2025 年 第 25 卷 第 11 期
2025, 25(11):04817-10

科 学 技 术 与 工 程
Science Technology and Engineering

ISSN 1671—1815
CN 11—4688/T

DOI:10.12404/j.issn.1671-1815.2404121

引用格式:陈淑, 庄越, 钱杨杨. 基于 MIMIC 模型的生成式人工智能公众风险感知[J]. 科学技术与工程, 2025, 25(11): 4817-4826.

Chen Shu, Zhuang Yue, Qian Yangyang. Public risk perception of generative artificial intelligence based on the MIMIC model[J]. Science Technology and Engineering, 2025, 25(11): 4817-4826.

# 基于 MIMIC 模型的生成式人工智能公众风险感知

陈淑, 庄越*, 钱杨杨

(武汉理工大学安全科学与应急管理学院, 武汉 430070)

**摘 要** 新一代人工智能技术的爆发式火热,将对风险社会中感知主体的风险体验产生深刻影响。运用因子分析和多指标多因素(multiple indicators and multiple causes, MIMIC)模型对生成式人工智能的 12 个风险场景进行研究,探究了反映公众风险感知程度的 4 个指标和影响公众风险感知的 5 个维度。结果显示,公众对生成式人工智能的风险感知程度可以由安全、技术、使用者和企业监管期望来反映;公众的风险感知受到其对技术、宏观、权益、主体和应用风险的主观评价的影响,其中,权益风险和宏观风险的影响最为显著。结果表明,公众对生成式人工智能具有"以我为主"和"未雨绸缪"的风险感知特点。在此基础上,进一步从历史与文化视角、风险沟通视角和科技治理视角对公众的生成式人工智能风险感知进行分析,并提出了相应的对策建议。

**关键词** 多指标多因素模型(MIMIC model); 生成式人工智能; 公众; 风险感知

**中图法分类号** X915.2; **文献标志码** A

# Public Risk Perception of Generative Artificial Intelligence Based on the MIMIC Model

CHEN Shu, ZHUANG Yue*, QIAN Yang-yang

(School of Safety Science and Emergency Management, Wuhan University of Technology, Wuhan 430070, China)

[**Abstract**] The explosive popularity of the new generation of artificial intelligence technologies will profoundly impact the risk experience of perceptual subjects within risk societies. Factor analysis and multiple indicators and multiple causes (MIMIC) model were used to study 12 risk scenarios of generative AI, besides four indicators reflecting the public's risk perception and five dimensions affecting the public's risk perception were explored. The results show that the public's perception of the risks of generative AI can be reflected by expectations of safety, technology, user and corporate regulatory. The public's risk perception is affected by its subjective evaluation of technology risks, macro risks, equity risks, subject risks and application risks, among which both equity risks and macro risks have the most significant impact. It shows that the public's risk perception of generative artificial intelligence is mainly characterized by "self-oriented" and "precautionary". On this basis, the public's risk perception of generative artificial intelligence from the perspectives of history and culture, risk communication and technology governance was analyzed further, and corresponding countermeasures was put forward.

[**Keywords**] multiple indicators and multiple causes (MIMIC) model; generative artificial intelligence; public; risk perception

从分析式人工智能(analytical artificial intelligence, AAI)到生成式人工智能(generative artificial intelligence,GAI)的跨越,互联网的内容生产模式在专业生成内容(professional generated content,PGC)与用户生产内容(user generated content,UGC)的基础上增加了全新的人工智能生产内容(artificial intelligence generated content,AIGC)模式,掀起了内容创作和信息检索的思维革命,具有划时代的意义。

目前正处于生成式人工智能技术蓬勃发展的阶段,同时人类社会也进入了风险社会[1],对新一代人工智能安全风险的研究也应聚焦在感知主体的风险体验上[2]。高风险感知容易催生社会不稳定风险,造成政治合法化危机[3],因此,有必要对生成式人工智能风险感知进行研究。此前,政府和技术创新企业能够并且已经根据其对人工智能安全的风险感知提出了风险应对措施,如各国政府出台一系列人工智能政策,以相互借鉴和学习,共同应对人工智能发展过程中存在的安全风险[4];Open AI 公司历时多年不断完

**收稿日期**: 2024-06-03 **修订日期**: 2024-09-25

**基金项目**: 国家社会科学基金(20BGL252)

**第一作者**: 陈淑(1996—),女,汉族,贵州遵义人,硕士研究生。研究方向:应急管理。E-mail:cs270804@163.com。

* **通信作者**: 庄越(1965—),男,汉族,湖北江陵人,博士,教授。研究方向:公共安全与应急管理。E-mail:zhuangyue@whut.edu.cn。

善 GPT(generative pre-trained transformer)中存在的技术缺陷和非技术问题,以预防新一代人工智能技术所带来的安全风险。公众作为又一风险感知主体,其风险感知会影响到智能技术的可持续发展,同时也会影响人工智能的科学传播和风险沟通[5],有必要对其风险感知加以关注。

本研究旨在通过社会调查,了解不同类型的公众对生成式人工智能在公共服务、社会生活中造成不利影响与伤害的风险认识情况,探究公众对生成式人工智能风险感知的维度及特点,并提出相应的风险防范对策,为生成式人工智能安全风险的治理提供重要的参考依据。

## 1 文献回顾

### 1.1 生成式人工智能安全风险研究

技术变革让人们在享受技术红利的同时,也对生成式人工智能技术存在的风险采取更加审慎的态度,继而催生了大量对生成式人工智能安全风险的研究。目前,学者们对生成式人工智能安全风险的研究聚焦于生成式人工智能的内在风险、主体风险和外在风险三大类。

内在风险即与技术本身相关的风险,这种风险源于技术内部因素,包括信息被泄露风险、技术失误风险、生成虚假错误信息风险和数据库污染的风险。首先,生成式人工智能作为人工智能大模型,需要大量数据进行支撑,数据来源合理合法对于输出结果的可靠性及结果的后续使用尤为重要,但目前生成式人工智能的数据来源存在着违法违规获取用户权限或隐私数据并造成信息泄露的风险[6];其次,生成式人工智能的信息处理过程是个“黑匣子”,无法得知数据处理过程中运用了何种方法和逻辑,存在着技术不可靠的安全风险[7];再次,生成式人工智能生成信息的质量影响着人们对生成式人工智能的态度,目前生成式人工智能生成的内容包含事实性虚假和幻觉性虚假信息,存在着生成虚假错误信息风险[8];最后,生成式人工智能随着社会进步和用户需求增加不断更新数据库,之前用生成式人工智能生成的存在误差的内容也会被纳入更新后的数据库,因此,数据库如何规避知识污染也是需考虑的风险[9]。

主体风险指生成式人工智能技术对其使用主体产生的诸多风险,如技术依赖、技术偏见、“信息茧房”和“机器换人”等风险。第一,人作为技术使用的主体,生成式人工智能的强大性能易让人产生信任感和依赖,一旦产生技术依赖,生成式人工智能输出不可信的结果时会产生更大的危害[10]。第二,生成式人工智能技术受到开发人员个体认知和习惯偏好的影响,训练数据选择等可能存在一定的价值偏见,经过多轮交替演进,易派生出偏见放大的问题,引发技术偏见或算法歧视风险[11]。第三,推荐算法的普遍使用带来了难以规避的信息茧房问题,生成式人工智能生成的内容会受到信息茧房的负面影响,也会引发技术使用主体信息茧房风险[12]。第四,随着人工智能的深入发展,不同于传统人工智能对就业的影响,生成式人工智能不再对低技能、低教育水平劳动力的就业产生威胁,而是凭借其较高的创造能力和较好的社交能力,冲击高技能、高教育水平且拥有高薪酬人员的就业机会[13],或是通过自身强大的学习能力替代中端脑力工作者的工作,导致中等技能工作岗位的减少[14],引发“机器换人”风险。

外在风险指生成式人工智能技术的广泛使用可能引发的宏观、长期性风险,包括经济风险、政治风险、伦理风险和法律风险等。生成式人工智能带来的经济风险主要源自技术垄断。生成式人工智能发展过程中不存在法律和技术壁垒,核心技术也完全公开,真正产生竞争壁垒的是训练成本,普通企业无力承担模型训练的高昂费用[15],因此,生成式人工智能技术将再一次巩固科技巨头的垄断地位,引发社会经济风险。生成式人工智能可能引发的政治风险源于技术生成过程的特点。生成式人工智能属于算法和数据驱动的深度合成技术,其生成的内容存在引发暴力冲突、意识形态渗透的可能,对国家安全造成威胁[16]。生成式人工智能带来的伦理风险重点在于人的主体性与生成式人工智能拟人化之间的冲突。生成式人工智能技术臻于完善,生成的内容越来越符合“人”的特点,若不承认其主体地位,将会造成主体缺位问题,若承认其主体地位,人类的主体价值将会被削弱,因此引发了人们对伦理风险的担忧[17]。生成式人工智能作为新兴技术,已有的法律规范还未能涵盖技术应用的各种场景。生成式人工智能投入规模化应用后,可能会引发创造性成果归属、刑事犯罪规制、数据滥用等多领域法律风险[18]。

### 1.2 风险感知理论

随着生成式人工智能技术的发展,科技领域掀起了新一轮的人工智能技术安全风险讨论浪潮,其中便涉及对风险感知角度的关注。风险感知是人们对某个特定风险的特征和严重性所做出的主观判断,是测量公众心理恐慌的重要指标[19]。不同主体的风险感知不同,对风险严重性的感知会影响主体的行为[20]。因此,有必要对公众的风险感知进行

分析,以改善不同主体之间的风险信息交流,帮助进行风险分析和政策制定[21]。

风险感知的研究方法不断延伸发展,已经相对成熟,形成了风险的心理测量理论、文化理论和社会放大理论三大流派。风险的心理测量理论由Slovic[21]提出,该理论认为风险是主观的风险,侧重于运用心理学方法对风险根源的主观特征和主观感受进行测量。文化理论由 Douglas 等[22]提出,是从社会学视角进行风险感知研究,认为风险是个体认知的结果,人类感知和适应风险的过程很大程度上是由社会背景和文化信仰决定的。社会放大理论侧重于研究风险社会中各个社会主体的参与和各类社会行为的联动反应,及其产生的涟漪效应[23]。

在新一代人工智能技术蓬勃发展背景下,学界对风险感知的研究大都结合风险的心理测量理论,从生成式人工智能的上位概念,即人工智能的角度出发进行探究,或聚焦于公众的人工智能伦理风险感知[24],抑或是关注青年人对人工智能的风险感知[25],等等。此外,在生成式人工智能安全风险感知研究方面,还有学者运用扎根理论方法,以生成式人工智能产品之一——Chat GPT 的用户为研究对象进行风险感知维度划分[26]。可见,现有研究集中于风险的心理测量范式,限定了风险类别及研究对象,或选取某一类客观风险进行分析,或探究特定群体的风险感知。总的来说,对生成式人工智能安全风险感知的研究仍是一个新领域。

公众是风险感知的重要主体,且生成式人工智能作为人工智能的一种新形态,在公众风险感知类别、风险感知维度和风险感知特点等方面与传统的人工智能不尽相同,因此,有必要对公众的生成式人工智能安全风险感知进行研究。目前,学者们对生成式人工智能安全风险感知的探析以单一的风险测量理论为主,风险感知的文化理论和社会放大理论方面的研究较为不足。从风险感知多理论结合的角度探索公众对生成式人工智能安全风险感知的维度及其影响,增强了理论解释度,细化了公众对新技术的风险感知研究,对公众正确感知和应对风险、技术创新企业弥补技术缺陷和公共部门制定有效风险应对措施具有重要意义。

## 2　研究设计

### 2.1　模型设计

#### 2.1.1　因子分析模型

本研究拟使用公众对生成式人工智能的监管期望来反映公众的风险感知,风险感知作为潜变量,监管期望作为显变量,共同构成因子分析模型。根据风险感知的文化理论,风险是个人认知的结果,风险感知受到个体期望水平的影响[27]。此外,对政府监管有效性的感知也是影响风险感知的主要因素[28],公众感知到生成式人工智能带来的风险,会期望政府对其进行有效监管,以降低自身风险感知程度。因此,可运用因子分析模型从监管期望角度来反映公众对生成式人工智能风险感知的程度。因子分析模型如图 1 所示,生成式人工智能风险感知为潜变量,监管期望为显变量。


图 1　GAI 风险感知因子分析模型

Fig. 1　GAI risk perception factor analysis model

使用因子分析模型来探究公众的生成式人工智能风险感知,首先需进行探索性因子分析,以考察选定的多个反映指标是否存在唯一一个公因子,既保证多项指标能够较为全面地反映风险感知,又不会产生多余变量。其次应进行验证性因子分析,检验整个因子分析模型的拟合度和多个反映指标因子载荷系数的显著性,进而验证风险感知因子分析模型的构建是否合理。

#### 2.1.2　MIMIC 模型

多指标多因素(multiple indicators and multiple causes, MIMIC)模型作为结构方程模型的一种特殊形式,与标准的结构方程模型既存在紧密联系,也有明显区别。它的解释变量为显变量,而被解释变量为潜变量。

本文在参考之前学者研究的基础上,运用 AMOS 28.0软件和 MIMIC 模型,初步绘制了生成式人工智能风险感知 MIMIC 模型图,具体如图 2 所示。该模型包括结构模型和测量模型两部分,左边的风险维度和中间的风险感知共同构成结构模型,由显变量风险维度来解释潜变量风险感知,中间的风险感知与右边的监管期望构成测量模型,由监管期望这一反映型指标变量来反映被解释变量风险感知。使用 MIMIC 模型来对公众的生成式人工智能风险感知进行测度,能够同时清晰地展现风险感知的客观影响因素及主观的反映型指标。

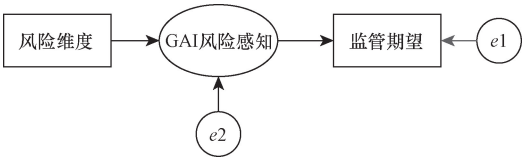



图 2　GAI 风险感知 MIMIC 模型

Fig. 2　GAI risk perception MIMIC model

## 2.2 量表设计

### 2.2.1 监管期望量表

风险感知潜变量作为公众对风险的认知和感受,由多种因素构成,因此,风险感知的度量也应该是多角度的。选用多个反映型指标来构造生成式人工智能风险感知潜变量,能够从多个方面对风险感知进行分析,降低传统单一指标带来的测量误差,有助于揭示风险感知的复杂性,为风险管理、沟通和决策提供更全面、深入的理论支持和实践指导,提高风险应对的有效性,降低潜在的负面影响。

度量风险感知可以使用多个反映型指标,这些指标能够从不同角度反映被研究的潜变量。一方面,根据风险感知的文化理论,公众的风险感知受个体期望水平的影响[27]。风险感知显露出公众的心理恐慌程度,所以公众的风险感知天然地包含了对安全的期望。另一方面,对政府监管有效性的感知和信任程度也是影响风险感知的主要因素[28-29],风险感知与公众的信任程度之间存在着相关关系[30]。国家的支持政策能够促进相关领域研究的发展,而政府的严格监管则会使公众采取审慎的态度[31]。从监管的对象来看,政府对生成式人工智能的安全风险监管包括对技术本身、技术使用者和技术创新企业进行监管。结合期望和监管这两方面反映指标,可以采用公众对安全的监管期望、对技术的监管期望、对使用者的监管期望及对技术创新企业的监管期望4个具体的反映指标来对公众的生成式人工智能风险感知进行反映。

为反映公众对生成式人工智能的风险感知程度,需对公众的生成式人工智能监管期望进行调查。选取的安全监管期望、技术监管期望、使用者监管期望和企业监管期望这四个反映型指标变量可以采用5分李克特量表(赋值为:1“完全不符合”,2“比较不符合”,3“一般符合”,4“比较符合”,5“完全符合”)进行测度,打分越高,说明公众对生成式人工智能的监管期望就越高,变量及其释义如表1所示。

### 2.2.2 风险类别量表

生成式人工智能的发展易带来多种风险,为了便于对学者们的风险研究情况进行汇总整理,前文按照其共通点将其概括为了内在风险、主体风险和外在风险。具体而言,前文指出目前学者们对生成式人工智能的研究主要涉及三大类共12种安全风险,分别为虚假错误信息风险、信息泄露风险、技术失误风险、数据库污染风险、机器换人风险、技术依赖风险、信息茧房风险、技术偏见风险、经济风险、政治风险、伦理风险和法律风险。

**表1 量表设计**

**Table 1 Scale design**

| 量表 | 变量 | 变量释义 |
|---|---|---|
| 风险类别量表 | 虚假错误信息风险 | 生成虚假或错误信息 |
| | 信息泄露风险 | 泄露个人数据信息 |
| | 技术失误风险 | 数据处理出现技术不可靠情况 |
| | 数据库污染风险 | 存在误差的内容被纳入更新后的数据库 |
| | 机器换人风险 | 工作被挤压或替代 |
| | 技术依赖风险 | 过度依赖技术,降低创新和思考意愿 |
| | 信息茧房风险 | 生成的内容具有同一性 |
| | 技术偏见风险 | 生成带有偏见或歧视性的内容 |
| | 经济风险 | 科技巨头的垄断地位加大贫富差距 |
| | 政治风险 | 对国家安全造成威胁 |
| | 伦理风险 | 技术冲击人的主体地位 |
| | 法律风险 | 技术应用中的法律规范不够完善 |
| 监管期望量表 | 安全监管期望 | 支持政府对生成式人工智能安全采取严格管制措施 |
| | 技术监管期望 | 支持政府对生成式人工智能技术采取严格管制措施 |
| | 使用者监管期望 | 支持政府对生成式人工智能使用者采取严格管制措施 |
| | 企业监管期望 | 支持政府对生成式人工智能企业采取严格管制措施 |

本文调查问卷设计中,通过文献分析法,将学者们研究的12种风险设置为对应的12道风险感知题目,尽量全面地对生成式人工智能安全风险发生的场景做出假设,以测度公众对生成式人工智能安全风险的感知维度及其特点。公众对各种风险的评价采用5分李克特量表(赋值为:1“完全不担忧”,2“比较不担忧”,3“一般担忧”,4“比较担忧”,5“完全担忧”)进行测度,打分越高,说明公众对相应风险场景的主观风险评价越严重。为避免公众对量表中各种风险的认知存在偏差,从而对调查结果的有效性产生影响,本研究统一对量表中的风险含义进行解释(表1)。随后,将对公众的主观风险评价进行因子分析,运用主成分分析方法提取公因子,提取的公因子即为风险维度。风险维度作为解释变量,主要探究其对公众的生成式人工智能风险感知的影响程度。

## 2.3 调查过程

为调查公众对生成式人工智能的安全风险感知情况,本研究采用线上社交媒体发放问卷和线下社区发放问卷的方式进行问卷调查获取数据信息。作为信息收集的手段之一,问卷调查在有效性和覆盖范围方面都具有显著的优势。一方面,线下问卷调查能够和受访者面对面沟通,减少被调查者对于问卷中各题项的认知误差,提高问卷回收的有效性;另一方面,线上问卷不受地区和时间的影响,覆盖人群更广,容易收集到各年龄段、受教育程度、职业类型和所在地区的样本。线上和线下两种问卷

调查方式相结合,更加契合研究所需。

本次调查共回收了 497 份问卷,根据问卷中设置的有效性检验题目“为了保证问卷的有效性,本题请选择一般担忧”,人工剔除了未按要求填写的无效问卷,得到了有效问卷共 436 份,问卷合格率达 87.7%。

本研究的控制变量主要包括性别、年龄、所在地区(常驻)、受教育程度和当前所从事职业类型,样本的描述性特征如表 2 所示。

样本特征描述性统计分析可以帮助更好地对本次调查样本的基本情况进行把握。由样本特征(表 2 )所示,样本的性别比较为均衡,年龄、职业、受教育程度和地区覆盖较为全面,因此样本符合本研究目的。

**表 2　样本特征描述性统计**

**Table 2　Descriptive statistics of sample characteristics**

| 项目 | 分类 | 频率 | 占比/% |
|---|---|---|---|
| 性别 | 男 | 217 | 49.8 |
| | 女 | 219 | 50.2 |
| 年龄 | 0~14 岁 | 1 | 0.2 |
| | 15~35 岁 | 340 | 78.0 |
| | 36~64 岁 | 95 | 21.8 |
| 所在地区 | 城市 | 381 | 87.4 |
| | 农村 | 55 | 12.6 |
| 受教育程度 | 高中及以下 | 21 | 4.8 |
| | 专科 | 53 | 12.2 |
| | 本科 | 287 | 65.8 |
| | 研究生及以上 | 75 | 17.2 |
| 职业类型 | 学生 | 107 | 24.5 |
| | 政府/企事业单位 | 58 | 13.3 |
| | 民营企业 | 232 | 53.2 |
| | 自由职业 | 28 | 6.4 |
| | 其他 | 11 | 2.5 |

# 3　实证结果分析

## 3.1　因子分析结果

运用 SPSS 28.0 软件对样本进行信度和效度检验,问卷整体的 Cronbach $\alpha$ 系数为 0.853,说明整体信度较好,问卷各变量内部一致性较高,可以作为数据分析的依据开展进一步研究。量表的 KMO 值为 0.916,Bartlett's 球形检验值为 2 486.989,在自由度为 120 的条件下以及 0.001 水平上显著,说明量表中的变量之间存在相关性,数据很适合进行因子分析。

### 3.1.1　风险维度的划分

运用因子分析方法对风险维度进行划分,能够提取公众对生成式人工智能风险主观评价的主要维度,便于对风险及公众感知特点进行全面分析。研究尝试了多个公因子的模型对比,分析发现五因子模型的总方差贡献率达到了 71.471%,且每一个公因子对应的因子数量大于 1,说明总体 70% 以上的信息可以由这 5 个公共因子来解释,取前 5 个公因子进行分析比较合理。

采用凯撒正态化最大方差法,各个题项经正交旋转后,得到旋转成分矩阵(表 3),通过主成分分析法提取 5 个公因子,即生成式人工智能安全风险的五个维度(旋转在 9 次迭代后已收敛)。第一个风险维度包括生成式人工智能技术发展过程中可能出现的生成虚假错误信息风险(T1)、技术失误风险(T2)和数据库被污染风险(T3),按其共通点可将该维度命名为“技术风险(T)”;第二个风险维度包括宏观的经济风险(M1)、政治风险(M2)和法律风险(M3),均为较为深远的长期性风险,可将该维度命

**表 3　旋转成分矩阵**

**Table 3　Rotated component matrix**

| 风险 | 成分 | | | | |
|---|---|---|---|---|---|
| | 技术风险(T) | 宏观风险(M) | 权益风险(R) | 主体风险(S) | 应用风险(U) |
| 虚假错误信息风险(T1) | 0.696 | 0.185 | 0.397 | 0.065 | 0.079 |
| 技术失误风险(T2) | 0.739 | 0.239 | -0.013 | 0.180 | 0.220 |
| 数据库污染风险(T3) | 0.757 | 0.195 | 0.167 | 0.154 | 0.127 |
| 经济风险(M1) | 0.221 | 0.636 | 0.416 | 0.093 | 0.108 |
| 政治风险(M2) | 0.305 | 0.619 | 0.244 | 0.322 | 0.049 |
| 法律风险(M3) | 0.220 | 0.755 | -0.014 | 0.122 | 0.332 |
| 信息泄露风险(R1) | 0.448 | 0.289 | 0.631 | 0.028 | 0.114 |
| 机器换人风险(R2) | 0.108 | 0.076 | 0.744 | 0.329 | 0.301 |
| 技术偏见风险(S1) | 0.409 | 0.055 | 0.085 | 0.594 | 0.441 |
| 伦理风险(S2) | 0.123 | 0.267 | 0.232 | 0.837 | 0.042 |
| 技术依赖风险(U1) | 0.078 | 0.276 | 0.485 | 0.248 | 0.540 |
| 信息茧房风险(U2) | 0.241 | 0.217 | 0.225 | 0.063 | 0.793 |

名为“宏观风险(M)”;第三个风险维度包括与个人权益息息相关的信息泄露风险(R1)和机器换人风险(R2),可将该维度命名为“权益风险(R)”;第四个风险维度包括挑战个人及全人类主体性的技术偏见风险(S1)和伦理风险(S2),可将该维度命名为“主体风险(S)”;第五个风险维度包括技术应用过程中对个人带来的技术依赖风险(U1)和信息茧房风险(U2),可将该维度命名为“应用风险(U)”。

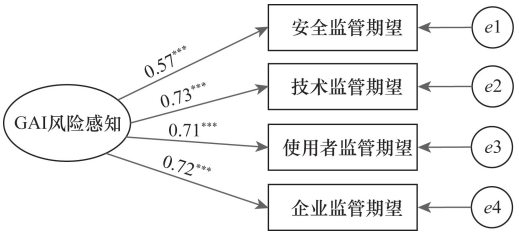
3.1.2 风险感知的测度

为测度公众对生成式人工智能的风险感知,首先采用探索性因子分析方法检验四个反映型指标是否在测量唯一一个生成式人工智能风险感知公因子。提取因子的方法为主成分因子法,同时采用极大似然因子法和主因子法做进一步佐证。

经分析,3 种方法均只提取了一个特征值大于 1 的公因子。就主成分因子法而言,提取的公因子特征值为 2.398,其累计方差贡献率为 59.954。同时,该公因子对四个指标的载荷系数分别为 0.702、0.801、0.794 和 0.796,均大于 0.75,具有高度相关性。这说明 4 个指标变量确实只能测度出唯一的公因子,初步表明选取的四个反映型指标能够很好地反映公众对生成式人工智能的风险感知。

在探索性因子分析的基础上,运用极大似然法继续进行验证性因子分析,以进一步检验模型的拟合度和指标因子载荷是否显著。

根据验证性因子分析结果(图 3)所示,4 个反映型指标的标准化因子载荷系数均大于 0.5,且在 0.1% 的水平上显著。在模型的拟合指标方面,模型总体拟合优度指标 CMIN/DF 的值为 0.073,SRMR 为 0.003 1,AGFI 为 0.999,均说明模型拟合较好。结合探索性因子分析的结果,认为测量模型成立。此外,该测量模型的拟合度较高,适合进行后续的模型估计。




*、** 和 *** 分别表示在 0.05、0.01 和 0.001 的统计水平上显著

图 3 GAI 风险感知的验证性因子分析

Fig. 3 Validated factor analysis of GAI risk perception

## 3.2 MIMIC 模型估计结果

运用 Amos 28.0 软件进行检验,获得生成式人工智能风险感知 MIMIC 模型整体拟合情况及各个测量变量之间的标准化路径系数,检验结果如表 4 和图 4 表示。由表 4 可知,MIMIC 模型拟合状况较好(CMIN/DF = 1.030, RMSEA = 0.008, SRMR = 0.027 1,CFI = 0.998,GFI = 0.986,TLI = 0.998),即构建的模型在统计上是成立的。

通过模型运行之后所展示出来的标准化路径系数,能更好地把握变量之间的关系。根据模型估计结果(图 4)所示,技术风险、主体风险和应用风险对应的路径系数分别为 0.10、0.11 和 0.16,结果不够显著。宏观风险和权益风险与风险感知之间的标准化路径系数分别为 0.31 和 0.26,且均在 0.1% 水平显著,即公众对宏观风险和权益风险的严重性判断分别增加 1 分,风险感知分别提高 0.31 和 0.26。综上,说明公众对宏观风险和权益风险的主观性风险评价越高,对生成式人工智能的风险感知程度也就越高。

**表 4 GAI 风险感知模型整体拟合度评价**

**Table 4 Evaluation of the overall fit of the GAI risk perception model**

| 评价指标 | 指标含义 | 评价标准 | 实际拟合值 | 拟合结果 |
|---|---|---|---|---|
| CMIN/DF | 卡方与自由度比值 | <3 | 1.030 | 符合标准 |
| RMSEA | 理论模型与饱和模型不拟合指数 | <0.05 | 0.008 | 符合标准 |
| SRMR | 标准化均方根残差 | <0.08 | 0.027 1 | 符合标准 |
| CFI | 在 NFI 基础上,考虑被检验模型与理论模型的中央卡方分布离散度 | >0.90 | 0.998 | 符合标准 |
| GFI | 模型拟合方差和协方差对样本方差和协方差解释度 | >0.90 | 0.986 | 符合标准 |
| TLI | Tucker-Lewis 指数 | >0.90 | 0.998 | 符合标准 |

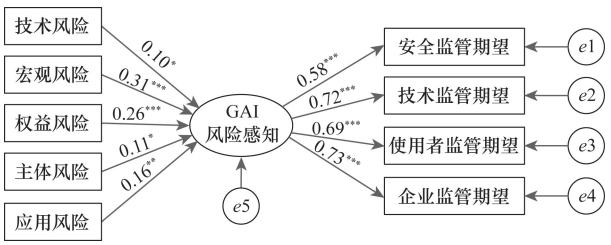


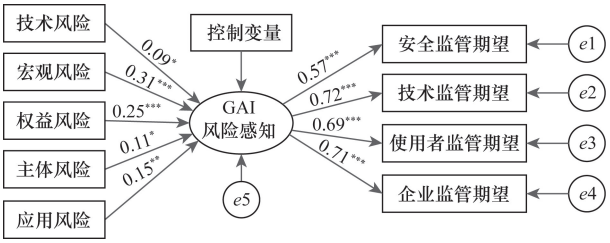

图 4 GAI 风险感知 MIMIC 模型估计结果

Fig. 4 Estimation results of the GAI risk perception MIMIC model

## 3.3 稳健性检验

为验证 MIMIC 模型估计结果的稳健性,选取了在原模型的基础上增加性别、年龄、所在地区、受教育程度和职业类型等控制变量的检验方法。增加了控制变量的扩展的 MIMIC 模型同样运用极大似然法进行分析,扩展模型整体拟合状况较好(CMIN/DF = 2.802,GFI = 0.925,SRMR = 0.071 7),检验结

果如图 5 所示。相较于原模型,扩展模型中的解释变量与被解释变量之间的路径系数有所变化,但总体变化幅度较小,系数显著性和作用方向未发生变化,这证明了模型估计结果的稳健性。




*、** 和 *** 分别表示在 0.05、0.01 和 0.001 的统计水平上显著

图 5　扩展的 MIMIC 模型估计结果

Fig. 5　Extended MIMIC model estimation results

### 3.4　结果分析

根据模型运行结果可知,公众对生成式人工智能的安全风险主观评价分为五个维度,通过对模型标准化路径系数的比较,能够发现公众的宏观风险和权益风险主观评价对其风险感知的影响显著。

公众对宏观风险的主观性风险评价对其生成式人工智能风险感知程度的影响最为显著,包括科技巨头技术垄断引发的社会经济公平风险,深度合成技术生成的内容对国家造成的安全威胁,以及生成式人工智能投入规模化应用后引发的多领域法律风险。随着社会数字鸿沟不断弥合,公众拥有了更多接触新技术的机会,对新技术风险的认识也较为长远和深刻,对宏观风险的关注度也就比较高。

公众对权益风险的主观性风险评价对其生成式人工智能风险感知程度也具有显著影响,风险内容包括公众隐私、数据信息被违法违规获取和使用的风险,以及冲击劳动力就业机会或替代某些工作的风险。得益于信息传播的便捷和整体教育水平的提高,公众对个人权益的关注日益增长,对个人权益保护提出了更高的要求。由于权益风险直接涉及公众的生活、隐私和职业安全等多个方面,因此人们对这类风险非常敏感,对权益风险严重性的感知程度也相对较高。

此外,对技术风险、主体风险和应用风险严重程度的评价并未显著强化公众对生成式人工智能的风险感知。技术风险包括因为生成式人工智能技术发展可能引发的生成虚假错误信息风险、技术失误风险和数据库污染风险。生成式人工智能作为一项新兴技术,正处于不断完善阶段,不可避免存在技术性问题。但公众目前对其工作原理和技术细节的了解有限,导致公众可能无法充分意识到潜在的技术风险。

主体风险涉及公众主体性与生成式人工智能拟人化之间的冲突风险,以及训练数据选择等因素导致的公众在生成式人工智能预测和决策时遭遇不公平结果的风险。当生成式人工智能存在主体风险,公众可能会对其产生不信任感,从而影响技术的接受度和公众的风险感知程度。但在科技产业的推动下,公众往往对新技术抱有乐观的态度,认为技术进步将为人类带来巨大的便利和福祉,这种乐观主义可能导致公众对潜在主体风险的忽视。

应用风险是公众在技术应用中因生成式人工智能的强大性能而产生的过度依赖风险,以及推荐算法应用于生成式人工智能中所导致的信息茧房风险。生成式人工智能聊天机器人 Chat GPT 发布仅两个月时间,活跃用户就已突破了 1 亿,成为史上增长最快的消费者应用,生成式人工智能的广泛应用也对公众的风险感知带来了一定的影响。但随着人工智能技术的普及,公众可能逐渐适应了这些风险,而不再把技术依赖和信息茧房等情况视为风险问题,从而降低了对应用风险的关注度。

综上可知,公众作为生成式人工智能风险感知主体,更加关注与自身密切相关的权益风险及较为长远和全面的宏观风险。相对抽象,不易被直观感知,或影响范围较小,尚未引起广泛关注的技术风险、主体风险和应用风险的主观评价对公众风险感知的影响不太显著,这意味着公众没有充分意识到生成式人工智能潜在的这些风险。总的来说,公众在应对生成式人工智能带来的风险时,具有“以我为主”和“未雨绸缪”的风险感知特点。

## 4　建议

### 4.1　融合历史文化因素,深化风险感知理解

公众的风险感知受到历史和文化的影响,不同历史时期的公众对风险的感知受到当时政策法规、经济发展、社会结构和科技水平等因素的影响,不同文化背景下公众的风险感知也各有特点。

最近欧洲议会通过了《人工智能法案》,其要点是把人的权利和隐私保护放在了最优先的位置,而我国新出台的《生成式人工智能服务管理暂行办法》则在之前发布的《生成式人工智能服务管理办法(征求意见稿)》的基础上还放宽了监管要求,增加了鼓励生成式人工智能技术发展的措施。两者相较可知,在促进生成式人工智能健康发展的过程中,西方国家更加侧重于公众的权益保护,而中国更加偏向于鼓励技术发展。通过中国是否会出台西方国家类似法案的思考,可知公众对生成式人工智能的风险感知会存在历史与文化的差异。

就当下而言,公众对生成式人工智能的权益风险

和宏观风险的主观性评价对其风险感知的影响较为显著。究其直接原因,权益风险与公众的生活息息相关,而宏观风险的影响较为广泛,这两种风险更容易吸引公众的注意力。透过历史与文化的视角来探究其深层次原因,能够发现中国的法治建设增强了公众的个人权益保护意识,但现行的生成式人工智能相关政策法规对公众的权益保护不足,导致了公众对权益风险的高关注度;“人无远虑,必有近忧”的传统思想则使公众用更加长远的眼光来考虑事情,自然而然对宏观风险呈现出较大程度的担忧。

相较之下,由于当前的技术发展水平及公众的风险认知水平受到历史和文化条件的限制,技术风险、主体风险和应用风险呈现出了专业性和隐蔽性特点,公众对这些风险知之甚少,难以察觉风险的存在,使得公众对生成式人工智能的技术风险、主体风险和应用风险的主观评价对其风险感知的影响不够显著。

### 4.2 优化风险沟通策略,构建风险共治机制

公众的风险感知并非自发产生的,从传播形式上日益受到网络媒体的影响,因此其风险感知往往受到诱导,继而产生偏差。对于生成式人工智能这种新技术,普通公众既无知识背景,又无实际体验,因此公众的风险感知很容易受到利益集团操纵,从而出现风险的放大效应,强化公众的不安全感,影响公众的接纳意愿和新技术的发展水平,亦或是出现忽视风险的现象,导致风险事件的发生对准备不足的公众带来更大的伤害。

要使公众降低对生成式人工智能感知风险与实际风险之间的偏差,形成科学的风险感知,就需进行有效的风险沟通。技术创新企业应该主动承担社会责任,做负责任的科普传播,让公众正确认知风险;政府需对网络虚假信息进行约束打击与规范,给公众营造良好的技术传播与技术接纳环境;技术创新企业与政府之间也有必要加强风险沟通,共同应对技术发展难题,防范潜在安全风险。

### 4.3 营造良好市场环境,完善风险治理格局

数据治理、算法治理的政策法规与市场执行在深刻地影响着公众的风险感知。公众对 Open AI 公司等技术创新企业所研发的生成式人工智能抱有科技助力生产生活的期待,对新技术带来的安全风险依然存在恐惧心理。如果技术创新型企业盗取和泄露公众隐私数据等现象屡禁不止,公众对新技术的恐惧心理则会影响到其对生成式人工智能的接纳意愿。这表现出了社会行为与公众风险感知的互为关联性,以及公众对生成式人工智能风险治理的迫切需求。

在生成式人工智能风险治理中,技术创新企业需主动承担起社会责任,在进行创新研发时除了要对其技术性因素进行测评外,还需要对其可能存在的社会风险进行评估,在政府部门的引领下,共同营造良好的市场环境。政府可以通过敏捷治理模式制定更加适配和完善的人工智能监管政策,以规制技术创新企业对公众权益的损害行为,防止宏观风险带来不利影响,形成完善的风险治理格局。

## 5 结论

本文中通过因子分析和 MIMIC 模型方法探究了公众对生成式人工智能的风险感知维度及特点。研究结果显示,公众对生成式人工智能的风险感知分为技术风险、宏观风险、权益风险、技术风险和应用风险五个维度,其中,公众对“权益风险”维度和“宏观风险”维度的主观评价对其风险感知的影响最大,说明公众在应对生成式人工智能带来的风险时,具有“以我为主”和“未雨绸缪”的风险感知特点。此外,通过探索性因子分析和验证性因子分析,得出用公众的安全监管期望、技术监管期望、使用者监管期望和企业监管期望来反映其生成式人工智能风险感知是合理的。据此,从历史与文化视角、风险沟通视角和科技治理视角对公众的生成式人工智能风险感知进行了深层次分析,并提出了相应的对策建议,为公众如何更有效地适应风险时代提供参考建议,为生成式人工智能的风险治理提供重要的参考依据。

与此同时,本文中还存在若干局限和不足:第一,本研究仅对公众在单一非变化的时间节点的主观评价进行探讨,缺少公众对生成式人工智能风险感知的动态变化的研究。今后的研究方法中可以纵向探索,通过了解动态变化过程进一步深入探讨作用机理。第二,目前的调查还未能覆盖所有的人口,例如占很大比例的老龄化人口,其数字鸿沟和数字失能使其风险感知又会呈现不同的面貌,所以接下来的研究还需考虑生成式人工智能发展的公平性与普惠性问题。第三,本研究仅围绕公众的生成式人工智能风险感知维度及其特点进行分析,缺少对风险感知与其他变量之间关系的探索。在今后的研究当中,可以结合技术接受模型对风险感知与个人认知、新技术接纳意愿等变量之间的关系进行探究,以增加研究的深度和广度。

## 参 考 文 献

[1] Beck U. 风险社会:新的现代性之路[M]. 张文杰, 何博闻, 译. 南京:译林出版社, 2018:1-5.
Beck U. Risk society: towards a new modernity [M]. Zhang Wenjie, He Bowen, translated. Nanjing: Yilin Press, 2018: 1-5.

[2] 黄和平, 邴振华, 钟伟, 等. 社区民宿邻避效应: 风险感知、权益冲突与形成机制研究——以上海地区为例[J]. 旅游科学, 2023, 37(2): 100-116.
Huang Heping, Bing Zhenhua, Zhong Wei, et al. The NIMBY effect of community homestays: a study on risk perception, rights conflict, and formation mechanism—taking the Shanghai area as an example[J]. Tourism Science, 2023, 37(2): 100-116.

[3] 蒲晓红, 赵海堂. 互联网使用对公众风险感知的影响机制——基于政府回应视角[J]. 中国行政管理, 2021(5): 146-154.
Pu Xiaohong, Zhao Haitang. The impact mechanism of internet use on public risk perception: from the perspective of government response[J]. Chinese Public Administration, 2021(5): 146-154.

[4] 赵景欣, 岳星辉, 冯崇朋, 等. 基于通用数据保护条例的数据隐私安全综述[J]. 计算机研究与发展, 2022, 59(10): 2130-2163.
Zhao Jingxin, Yue Xinghui, Feng Chongpeng, et al. A review of data privacy security based on the general data protection regulation [J]. Journal of Computer Research and Development, 2022, 59(10): 2130-2163.

[5] 朱依娜, 邱紫薇. 人工智能公众态度: 概念及测量研究综述[J]. 科学与社会, 2023, 13(3): 127-138.
Zhu Yina, Qiu Ziwei. Public attitudes towards artificial intelligence: a review of concepts and measurement research[J]. Science and Society, 2023, 13(3): 127-138.

[6] 朱禹, 陈关泽, 陆泳溶, 等. 生成式人工智能治理行动框架: 基于 AIGC 事故报道文本的内容分析[J]. 图书情报知识, 2023, 40(4): 41-51.
Zhu Yu, Chen Guanzhe, Lu Yongrong, et al. A governance action framework for generative artificial intelligence: content analysis based on AIGC accident reporting texts[J]. Documentation, Information & Knowledge, 2023, 40(4): 41-51.

[7] 李艳燕, 郑娅峰. 生成式人工智能的教育应用[J]. 人民论坛, 2023(23): 69-72.
Li Yanyan, Zheng Yafei. The educational application of generative artificial intelligence[J]. People's Forum, 2023(23): 69-72.

[8] 莫祖英, 盘大清, 刘欢, 等. 信息质量视角下 AIGC 虚假信息问题及根源分析[J]. 图书情报知识, 2023, 40(4): 32-40.
Mo Zuying, Pan Daqing, Liu Huan, et al. Analysis of the problem and root causes of AIGC false information from the perspective of information quality[J]. Documentation, Information & Knowledge, 2023, 40(4): 32-40.

[9] 任宇, 罗剑萍, 甘甜. ChatGPT 对学术出版实践的影响及其应对[J]. 武汉理工大学学报(社会科学版), 2023, 36(4): 113-119.
Ren Yu, Luo Jianping, Gan Tian. The impact of ChatGPT on academic publishing practices and its response[J]. Journal of Wuhan University of Technology (Social Science Edition), 2023, 36(4): 113-119.

[10] 郁建兴, 刘宇轩, 吴超. 人工智能大模型的变革与治理[J]. 中国行政管理, 2023, 39(4): 6-13.
Yu Jianxing, Liu Yuxuan, Wu Chao. Transformation and governance of large AI models[J]. Chinese Public Administration, 2023, 39(4): 6-13.

[11] 谢梅, 王世龙. ChatGPT 出圈后人工智能生成内容的风险类型及其治理[J]. 新闻界, 2023(8): 51-60.
Xie Mei, Wang Shilong. Risk types and governance of AI-generated content after ChatGPT goes mainstream[J]. Journalism and Mass Communication, 2023(8): 51-60.

[12] 房娇娇, 高天书. 生成式人工智能辅助行政决策的算法隐患及其治理路径[J]. 湖湘论坛, 2024, 37(1): 99-111.
Fang Jiaojiao, Gao Tianshu. The algorithmic hidden dangers of generative artificial intelligence in assisting administrative decision-making and their governance path[J]. Huxiang Forum, 2024, 37(1): 99-111.

[13] 陈永伟. 超越 ChatGPT: 生成式 AI 的机遇、风险与挑战[J]. 山东大学学报(哲学社会科学版), 2023(3): 127-143.
Chen Yongwei. Beyond ChatGPT: opportunities, risk, and challenges from generative AI[J]. Journal of Shandong University (Philosophy and Social Sciences), 2023(3): 127-143.

[14] 郑世林, 姚守宇, 王春峰. ChatGPT 新一代人工智能技术发展的经济和社会影响[J]. 产业经济评论, 2023(3): 5-21.
Zheng Shilin, Yao Shouyu, Wang Chunfeng. The economic and social impact of the development of the new generation of AI technology ChatGPT[J]. Review of Industrial Economics, 2023(3): 5-21.

[15] 唐林垚. 具身伦理下 ChatGPT 的法律规制及中国路径[J]. 东方法学, 2023(3): 34-46.
Tang Linyao. Legal regulation of ChatGPT under embodied ethics and China's path[J]. Oriental Law, 2023(3): 34-46.

[16] 商建刚. 生成式人工智能风险治理元规则研究[J]. 东方法学, 2023(3): 4-17.
Shang Jiangan. Research on the meta-rules of generative artificial intelligence risk governance[J]. Oriental Law, 2023(3): 4-17.

[17] 冯子轩. 生成式人工智能应用的伦理立场与治理之道: 以 ChatGPT 为例[J]. 华东政法大学学报, 2024, 27(1): 61-71.
Feng Zixuan. The ethical stance and governance approach of generative artificial Intelligence application: a case study of ChatGPT [J]. East China University of Political Science and Law Journal, 2019, 27(1): 61-71.

[18] 袁曾. 生成式人工智能的责任能力研究[J]. 东方法学, 2023(3): 18-33.
Yuan Zeng. Research on the liability capacity of generative artificial intelligence[J]. Oriental Law, 2023(3): 18-33.

[19] 程方明, 邵杰, 苏畅, 等. 风险感知对非适应性应急疏散行为的影响[J]. 中国安全生产科学技术, 2023, 19(1): 176-182.
Cheng Fangming, Shao Jie, Su Chang, et al. Impact of risk perception on non-adaptive emergency evacuation behavior[J]. Journal of Safety Science and Technology, 2023, 19(1): 176-182.

[20] 李文琴, 刘荣敏, 孙林辉, 等. 跨层次视角下信息安全氛围对员工信息安全制度遵守意愿的影响作用[J]. 科学技术与工程, 2024, 24(8): 3479-3487.
Li Wenqin, Liu Rongmin, Sun Linhui, et al. The influence of information security climate on employees' willingness to comply with information security systems from a cross-level perspective[J]. Science Technology and Engineering, 2024, 24(8): 3479-3487.

[21] Slovic P. Perception of risk[J]. Science, 1987, 236(4799): 280-285.

[22] Douglas M, Wildavsky A. Risk and culture: an essay on the selection of technological and environmental dangers[M]. Berkeley: University of California Press, 1983: 1-15.

[23] 宋宪萍, 曹宇驰. 风险的社会放大框架: 逻辑进路与趋向研判[J]. 甘肃社会科学, 2022(5): 130-139.

Song Xianping, Cao Yuchi. Social magnification framework of risk: logical approach and trend research[J]. Gansu Social Sciences, 2022(5): 130-139.

[24] 宋艳, 陈琳, 李琴, 等. 人工智能伦理风险感知、信任与公众参与[J]. 科学学研究, 2022, 40(7): 1153-1162, 1171.
Song Yan, Chen Lin, Li Qin, et al. Public perception of AI ethics risks, trust, and participation[J]. Studies in Science of Science, 2022, 40(7): 1153-1162, 1171.

[25] 李森林, 张乐, 李瑾. 当代青年人工智能风险感知的测度与解析[J]. 科学学研究, 2023, 41(10): 1737-1746.
Li Senlin, Zhang Le, Li Jin. Measurement and analysis of contemporary youth's perception of artificial intelligence risks[J]. Studies in Science of Science, 2023, 41(10): 1737-1746.

[26] 刘亚丽, 范逢春. ChatGPT-AIGC 用户风险感知维度识别与治理研究——基于扎根理论的探索性分析[J]. 情报理论与实践, 2024, 47(3): 121-129.
Liu Yali, Fan Fengchun. ChatGPT-AIGC user risk perception dimension identification and governance research - exploratory analysis based on grounded theory[J]. Information Studies: Theory & Application, 2024, 47(3): 121-129.

[27] 刘金平, 周广亚, 黄宏强. 风险认知的结构, 因素及其研究方法[J]. 心理科学, 2006(2): 370-372.
Liu Jinping, Zhou Guangya, Huang Hongqiang. The structure, factors, and research methods of risk perception[J]. Psychological Science, 2006(2): 370-372.

[28] 王永强, 解强. 消费者对生鲜果蔬农药残留风险感知研究[J]. 大连理工大学学报(社会科学版), 2017, 38(2): 93-97.
Wang Yongqiang, Xie Qiang. A study on consumers' perception of pesticide residue risks in fresh fruits and vegetables[J]. Journal of Dalian University of Technology (Social Sciences), 2017, 38(2): 93-97.

[29] 姜泽玮. 人机交互中隐私风险感知的影响因素模型构建——基于智能音箱用户使用的实证研究[J]. 新闻界, 2023(8): 83-96.
Jiang Zewei. The construction of an influence factor model for privacy risk perception in human-computer interaction: an empirical study based on the use of smart speakers[J]. Journalism and Mass Communication, 2023(8): 83-96.

[30] 赵传林, 靳思缘, 武海娟, 等. 自动代客泊车接受度及停车选择行为影响因素分析[J]. 科学技术与工程, 2023, 23(35): 15259-15268.
Zhao Chuanlin, Jin Siyuan, Wu Haijuan, et al. Influence factors of automated valet parking acceptance and parking choice behavior[J]. Science Technology and Engineering, 2023, 23(35): 15259-15268.

[31] 褚旭龙, 史冬梅, 刘进长. 近 20 年中国智能机器人领域研究热点——基于 CiteSpace 的文献计量分析[J]. 科学技术与工程, 2023, 23(6): 2477-2484.
Chu Xulong, Shi Dongmei, Liu Jinchang. Research hotspots in China's intelligent robot field in the past 20 years: a bibliometric analysis based on CiteSpace[J]. Science Technology and Engineering, 2023, 23(6): 2477-2484.