 进行知识创造的意向具有显著的正向影响。

H1d:现时性对用户使用 AI 进行知识创造的意向具有显著的正向影响。

1.1.3　感知有用性与感知易用性的中介作用

用户视角的可解释 AI 研究指出 AI 提供的可解释性要与用户动机相关联,因为解释只有在满足用户动机时才能驱动用户意向与行为[10]。技术接受理论指出感知有用性与感知易用性是驱动技术接受的主要动机,它们中介了外部因素对技术接受的影响,其中感知有用性是指用户认为使用特定系统会提高工作绩效的程度、感知易用性是指用户认为使用特定系统的容易程度[19]。学者发现感知有用性与感知易用性也是驱动用户接受 AI 的重要动机[20]。例如,Lim 和 Zhang[21]发现感知有用性与感知易用性促使 AI 用于新闻服务。然而,这些研究鲜有探讨可解释性如何通过感知有用性与感知易用性影响用户接受 AI。此外,这些研究主要关心用户接受 AI 用于自动化,较少探讨用户接受 AI 用于知识创造。按照技术接受理论[20],感知有用性与感知易用性能够中介可解释性对用户接受 AI 用于知识创造的影响。接下来,论证它们的中介作用。

AI 用于知识创造并不是为了取代用户的知识创造而是为了增强用户的知识创造,从而实现 AI 与用户协同与互补[15, 22]。这就需要用户学习如何与 AI 互动,包括掌握 AI 的算法性能、知识创造行为、输出特性等新的知识[23],也包括如何整合 AI 创造的新知识[24]。解释是社会性的,它们是知识的传递,能够帮助用户获得互动所需的新知识[25]。解释的完整性、准确性、现时性与格式化进一步确保了用户获得互动所需的新知识是完整的、准确的、实时的与易吸收的[6]。这些有利于用户与 AI 实现知识创造的协同与互补,进而提升了用户对 AI 是有用与易用的认知。显然,感知有用性与感知易用性又将促进用户接受 AI 用于知识创造。

感知有用性与感知易用性的中介作用也可以从信任角度来理解。可解释性能够帮助用户理解 AI,因而被认为是用户信任 AI 的前提[5, 11, 26]。例如,先前的研究表明用户对机器学习模型的信任会根据陈述和观察到的准确性信息的增加而增加(准确性),为用户提供情境信息、历史数据和最新参考可以增强对系统的信任(完整性与现时性),可视化解释可以提高系统的可见性、可理解性、可观察性和信任度(格式化)[6]。而信任会增进用户的感知有用性与感知易用性进而导致用户接受 AI 用于知识创造。综上,提出如下假设:

H2:感知有用性与感知易用性能在可解释性影响用户使用 AI 进行知识创造意向中起中介作用。

H2a:感知有用性与感知易用性能在完整性影响用户使用 AI 进行知识创造意向中起中介作用。

H2b:感知有用性与感知易用性能在准确性影响用户使用 AI 进行知识创造意向中起中介作用。

H2c:感知有用性与感知易用性能在格式化影响用户使用 AI 进行知识创造意向中起中介作用。

H2d:感知有用性与感知易用性能在现时性影响用户使用 AI 进行知识创造意向中起中介作用。

1.1.4　用户特征的调节作用

近来,一些学者指出 AI 的可解释性主要服务于用户[10]。他们进一步指出不同特征的用户在利用 AI 方面具有不同的领域知识、目标与认知负荷,因而,不同特征的用户对可解释性的需要并不相同[27]。参照这些观点,本文认为用户特征是可解释性作用发挥所依赖的情境。学历、职位与经验等常常被用来表征用户特征而被引入 AI 接受研究中[28-29]。不同于这些研究将它们作为控制变量,本文将它们作为调节变量。

学历反映了用户所具备的工作领域专业知识。一般来说,在同等条件下,学历越高,用户所具备的专业知识越多。无论 AI 在解释自己的行为、模型和过程方面有多好,用户如果学历不高、没有足够的工作领域专业知识,他们理解 AI 的解释就变得非常困难[7]。在此情况下,解释的完整性、准确性、格式化与现时性并不能实现解释的关键目标,如增进理解、减少不确定性意识和校准信任[6]。这阻碍了用户对 AI 是否有用、易用的感知。而当用户学历高时,则是相反的情况。综上,提出如下假设:

H3:用户学历分别正向调节了可解释性对感知有用性与感知易用性的影响。

H3a:用户学历分别正向调节了完整性对感知有用性与感知易用性的影响。

H3b:用户学历分别正向调节了准确性对感知有用性与感知易用性的影响。

H3c:用户学历分别正向调节了格式化对感知有用性与感知易用性的影响。

H3d:用户学历分别正向调节了现时性对感知有用性与感知易用性的影响。

使用经验反映了用户对 AI 的认知负荷与知识情况。可解释 AI 研究通常假设用户对 AI 具有一定的知识与认知,这样才能确保 AI 提供的解释被理解[18]。一般来说,在同等条件下,使用经验越多,用户具有的 AI 方面的知识越多、对 AI 的认知负荷越小。在用户能够认知 AI 的情况下,解释的有用性才能体现出来[7]。相应地,用户将会感知到 AI 用于知识创造的有用性与易用性。反之,如果使用经验较少,用户对 AI 将会处于一种高认知负荷状态。在此状态下,由于没有能力处理 AI 提供的解释信息,用户对 AI 的信任度会降低[11],相应地,解释的完整性、准确性、格式化与现时性也失去了意义。这些将阻碍用户对 AI 是有用与易用的感知。综上,提出如下假设:

H4:用户经验分别正向调节了可解释性对感知有用性与感知易用性的影响。

H4a:用户经验分别正向调节了完整性对感知有用性与感知易用性的影响。

H4b:用户经验分别正向调节了准确性对感知有用性与感知易用性的影响。

H4c:用户经验分别正向调节了格式化对感知有用性与感知易用性的影响。

H4d:用户经验分别正向调节了现时性对感知有用性

与感知易用性的影响。

不同职位的用户,对 AI 提供解释的态度并不相同。一般来说,职位越高,用户可能越不需要 AI 提供的解释,因为 AI 提供的解释在促进用户理解 AI 的同时也会使得用户的工作、习惯甚至生活变得更加透明[18]。这将减少 AI 提供解释的意义甚至困扰职位高的用户。因此,职位越高,用户可能越反对 AI 提供解释。面对这样的用户,如果继续强调解释的完整性、准确性、格式化与现时性,用户对 AI 的有用性与易用性感知水平就会降低。综上,提出如下假设:

H5:用户职位分别负向调节了可解释性对感知有用性与感知易用性的影响。

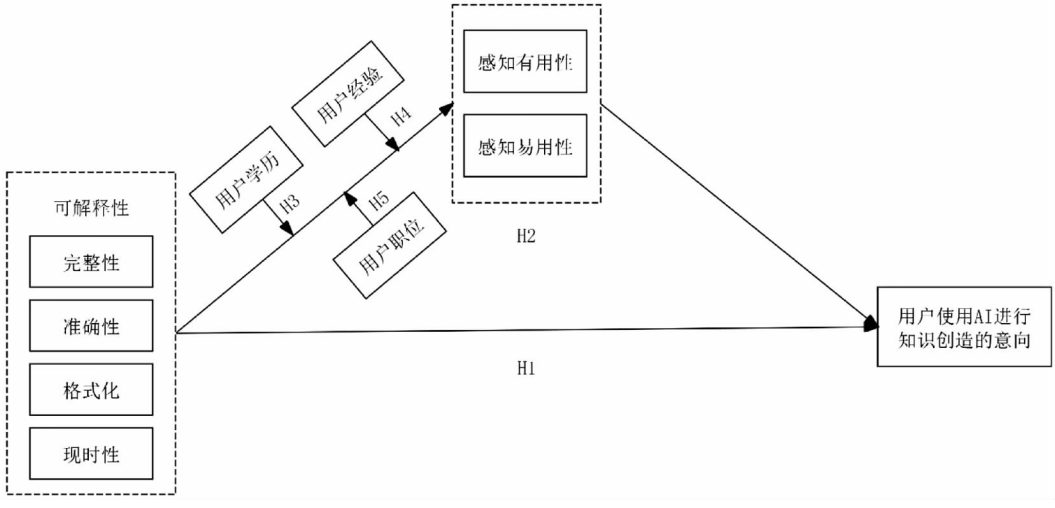
H5a:用户职位分别负向调节了完整性对感知有用性与感知易用性的影响。

H5b:用户职位分别负向调节了准确性对感知有用性与感知易用性的影响。

H5c:用户职位分别负向调节了格式化对感知有用性与感知易用性的影响。

H5d:用户职位分别负向调节了现时性对感知有用性与感知易用性的影响。

汇总以上假设,形成研究模型(见图 1)。



注:H2 表示感知有用性与感知易用性的中介作用。

图 1　研究模型

Figure 1　Research model

1.2　研究方法

1.2.1　数据收集

考虑到 AI 应用并未普及,为了能够接触到被调查对象以及为了能够增加样本容量,AI 领域的研究常常使用基于滚雪球的方便抽样方法收集数据[30]。跟随这些研究,本研究也采用这种方法:首先,面向熟悉的已经在工作中将 AI 应用于知识创造的 MBA 学员等发放问卷;其次,委托他们再向他们熟悉的已在工作中将 AI 应用于知识创造的其他用户发放问卷。最终,我们回收问卷 638 份。我们剔除了工作中无 AI 应用、答题时间过短与过长、回答有规律等无效问卷 213 份,最终得到有效问卷 425 份,样本特征如表 1 所示。

表 1　样本特征

Table 1　Characteristics of the sample

基本特征	样本数	百分比	基本特征	样本数	百分比
用户性别			用户职位		
男性	174	40.9%	员工	240	56.5%
女性	251	59.1%	管理者	185	43.5%
用户学历			使用时长		
专科及以下	41	9.6%	1 年以下	152	35.8%
本科	212	49.9%	1 ~ 3 年	197	46.3%
研究生	172	40.5%	3 年以上	76	17.9%
用户年龄			用户行业		
25 岁及以下	211	49.7%	来自制造业	77	18.1%
26 ~ 30 岁	105	24.7%	来自服务业	348	81.9%
31 岁及以上	109	25.6%			

1.2.2　变量测量

采用 Likert 七级打分法进行变量测量,邀请受访者在“1 =完全不同意”至“7 = 完全同意”之间打分。可解释性的四维度是由 Haque 等[6]提出。为此,基于他们的研究,共计采用 12 个题项测量这四个维度,如表 2 所示。基于 Venkatesh 和 Davis[31],各自使用 4 个题项测量感知有用性与感知易用性,如表 2 所示。基于 Venkatesh 等[32],使用 3 个题项测量用户使用 AI 进行知识创造的意向。本研究将用户学历划分为 3 个等级——专科及以下、本科与研究生;通过用户使用 AI 的时长来表征用户使用 AI 的经验;使用 0 - 1 变量测量用户职位,将一般员工赋值为 0、将管理者赋值为 1。

1.2.3　偏差分析

Harman 单因子法被用于检测共同方法偏差。本研究将自变量、中介变量与因变量的所有题项一起进行探索性因子分析。由于调节变量都是单一题项变量,故探索性因子分析未包括它们。结果显示 7 个因子被提取出来,它们累计解释了 70.020% 的总方差。其中,第一个因子解释了 29.382% 的方差、未超过解释总方差的 50% 。这些结果表明共同方法偏差在本研究中并不显著。

1.2.4　信度与效度

使用 Cronbach′s α 与 CR 值检验变量测量信度。3 个调节变量都是单一题项变量,故信效度分析未包括它们。结果显示(如表 2),各个变量的 Cronbach′s α 与 CR 值都大于 0.7,表明对它们的测量符合信度要求。使用验证性因子分析法检验测量效度。结果显示,包含各个变量的 7 因子模拟拟合结果良好:$\chi^2/df = 1.797$,小于 3;RMSEA = 0.043,小于 0.08;IFI = 0.964,大于 0.9;GFI = 0.927,大于 0.9;CFI = 0.964,大于 0.9;NFI = 0.922,大于 0.9;TLI = 0.956,大于 0.9;RFI = 0.906,大于 0.9。各个变量对应题项的因子载荷均大于 0.5(如表 2 所示)。进一步地,根据因子载荷计算各个变量的 AVE,结果显示 AVE 均接近或超过 0.5(如表 2 所示),表明对它们的测量符合收敛效度要求。通过比较各个变量 AVE 的平方根与变量间的相关系数,发现前者大于后者(如表 3 所示),表明对它们的测量符合区分效度要求。

表 2　信度与效度

Table 2　Reliability and validity tests results

变量与题项	因子载荷
完整性(Cronbach′s α = 0.760, CR = 0.7745, AVE = 0.5345)	
1. 该 AI 系统为我提供了一整套解释信息	0.731
2. 该 AI 系统为我提供了完整的解释信息	0.779
3. 该 AI 系统给我提供了我需要的所有解释信息	0.680
准确性(Cronbach′s α = 0.743, CR = 0.7458, AVE = 0.4951)	
1. 该 AI 系统能提供给我正确的解释信息	0.702
2. 我从该 AI 系统那里获得的解释信息很少会有错误	0.658
3. 由该 AI 系统提供的解释信息是精确的	0.748
格式化(Cronbach′s α = 0.781, CR = 0.7818, AVE = 0.5444)	
1. 该 AI 系统能为我提供格式良好的解释信息	0.734
2. 该 AI 系统提供给我的解释信息能被很好地展示出来	0.719
3. 该 AI 系统提供的解释信息能被清晰地展示出来	0.760
现时性(Cronbach′s α = 0.762, CR = 0.7661, AVE = 0.5222)	
1. 该 AI 系统能为我提供最近产生的解释信息	0.739
2. 该 AI 系统能为我提供最新产生的解释信息	0.735
3. 来自该 AI 系统的解释信息总是及时更新的	0.693
感知有用性(Cronbach′s α = 0.843, CR = 0.8439, AVE = 0.5748)	
1. 使用该 AI 系统提高了我的工作绩效	0.769
2. 在我的工作中使用该 AI 系统提高了我的生产率	0.779
3. 使用该 AI 系统提高了我的工作有效性	0.740
4. 该 AI 系统在我的工作中很有用	0.744
感知易用性(Cronbach′s α = 0.803, CR = 0.8047, AVE = 0.5075)	
1. 我与该 AI 系统的交互是明确与易懂的	0.718
2. 与该 AI 系统交互不需要我付出很多精力	0.689
3. 我发现该 AI 系统是容易使用的	0.710
4. 让该 AI 系统做我想让它做的事情很容易	0.732
使用意向(Cronbach′s α = 0.852, CR = 0.8546, AVE = 0.6628)	
1. 我想要继续将该 AI 系统用于知识创造	0.745
2. 我预测我会在知识创造中继续使用该 AI 系统	0.837
3. 我计划继续将该 AI 系统用于知识创造	0.856

表 3　变量相关系数

Table 3　Correlation coefficients of variables

变量	*CO*	*AC*	*FO*	*CU*	*PU*	*PEU*	*KBI*
完整性(*CO*)	0.731						
准确性(*AC*)	0.619***	0.704					
格式化(*FO*)	0.642***	0.594***	0.738				
现时性(*CU*)	0.547***	0.562***	0.582***	0.723			
感知有用性(*PU*)	0.607***	0.534***	0.622***	0.528***	0.758		
感知易用性(*PEU*)	0.546***	0.497***	0.614***	0.540***	0.562***	0.712	
使用意向(*KBI*)	0.388***	0.325***	0.389***	0.389***	0.452***	0.392***	0.814

注:* 表示 $p < 0.05$, 表示 $p < 0.01$,*** 表示 $p < 0.001$;对角线数据为相应变量的 *AVE* 平方根。**

2　实证分析

2.1　直接效应与中介效应分析

通过构造结构方程模型，检验可解释性的四个维度对用户使用 AI 进行知识创造意向的直接影响（假设 1a、1b、1c、1d）与间接影响（假设 2a、2b、2c、2d）。模型拟合良好：$\chi^2/df = 1.809$，小于 3；RMSEA = 0.044，小于 0.08；IFI = 0.963，大于 0.9；GFI = 0.927，大于 0.9；CFI = 0.963，大于 0.9；NFI = 0.921，大于 0.9；TLI = 0.955，大于 0.9；RFI = 0.905，大于 0.9。然而，与假设预期不一致，研究结果显示完整性（$p = 0.385 > 0.1$）、准确性（$p = 0.487 > 0.1$）、格式化（$p = 0.814 > 0.1$）与现时性（$p = 0.130 > 0.1$）都不能直接影响用户使用 AI 进行知识创造的意向。因而假设 1a，1b，1c，1d 均不成立。一个可能原因是在面对 AI 提供的可解释性刺激时，考虑到知识创造极其重要，用户接受 AI 用于知识创造需要依赖一些心理认知与动机产生。这一原因也与用于解释个体对外部刺激响应的“刺激—有机体—反应”（SOR）理论观点一致。SOR 理论认为刺激并不直接导致反应，而是通过认知与动机等中介来实现。

本研究通过删除这四条路径进行模型修正、得到新的结构方程模型。修正后的模型拟合良好且较修正前的模型改善：$\chi^2/df = 1.790$，小于 3；RMSEA = 0.043，小于 0.08；IFI = 0.963，大于 0.9；GFI = 0.926，大于 0.9；CFI = 0.963，大于 0.9；NFI = 0.921，大于 0.9；TLI = 0.956，大于 0.9；RFI = 0.906，大于 0.9。路径系数如表 4 与图 2 所示。从表 4 可知，完整性显著正向影响感知有用性（$\beta = 0.330$，$p < 0.01$）且感知有用性显著正向影响使用意向（$\beta = 0.394$，$p < 0.001$），这表明完整性能够通过感知有用性的中介作用影响使用意向、效应为 0.130，这与假设预期一致。与此同时，表 4 显示完整性不能显著影响感知易用性（$\beta = 0.095$，$p > 0.1$），这表明完整性不能通过感知易用性的中介作用影响使用意向，这与假设预期不一致。一个可能的原因是完整性强调提供很多的解释与信息[6]，这会增加用户的工作量进而难以产生 AI 是易用的感知。综上，假设 2a 得到部分验证。

表 4　路径系数及其显著性

Table 4　Path coefficients and their significance

关系路径	标准化系数	*S. E.*	*C. R.*	*P*
感知有用性 < - - - 完整性	0.330	0.113	2.619	0.009
感知易用性 < - - - 完整性	0.095	0.121	0.716	0.474
感知有用性 < - - - 准确性	-0.030	0.103	-0.250	0.803
感知易用性 < - - - 准确性	-0.074	0.112	-0.580	0.562
感知有用性 < - - - 格式化	0.419	0.127	3.295	***
感知易用性 < - - - 格式化	0.601	0.147	4.145	***
感知有用性 < - - - 现时性	0.148	0.082	1.587	0.113
感知易用性 < - - - 现时性	0.227	0.090	2.248	0.025
使用意向 < - - - 感知有用性	0.394	0.101	5.178	***
使用意向 < - - - 感知易用性	0.211	0.099	2.810	0.005

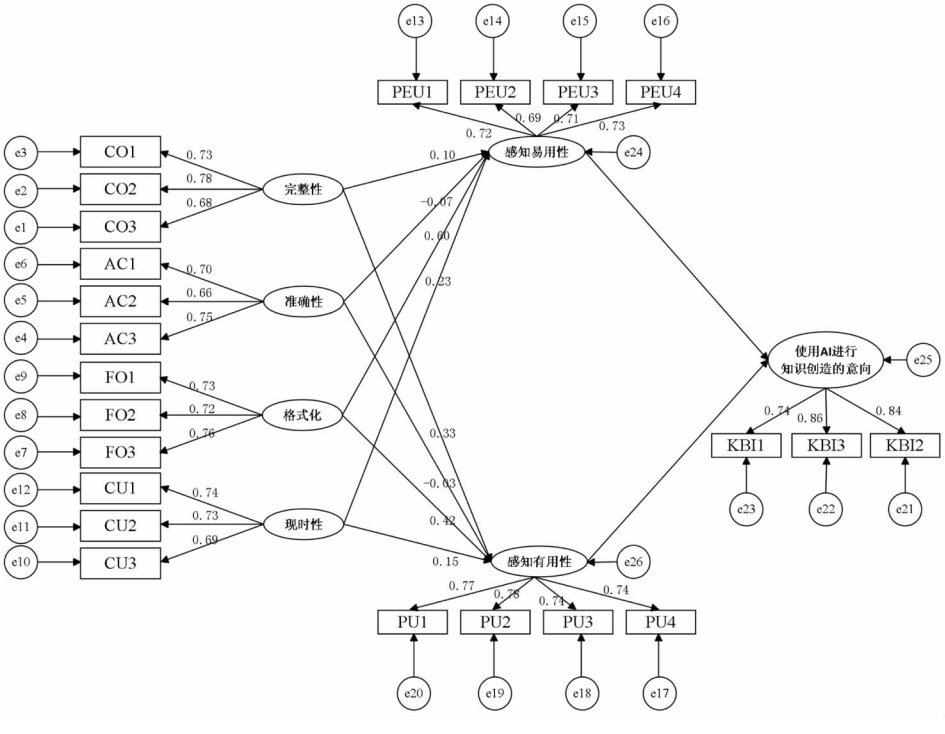


图 2　修正后的结构方程模型及其路径系数

Figure 2　Revised structural equation model and its path coefficients

从表 4 可知,尽管感知有用性($\beta = 0.394, p < 0.001$)与感知易用性($\beta = 0.211, p < 0.01$)均能显著正向影响使用意向,但准确性不能显著正向影响感知有用性($\beta = -0.030, p > 0.1$)与感知易用性($\beta = -0.074, p > 0.1$),这表明准确性不能通过它们的中介作用影响使用意向,因而,假设 2b 未得到证实。一个可能的原因是,用户参与是提高解释准确性的重要方法[10],这就意味着越是提高解释准确性、越是需要用户深度参与(例如时间精力投入、隐性知识投入),相应地,会增加用户关于 AI 是成本高昂的、是复杂的认识以及被替代的担忧。

从表 4 可知,格式化均能显著正向影响感知有用性($\beta = 0.419, p < 0.001$)与感知易用性($\beta = 0.601, p < 0.001$)且感知有用性($\beta = 0.394, p < 0.001$)与感知易用性($\beta = 0.211, p < 0.01$)均能显著正向影响使用意向,这表明格式化不仅能够通过感知有用性的中介作用影响使用意向、效应为 0.165,也能通过感知易用性的中介作用影响使用意向、效应为 0.127,因而,假设 2c 得到证实。

从表 4 可知,现时性不能显著影响感知有用性($\beta = 0.148, p > 0.1$),这表明现时性不能通过感知有用性的中介作用影响使用意向,这与假设预期不一致。一个可能的原因是,现时性强调将最新的信息用于解释[6],但最新的信息往往未经过验证是有价值的,这制约了用户产生 AI 是有用的感知。与此同时,表 4 显示现时性显著正向影响感知易用性($\beta = 0.227, p < 0.05$)且感知易用性显著正向影响使用意向,这表明现时性能够通过感知易用性的中介作用影响使用意向、效应为 0.050,这与假设预期一致。因而,假设 2d 得到部分验证。

2.2　调节效应分析

使用层次回归分析检验调节效应。建立模型 1 - 8 检验假设 3(3a,3b,3c,3d)的成立情况,如表 5 所示。这 8 个模型最大的 VIF 为 2.216、小于 10,表明它们不存在共线性问题。如表 5 所示,模型 1 与 5 为基础模型,检验控制变量的影响;模型 2 与 6 分别在模型 1 与 5 的基础上加入完整性、准确性、格式化与现时性这四个自变量,模型 3 与 7 分别在模型 2 与 6 的基础上加入用户学历这个调节变量;模型 4 与 8 分别在模型 3 与 7 的基础上加上学历 × 完整性、学历 × 准确性、学历 × 格式化、学历 × 现时性这四个调节项。如表 5 所示,学历 × 完整性对感知有用性具有显著的正向影响($\beta = 0.184, p < 0.001$),表明用户学历正向调节了完整性对感知有用性的影响,这与假设预期一致。与此同时,学历 × 格式化对感知有用性具有显著的负向影响($\beta = -0.083, p < 0.1$)、学历 × 格式化对感知易用性具有显著的负向影响($\beta = -0.105, p < 0.05$),表明用户学历负向调节了格式化对感知有用性与感知易用性的影响,这些与假设预期相反。可能的原因是,对于拥有较高学历的用户来说,他们由于具有更多专业知识而更加注重解释的内容而非解释的格式。追求解释的呈现形式(如多种模态的解释)可能还会为这些用户带来冗余。

表 5　用户学历的调节效应

Table 5　Moderating effect of the user's educational background

自变量	因变量							
	感知有用性				感知易用性			
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8
用户性别	0.004	-0.021	-0.023	-0.029	0.027	0.002	0.008	0.011
用户年龄	0.021	-0.028	-0.026	-0.026	0.077	0.036	0.026	0.028
企业规模	0.098 *	-0.003	-0.005	0.008	0.112 *	0.025	0.032	0.037
完整性		0.273 ***	0.272 ***	0.304 ***		0.158 **	0.162 **	0.173 **
准确性		0.103 *	0.104 *	0.079		0.075	0.071	0.059
格式化		0.303 ***	0.302 ***	0.278 ***		0.339 ***	0.342 ***	0.328 ***
现时性		0.145 **	0.145 **	0.159 **		0.213 ***	0.214 ***	0.224 ***
用户学历			0.011	0.008			-0.047	-0.049
学历 × 完整性				0.184 ***				0.043
学历 × 准确性				-0.044				0.036
学历 × 格式化				-0.083 +				-0.105 *
学历 × 现时性				-0.015				0.009
R^2	0.010	0.485	0.485	0.507	0.019	0.452	0.454	0.460
调整后 R^2	0.003	0.476	0.475	0.493	0.012	0.443	0.443	0.444
R^2 变化	0.010	0.475	0.000	0.022	0.019	0.433	0.002	0.006
F 统计值	1.453	56.101 ***	48.991 ***	35.302 ***	2.675 *	49.128 ***	43.231 ***	29.245 ***

注:+ 表示 $p < 0.1$, * 表示 $p < 0.05$, ** 表示 $p < 0.01$, * 表示 $p < 0.001$。**

在模型 2 与 6 的基础上，建立模型 9 – 12 检验假设 4(4a，4b，4c，4d)的成立情况，如表 6 所示。这 4 个模型最大的 VIF 为 2.156、小于 10，表明它们不存在共线性问题。模型 10 的结果显示(如表 6 所示)，经验 × 完整性、经验 × 准确性、经验 × 格式化、经验 × 现时性这四个调节项均未显著影响感知有用性。模型 12 的结果显示(如表 6 所示)，经验 × 完整性对感知易用性具有显著的负向影响($\beta = -0.097$，$p < 0.1$)，表明用户经验负向调节了完整性对感知易用性的影响，但这与假设预期相反。一个可能的原因是，对于拥有丰富使用经验的用户来说，提供很多解释并不必要甚至增加用户负担。表 6 亦显示，经验 × 现时性对感知易用性具有显著的正向影响($\beta = 0.120$，$p < 0.05$)，表明用户经验正向调节了现时性对感知易用性的影响，这与假设预期一致；经验 × 准确性、经验 × 格式化这两个调节项未显著影响感知易用性。

表 6　用户经验与用户职位的调节效应

Table 6　Moderating effects of the user's experiences and positions

自变量	因变量							
	感知有用性		感知易用性		感知有用性		感知易用性	
	模型 9	模型 10	模型 11	模型 12	模型 13	模型 14	模型 15	模型 16
用户性别	−0.028	−0.029	−0.003	0.001	−0.022	−0.021	0.001	0.013
用户年龄	0.005	0.004	0.061	0.062	−0.023	−0.030	0.045	0.037
企业规模	0.004	0.004	0.031	0.040	−0.001	0.002	0.028	0.029
完整性	0.271***	0.275***	0.156**	0.162**	0.272***	0.274***	0.155**	0.178**
准确性	0.112*	0.106*	0.082	0.065	0.103*	0.099*	0.075	0.063
格式化	0.307***	0.306***	0.342***	0.346***	0.305***	0.308***	0.343***	0.348***
现时性	0.136**	0.141**	0.206***	0.208***	0.144**	0.143**	0.210***	0.197***
用户经验	−0.073+	−0.074+	−0.055	−0.056				
用户职位					−0.015	−0.014	−0.028	−0.027
经验 × 完整性		−0.032		−0.097+				
经验 × 准确性		−0.023		−0.017				
经验 × 格式化		0.051		0.006				
经验 × 现时性		0.017		0.120*				
职位 × 完整性						−0.016		−0.171**
职位 × 准确性						0.036		0.052
职位 × 格式化						0.033		0.105*
职位 × 现时性						−0.020		−0.006
R^2	0.489	0.491	0.454	0.465	0.485	0.487	0.453	0.468
调整后 R^2	0.479	0.476	0.444	0.450	0.475	0.472	0.442	0.452
R^2 变化	0.004	0.002	0.002	0.011	0.000	0.002	0.001	0.015
F 统计值	49.737***	33.096***	43.265***	29.878***	49.009***	32.587***	43.003***	30.151***

注：+ 表示 $p < 0.1$，* 表示 $p < 0.05$， 表示 $p < 0.01$，*** 表示 $p < 0.001$。**

在模型 2 与 6 的基础上，建立模型 13 – 16 检验假设 5(5a，5b，5c，5d)的成立情况，如表 6 所示。这 4 个模型最大的 VIF 为 2.200、小于 10，表明它们不存在共线性问题。模型 14 的结果显示(如表 6 所示)，职位 × 完整性、职位 × 准确性、职位 × 格式化、职位 × 现时性这四个调节项均未显著影响感知有用性。模型 16 的结果显示(如表 6 所示)，职位 × 完整性对感知易用性具有显著的负向影响($\beta = -0.171$，$p < 0.001$)，表明用户职位负向调节了完整性对感知易用性的影响，这与假设预期一致；职位 × 准确性、职位 × 现时性这两个调节项未有显著影响感知易用性。表 6 亦显示职位 × 格式化对感知易用性具有显著的正向影响($\beta = 0.105$，$p < 0.05$)，表明用户职位正向调节了格式化对感知易用性的影响，但与假设预期相反。一个可能的原因是，格式化强调解释的良好呈现形式，相对其他维度而言，为职位高的用户造成的困扰较小，且能满足职位高的用户对良好表达形式的偏好。

3　主要研究结论及启示

3.1　主要结论

为了破解黑箱问题对用户接受 AI 用于知识创造的制约，本研究着重从用户视角探讨了 AI 的可解释性影响用户接受 AI 用于知识创造的机制，主要取得了如下研究

结论:

(1)可解释性对用户接受 AI 用于知识创造的影响表现为完整性、格式化与现时性三个维度的效应。这表明并不是所有的可解释性维度都能影响用户接受 AI 用于知识创造,例如准确性维度。这一结论揭示了可解释性的各个维度与用户接受 AI 用于知识创造的总体关系,不仅为用户接受 AI 用于知识创造识别出了前因而贡献了 AI 知识创造理论,而且为 AI 可解释性各个维度识别出了结果而贡献了可解释 AI 理论。

(2)可解释性对用户接受 AI 用于知识创造的影响是间接的,需要感知有用性与感知易用性的中介。具体来说,可解释性的完整性维能够通过感知有用性间接影响用户接受 AI 用于知识创造、可解释性的格式化维能够分别通过感知有用性与感知易用性间接影响用户接受 AI 用于知识创造、可解释性的现时性维能够通过感知易用性间接影响用户接受 AI 用于知识创造。这些结论揭示出了可解释性的各个维度影响用户接受 AI 用于知识创造的路径机制,为 AI 知识创造研究提供了一个基于可解释性的用户接受模型。

(3)可解释性对用户接受 AI 用于知识创造的一些影响受学历、使用经验与职位等用户特征的调节。表现在,用户学历正向调节了完整性维对感知有用性的影响、负向调节了格式化维对感知有用性与感知易用性的影响;用户经验负向调节了完整性维对感知易用性的影响、正向调节了现时性维对感知易用性的影响;用户职位负向调节了完整性维对感知易用性的影响、正向调节了格式化维对感知易用性的影响。这些结论揭示出了可解释性的各个维度影响用户接受 AI 用于知识创造的权变机制,连同前文结论一起系统回答了用户如何面对可解释性做出接受 AI 用于知识创造反映的问题,这也为突破计算机科学、使用社会科学框架与方法研究可解释 AI 提供了一个有益尝试与示范。

3.2　管理启示

本文结论具有如下管理启示:

(1)应从用户的角度评估与满足对 AI 可解释性的多样化需求。为了促进用户接受 AI 用于知识创造,企业应该意识到用户对 AI 可解释性的需求是具体的而不是笼统的,包括完整性、准确性、格式化与现时性等多个方面。在此基础上,企业应该邀请用户去评估与分析他们在解释的各个方面的需求,进而组织技术专家满足这些需求。需要指出的是,企业过多追求解释的准确性并不必要,而且需要付出较多成本甚至促使用户产生被替代的担忧。

(2)应驱动用户产生 AI 是有用与易用的认知。满足了用户在解释的完整性、格式化与现时性方面的需求并不能直接导致用户接受 AI 用于知识创造。企业还需驱动用户产生 AI 是有用与易用的认知。具体来说,企业应以解释完整性促进用户产生 AI 是有用的认知,应以解释格式化促进用户产生 AI 是有用与易用的认知,应以解释现时性促进用户产生 AI 是易用的认知。

(3)要进行 AI 的可解释性与用户学历、经验与职位的适配。具体来说,用户学历越高,企业应该越是提升解释的完整性与减少解释的格式化;用户学历越低,则相反。用户经验越多,企业应该越是减少解释的完整性与提升解释的现时性;用户经验越少,则相反。用户职位越高,企业应该越是减少解释的完整性与增强解释的格式化;用户职位越低,则相反。

3.3　不足与展望

本研究存在一些有待进一步研究之处。(1)本研究未考虑可解释性多个维度并发的效应。然而,用户接受 AI 用于知识创造也可能受到可解释性多个维度的并发影响。未来可以用 QCA 方法对可解释性多个维度进行组态分析。(2)本研究未考虑用户接受 AI 用于知识创造的行业间差异性。未来针对一些重点行业,如医疗、服务等,研究它们的用户如何针对可解释性做出接受 AI 用于知识创造的反应,提升研究的针对性。

参考文献:

[1] OLAN F, ARAKPOGUN E O, SUKLAN J, et al. Artificial intelligence and knowledge sharing: Contributing factors to organizational performance[J]. Journal of Business Research, 2022, 145:605 - 615.

[2] 张志学,华中生,谢小云. 数智时代人机协同的研究现状与未来方向[J]. 管理工程学报,2024,38(1):1 - 13.
ZHANG Zhixue, HUA Zhongsheng, XIE Xiaoyun. Research status and future directions of human - computer collaboration in the era of digital intelligence[J]. Journal of Industrial Engineering and Engineering Management, 2024, 38(1): 1 - 13.

[3] HARFOUCHE A, QUINIO B, SABA M, et al. The recursive theory of knowledge augmentation: Integrating human intuition and knowledge in artificial intelligence to augment organizational knowledge[J]. Information Systems Frontiers, 2023, 25(1):55 - 70.

[4] MAKARIUS E E, MUKHERJEE D, FOX J D, et al. Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization[J]. Journal of Business Research, 2020, 120:262 - 273.

[5] LOVE P E D, FANG W, MATTHEWS J, et al. Explainable artificial intelligence (XAI): Precepts, models, and opportunities for research in construction[J]. Advanced Engineering Informatics, 2023, 57:102024.

[6] HAQUE A B, ISLAM A K M N, MIKALEF P. Explainable artificial intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research[J]. Technological Forecasting and Social Change,

2023, 186:122120.

[7] DIKMEN M, BURNS C. The effects of domain knowledge on trust in explainable AI and task performance: A Goyal case of peer - to - peer lending[J]. International Journal of Human - Computer Studies, 2022, 162:102792.

[8] ARRIETA A B, 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.

[9] DONG J, CHEN S, MIRALINAGHI M, et al. Why did the AI make that decision? Towards an explainable artificial intelligence (XAI) for autonomous driving systems[J]. Transportation Research Part C: Emerging Technologies, 2023, 156:104358.

[10] KIM M, KIM S, KIM J, et al. Do stakeholder needs differ? Designing stakeholder - tailored explainable artificial intelligence (XAI) interfaces[J]. International Journal of Human - Computer Studies, 2024, 181:103160.

[11] SAEED W, OMLIN C. Explainable AI (XAI): A systematic meta - survey of current challenges and future opportunities[J]. Knowledge - Based Systems, 2023, 263:110273.

[12] MESKE C, BUNDE E, SCHNEIDER J, et al. Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities[J]. Information Systems Management, 2022, 39(1):53 - 63.

[13] CONATI C, BARRAL O, PUTNAM V, et al. Toward personalized XAI: A case study in intelligent tutoring systems[J]. Artificial Intelligence, 2021, 298:103503.

[14] NONAKA I. A dynamic theory of organizational knowledge creation[J]. Organization Science, 1994, 5(1):14 - 37.

[15] LI J, HUANG J, LIU J, et al. Human - AI cooperation: Modes and their effects on attitudes[J]. Telematics and Informatics, 2022, 73:101862.

[16] VÖSSING M, KÜHL N, LIND M, et al. Designing transparency for effective human - AI collaboration[J]. Information Systems Frontiers, 2022, 24(3):877 - 895.

[17] 张成洪,陈刚,陆天,等. 可解释人工智能及其对管理的影响:研究现状和展望[J]. 管理科学,2021,34(3):63 - 79.
ZHANG Chenghong, CHEN Gang, LU Tian, et al. Explainable artificial intelligence and its impact on management: Research status and prospects[J]. Journal of Management Science, 2021, 34(3):63 - 79.

[18] DE BRUIJN H, WARNIER M, JANSSEN M. The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision - making[J]. Government Information Quarterly, 2022, 39(2):101666.

[19] DAVIS F D. Perceived usefulness, perceived ease of use, and user acceptance of information technology[J]. MIS Quarterly, 1989, 13(3):319 - 340.

[20] 吴俊,张迪,刘涛,等. 人类对人工智能信任的接受度及脑认知机制研究:实证研究与神经科学实验的元分析[J]. 管理工程学报,2024,38(1):60 - 73.
WU Jun, ZHANG Di, LIU Tao, et al. A study on the acceptance of human trust in artificial intelligence and brain cognitive mechanism: A meta - analysis of empirical studies and neuroscience experiments[J]. Journal of Industrial Engineering and Engineering Management, 2024, 38(1):60 - 73.

[21] LIM J S, ZHANG J. Adoption of AI - driven personalization in digital news platforms: An integrative model of technology acceptance and perceived contingency[J]. Technology in Society, 2022, 69:101965.

[22] 杨祎,刘嫣然,李垣. 替代或互补:人工智能应用管理对创新的影响[J]. 科研管理,2021,42(4):46 - 54.
YANG Yi, LIU Yanran, LI Yuan. Substitution or complementation: The impact of AI application and management on innovation[J]. Science Research Management, 2021, 42(4):46 - 54.

[23] JARRAHI M H, ASKAY D, ESHRAGHI A, et al. Artificial intelligence and knowledge management: A partnership between human and AI[J]. Business Horizons, 2023, 66(1): 87 - 99.

[24] LEBOVITZ S, LIFSHITZ - ASSAF H, LEVINA N. To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis[J]. Organization Science, 2022, 33(1):126 - 148.

[25] WYSOCKI O, DAVIES J K, VIGO M, et al. Assessing the communication gap between AI models and healthcare professionals: Explainability, utility and trust in AI - driven clinical decision - making[J]. Artificial Intelligence, 2023, 316:103839.

[26] 孔祥维,王子明,王明征,等. 人工智能使能系统的可信决策:进展与挑战[J]. 管理工程学报,2022,36(6):1 - 14.
KONG Xiangwei, WANG Ziming, WANG Mingzheng, et al. Trustworthy decision - making in artificial intelligence - enabled systems: Progress and challenges[J]. Journal of Industrial Engineering and Engineering Management, 2022, 36(6):1 - 14.

[27] LANGER M, OSTER D, SPEITH T, et al. What do we want from explainable artificial intelligence (XAI)? A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research[J]. Artificial Intelligence, 2021, 296: 103473.

[28] CHONG A Y L, BLUT M, ZHENG S. Factors influencing the acceptance of healthcare information technologies: A meta - analysis[J]. Information & Management, 2022, 59(3):103604.

[29] JANSSEN M, HARTOG M, MATHEUS R, et al. Will algorithms blind people? The effect of explainable AI and decision - makers' experience on AI - supported decision - making in government[J]. Social Science Computer Review, 2022, 40(2):478 - 493.

[30] ULLAH R, BIN ISMAIL H, KHAN M T I, et al. Nexus between ChatGPT usage dimensions and investment decisions making in Pakistan: Moderating role of financial literacy[J]. Technology in Society, 2024, 76: 102454.

[31] VENKATESH V, DAVIS F D. A theoretical extension of the technology acceptance model: Four longitudinal field studies [J]. Management Science, 2000, 46(2): 186 - 204.

[32] VENKATESH V, MORRIS M G, DAVIS G B, et al. User acceptance of information technology: Toward a unified view [J]. MIS Quarterly, 2003, 27(3): 425 - 478.

Research on the impact of explainability on users′ acceptance of AI for knowledge creation

Hu Baoliang, Wang Jiawen, Yan Shuai

(School of Management, Hangzhou Dianzi University, Hangzhou 310018, Zhejiang, China)

Abstract: The black-box problem of artificial intelligence (AI) is troubling users to accept AI for knowledge creation. The explainable AI is one of the important solutions to solve the problem. However, existing literature has rarely explored how the explainability of AI affects users′ acceptance of AI for knowledge creation. Therefore, this study focused on exploring the question, including the path mechanism of explainability affecting users′ acceptance of AI for knowledge creation, and the moderating effect of user characteristics on the path. This paper proposed some theoretical hypotheses and conducted the structural equation modeling and hierarchical regression analysis on 425 questionnaire data to test the hypotheses. The results showed that the three dimensions of explainability, i.e., completeness, format, and currency, have an influence on users′ acceptance of AI for knowledge creation; the influence of explainability on users′ acceptance of AI for knowledge creation is indirect, with perceived usefulness and perceived ease of use playing a mediating role. The results also showed that the influence of explainability on users′ acceptance of AI for knowledge creation is moderated by user characteristics such as education level, usage experience, and position. This study will not only contribute to the theories of AI knowledge creation and AI explainability theory by providing a user acceptance model based on the explainability, but also provide insights for enterprises to correctly play the role of AI explainability and promote AI knowledge creation.

Keywords: artificial intelligence; explainability; knowledge creation; user acceptance