中国心理学前沿
2022 年 8 月第 4 卷第 8 期

大众对人工智能道德担忧的双层结构分析
——基于扎根理论的探索

王云霄　龙　帅　黄译萱　陈　华

西南交通大学心理研究与咨询中心，成都

摘　要 | 人工智能的道德系统是动态的，不断更新的，以社会共同的道德担忧为基础来构建道德模型是重要研究方向，本文旨在探究大众对于人工智能的道德担忧模型。运用扎根理论对17个访谈资料进行分析，共形成23个范畴和8个主范畴。从主范畴的结构关系可以得出大众对人工智能道德担忧模型由两个层次组成，即因果层和影响层。其中因果层由道德认知和争议点构成的，争议点是根本原因，道德认知是直接原因；影响层由道德规范和争议点构成的，通过中介和调节来影响道德担忧。研究提供了减少大众对人工智能道德担忧的具体路径。

关键词 | 人工智能；道德担忧；道德规则；扎根理论

1　引言

随着人工智能系统的多领域发展，道德问题日益凸显。在自动驾驶汽车的例子中，如何对情景作出选择判断，对于乘客、场景中的其他人及社会环境有着不同的影响（Bonnefon，Shariff，and Rahwan，2016）。人工智能已经产生潜在的伦理风险，因此学界强调人工智能的道德代码与伦理嵌入的重要性。现行的道德设计包含三种理念：自上而下、自下而上及混合取向（Arkin，2009；Bringsjord and Taylor，2012）。自上而下又称基于规则，将人类认可的道德标准编码作为机器处理方式，使智能产品行动模式考虑到道德。自下而上的方案是提供恰当环境，使其通过经验而学习道德意识，类似于儿童一般。混合取向则认为，需要通过自上而下的方法设定一定的道德语法，并通过自下而上的情景学习而发展进化（Allen et al.，2005）。机器伦理的最终目标是：通过赋予机器道德能力以使其像人类一样自主地进行道德判断与行动（陈锐、孙庆春，2020）。为此，

基金项目：四川省心理学会年度科研规划重点项目成果（项目编号：SCSXLXH2021001）。
通讯作者：陈华，西南交通大学教授、硕士生导师，研究方向：健康心理学、人格心理学、发展心理与职业心理学，E-mail：chenhua991115@126.com。
文章引用：王云霄，龙帅，黄译萱，等. 大众对人工智能道德担忧的双层结构分析——基于扎根理论的探索［J］. 中国心理学前沿，2022，4（8）：886-897.
https://doi.org/10.35534/pc.0408107

机器伦理的研究者提出了诸多实现这一目标的方法，如康德式方法、范畴理论方法、博弈法，以及表面义务的混合方法等（Hooker and Kim，2018）。然而一个个问题随之产生，伦理规则及学习情境从哪里来？谁的道德认识应该被应用到道德框架中？在道德设置中人们担忧的问题是什么呢？

有研究者已经提出将我们的多人道德价值模型置于相关备选方案之上，或者使用所有人共同的道德忧虑（Brandt，Conitzer，and Endriss，2012；Brandt et al.，2016）。李楠通过对专家直觉及康德的道义论规则进行论述，认为智能机器的道德规则应该从普通大众的道德认识和担忧及行为样本中使用学习方法得出，从而得到最接近于普通人的人工智能。将多人的道德观念（通过机器学习）聚合在一起可能产生比个人或者单个群体更完善的道德系统，因为这减少了个人经验造成的特殊道德错误（Conitzer V et al.，2017），从而减少人们的道德担忧。基于人为基础的人工智能的决策过程可以更容易地被可视化和更新（Van Berkel et al.，2020）。但是目前对于大众的道德担忧探索较少。无法为人工智能的道德模型提供充足的理论及情景材料。另外对于道德情景及忧虑点的设置不是静态的，是通过多种方式不断进行补充（Van Berkel et al.，2020）。因此本研究利用扎根理论研究方法基于现实材料探究大众对于人工智能的道德担忧及道德约束等方面，构建出大众对于人工智能的道德担忧模型，探究减少道德担忧的路径，为人工智能道德模型增加理论基础。

2　研究设计与方法

2.1　研究方法

本文采用程序化扎根理论的质性研究方法，该方法能借助一条故事线将材料中分散的变量串连分析，可以知道研究以实证的方式对样本数据进行详细且反复地编码、归类和分析比较。这是一个自下而上对数据资料进行不断浓缩的过程，主要采用开放性编码、主轴性编码和选择性编码这三级编码程序。开放式编码将数据资料抽象形成概念或范畴；主轴式编码是发现和建立概念类属之间的各种联系，挖掘主范畴、范畴及其之间的关系；选择性编码是在所有已发现的概念类属中经过系统地分析以后确定核心范畴及范畴之间的网络关系（Strauss，1987）。

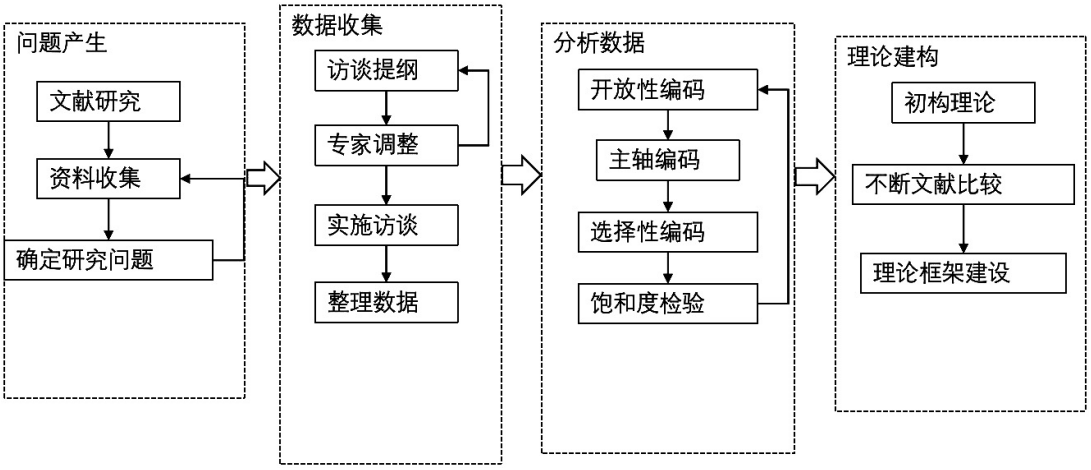

图 1　研究实施流程

Figure 1　Research flow chart

https://doi.org/10.35534/pc.0408107

2.2 数据收集

为了通过访谈获取更深入、更有用的信息，在正式访谈开始前，研究者先后向 5 位心理咨询方向的硕士生进行预访谈。并根据预访谈受访者反馈意见先后修改访谈提纲，形成最终版访谈提纲。本研究遵循理论抽样原则使用深度访谈的方法对 17 名对象进行了数据收集，采用一对一访谈的方式，时间为 30 ~ 50 分钟，受访者信息如表 1 所示。半结构化访谈提纲包括但不限于以下问题：（1）当智能产品普及到生活方方面面，您比较担心哪方面？（2）您愿意从个人和社会的角度详细谈一下对于人工智能的发展担忧吗？（3）如果发生了您担心的问题，您认为是什么原因呢？（4）您认为通过什么方法可以减少您的担忧么？每次访谈结束，立即撰写关于该访谈的备忘录。

表 1 受访者信息表

Table 1 Information of interviewees

项目	分类	人数	比例
性别	男	9	53.0%
	女	8	47.0%
年龄	21 ~ 25 岁	11	64.0%
	25 岁以上	6	36.0%
职业	学生	5	29.4%
	老师	2	11.8%
	人力资源	2	11.8%
	外贸	1	5.8%
	少儿编程	2	11.8%
	工程师	2	11.8%
	自由职业	3	17.6%

3 数据分析与结果

3.1 开放性编码

开放性编码是程序化扎根理论的第一步，是对获得的数据进行标签化、概念化和范畴化的处理过程。采用编码一致性及小组讨论的方式进行，由两名心理学研究生对于材料进行编码与讨论，一致性超过 80% 时停止，整个编码过程通过不断地新增、删除、合并，除去三份原始材料做饱和度检验之外，将数据逐步概括为 131 个概念，23 个范畴，如表 2 所示。

表 2 开放式编码过程

Table 2 The process of open coding

范畴化	概念化	原文示例
安全隐患	人类生存、安全隐患、生命代价的智能、反人类	人工智能过于智能，超过人类的智慧会对人类的生存有的影响
发展态度	前景广泛、被动接受、超出认知范围、技术无错、双刃剑，但利更多	对于整个社会性发展而言，机器人是必要的存在。如果是我的话，就放在一个天平上权衡，利益那边要更重要一些吧

续表

范畴化	概念化	原文示例
工具设定	无法代替决策、提高效率、核心功能为帮助、便利生活	机器人就相当于之前人工的一种工具，给我们社会变革的一种工具。其实最核心的目的要使我的生活更加方便，更加舒适。使我的工作效率更快更有效，避免一些重复性的工作
功能局限	过度解读、不确定性、需要更正、应对复杂情况、数据囊括有限、缺乏灵活性	所以在设计的时候，一个可能就是它大脑的评判系统始终还是没有人脑那么灵活多变，所以在处理一些比较复杂的车况时可能会出现一些问题
人机关系	关系定义、距离感、引导、减少道德压力、社会化交往的另一种方式、被动接受、从属关系	但是和人有一定的距离就会让人有一种心理上的安全感，谈不上一种对等，更重视服务从属的一种关系
规范来源	亲人亲社会、大众偏好为主、普遍道德认知、满足需求、三方赋予	首先一定要符合当前人们法律法规及道德的标准，符合当前人们最普遍的认知
规范设置难点	统一标准、规则限制决策权、规则滞后性、独立规则、符合期待	也很难在设计的时候就把这些东西都考虑得那么全面，它总是要涉及到更复杂的部分
规则的发展性	惩罚、分情景综合判断、规则灵活性、亡羊补牢为时不晚	需要不断地去学习，去完善，然后去教会它，我觉得这个规则需要不断更新
国家和法律	国家底线、法律规范、政府把控、官方引导、宏观角度	政府的敏感性。因为很多社会性的工作，无论是中国还是欧美，决策一块都是需要政府去把控
过度依赖	较早接触、负面影响、交往方式、依赖人工智能、现实脱节、成瘾	方方面面不可或缺，对人工智能的依赖会加重，个人价值何去何从
就业冲击	就业结构、底层就业、能力超越人类、行业人才把控、行业界限模糊、创造性工作	人就应该做创造性的工作。单调重复的工作机器人会做得比我更好。毫无疑问，肯定就会导致一部分人失业
控制权	决策在人、可控状态、学习情景不可控、控制行为、失控恐慌	这些产品反客为主的可能要考虑在内，人不想被它们主导，所以我想在使用这些产品的时候，还是能够保留自己的主导权
理性优势	后果最优解、功利性判断期望、理性道路	它就像是不会犯错的人的代替品
情感层面	情绪不可控、负面情绪、工具服务、信息集合、情感捕捉	拥有高级的认知水平，它有可能会变得很麻木，也有可能会发展出来更多的情绪，这是不可预测的
人机异同	人类智慧、道德标准、机器不属生物、行为原因、生活情景	我们做出某个决定是出于我们的良心，但在机器人看来，只是所有的解决方法中的一种而已，导致的后果是最优解的未知性
设计者和企业	研发公司、启发民智、市场自我感染能力、设计者规范带入产品	科学家应该预先设计怎么样能够让它更好地学习，更智能，同时又降低它出事故的概率
使用者责任	选择后果、风险承担、最底端责任、赋予权限	它做出的决策就是人赋予它的决策。所以它做错的事情就是因为人教坏了它。如果它出现问题就是这个教的人的问题
数据获取	数据便利、失控感觉、服务来源于数据、提供思路	提供一些数据给人工智能，才能更加合理地分析，或者更加全面地分析出来人到底是要提供什么样更好的服务
为利益服务	获取利益、利益和服务博弈、企业营利性、优势触发购买	利益是发展的原动力，获取利益的同时，也应该要做好相应的保障
信息茧房	影响认知、信息阻拦、信息困境、信息决策	感觉像是把人围在里面了，很难走出这样的一个困境，很难说……比如很难去见识到更多的产品
隐私泄露	生活透明化、侵犯个人隐私、不可避免	在我个人看来，非专业地看来，科技更加先进的同时，人工智能数据暴露的概率会更高一点
自主能力	自主学习进步、不断模拟、紧急情况自主处理、道德学习	机器要有自主能力，并且得不断更新才能适应时代
价值观影响	多数人利益、价值观监控、人类价值思考、社会主流价值观、思考方向局限	你思考的内容、你的价值观会被智能监听，也会通过各种方式受到影响

https://doi.org/10.35534/pc.0408107

3.2 主轴编码

主轴编码是对初始范畴进行关系梳理的步骤，包括因果、脉络结构等。对开放性编码中 23 个范畴进行联结、归纳和整合，最终形成 8 个主范畴，囊括在道德认知、道德担忧、道德规范及争议点这 4 个维度中，如表 3 所示。

表 3 主范畴结果表

Table 3 Main category result table

维度	主范畴	范畴	内涵
道德认知	人机定位	人机异同	人们从人类智慧、道德标准、行为原因和生活情景几个方面对智能机器人进行对比认识
		人机关系	人们从个人和社会的角度对人机关系进行定义理解，以寻找更好的相处方式
	认识角度	发展态度	人们从当前经验对于人工智能未来发展的态度和前景预估
		工具设定	人们对于人工智能的本质认识是辅助工具
		理性优势	人们从决策角度认为人工智能的最大优势是完全理性
道德担忧	社会影响	隐私泄露	担忧个人隐私的泄漏，但认为是不可避免的
		功能局限	对智能产品本身功能不足和由于缺乏灵活性引发的不信任
		就业冲击	人工智能对就业产生多样化的影响
		为利益服务	人工智能在服务和利益之间的博弈促进发展
	个体影响	安全隐患	对于人类自身安全的担忧
		价值观影响	人工智能的普及对人类现有价值观产生影响
		过度依赖	由于人工智能的便利性人类产生过度依赖
		信息茧房	人工智能影响人类对信息的正常筛选，限制多方面的信息来源
道德规范	规范设置	规范来源	规范设置应该来源于需求、大众道德共识等部分，拒绝片面设置
		规范设置难点	规范设置多方面考虑产生统一标准、责任分配等难点
		规则的发展性	人工智能规则是灵活的、可发展的
	外部信任	国家和法律	国家法律以人类利益为基础宏观设置规则
		设计者和企业	企业设计者以产品安全和需求为基础设置规则
	内部控制	使用者责任	使用者在外部规则下承担个人责任
		控制权	人类拥有控制权保障人工智能在可控条件下发展
争议点	争议点	情感层面	人工智能能否有情感层面的设计是具有争议的
		数据获取	人工智能对数据的获取没有标准，在隐私与完善服务之间产生矛盾
		自主能力	人工智能可拥有自主能力程度是较大争议点

3.3 选择性编码

选择性编码是对核心范畴的选择与设定过程。在对主副范畴进行完善的基础上，将其反馈到关键概念的类型特征中，从而建立核心范畴与其他部分的联系。其主要结构关系如表 4 所示。

表 4 主范畴关系结构表

Table 4 The structure table of main category relation

典型关系	关系结构图	关系内涵
人机定位 ➡ 道德担忧	因果关系	人们对于人机的异同和关系定位是产生道德担忧的直接原因
认识角度 ➡ 道德担忧	因果关系	人们对人工智能的认识角度不同产生的道德担忧维度也不同
道德担忧 ➡ 社会影响	表现关系	人们对于人工智能的道德担忧不同程度地表现在社会影响中

续表

典型关系	关系结构图	关系内涵
道德担忧 ➡ 个体影响	表现关系	人们对于人工智能的道德担忧从对个体影响表现出来
规范设置 ➡ 争议点 ➡ 道德担忧	调节关系	道德规范通过对争议点的规整来调节道德担忧
外部信任 ➡ 社会影响	中介关系	外部信任通过减少社会影响而对道德担忧起中介作用
内部控制 ➡ 个体影响	中介关系	内部控制通过减少个体影响而对道德担忧起中介作用
争议点 ➡ 道德担忧	因果关系	对于人工智能关于情感、自主能力等方面的争议是道德担忧产生的根本原因

本研究以大众对于人工智能的道德担忧为核心范畴建立故事线：（1）人们通过各种渠道了解到人工智能，从人机定位及其认识角度产生道德认知，这是道德担忧产生的直接原因，而对于人工智能关于情感、自主能力等的争议点是产生道德担忧的根本原因。（2）人工智能的社会影响构成了道德担忧的外部表现层，而对于个体影响的考虑构成了道德担忧的内部表现层。（3）规范设置可以通过规整争议点来对道德担忧进行调节。（4）通过国家法律等形成外部信任，通过减少社会影响而对道德担忧起作用，同样内部控制通过对个体的影响来对道德担忧起作用。

3.4 理论饱和度检验

理论饱和度状态是指在无法继续提炼和抽取初始概念的范畴下不再产生新概念。对已构建的理论框架能够较好地解释，理论就是“饱和”的状态。本文对于三份原始材料通过小组讨论和专家检验的方法进行饱和度检验，发现并无可提炼的新概念范畴，表明上述理论模型是饱和的。

4 大众对人工智能道德担忧的双层结构模型

根据访谈后的三级编码过程，通过进一步的凝练概括，构建了大众对人工智能道德担忧的理论框架。如图 2 所示，该框架包含因果层和影响层两个部分，包含因果、表现、调节和中介四种结构关系，可以从多个角度解释人工智能的道德担忧。

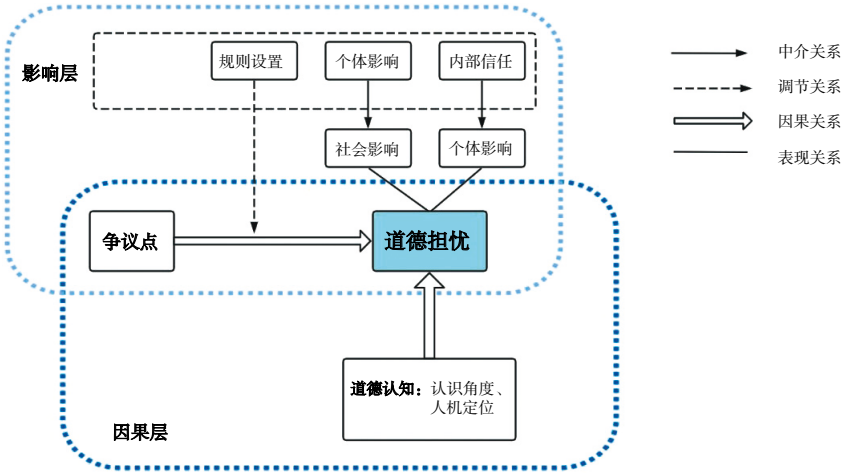

图 2 大众对人工智能道德担忧模型图

Figure 2 The model of popular moral concerns about AI

https://doi.org/10.35534/pc.0408107

4.1 人工智能道德担忧因果层分析

人工智能道德担忧产生的直接原因是人机定位和认知角度，包括人机异同、人机关系、工具设定、发展态度和理性优势 5 个范畴。涉及到智能，大众通过比对自身产生道德认知，冈克尔对于人机定位曾提出人工智能的道德是人为规范出来的，而人工智能也被定义为“他者”（Gunkel，2012）。大众对于人和机器的异同及关系的理解产生了价值观影响、信息茧房和过度依赖的个体道德担忧；对于人工智能的不同认知角度产生了安全隐患、就业冲击和利益服务等的社会道德担忧。而道德担忧产生的根本原因是人们对于人工智能的争议点，其中在情感层面对应着未知恐惧，而自主能力是构建人工智能主体的动力源，这些无法确定的争议点造成了道德担忧。

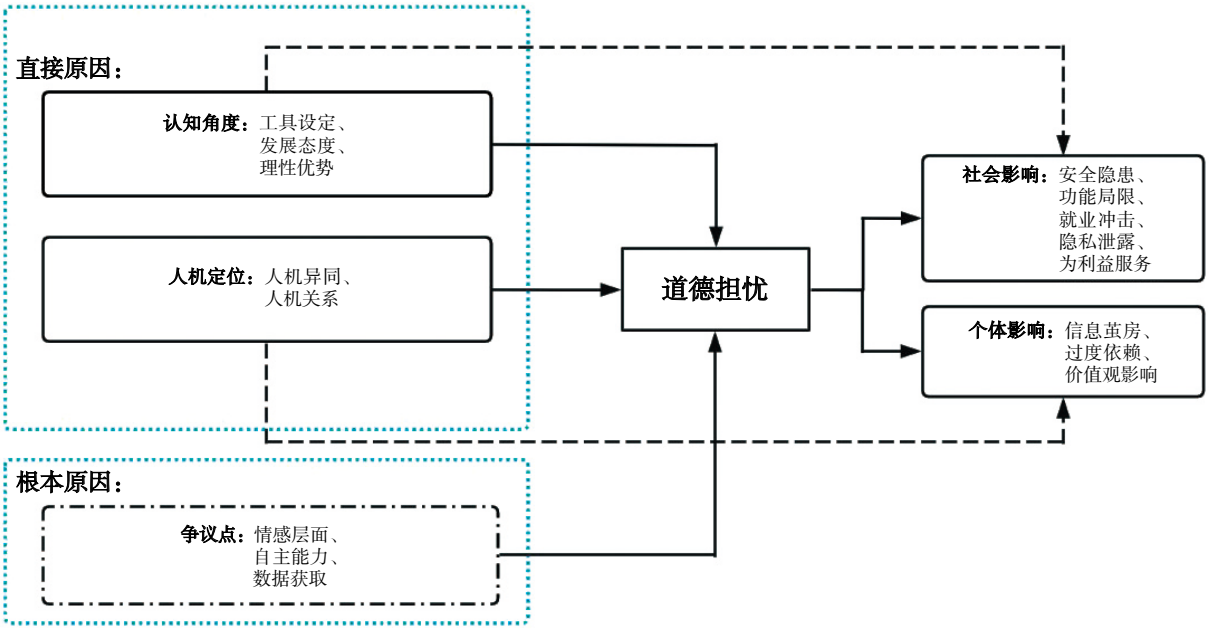

图 3 人工智能道德担忧因果模型图

Figure 3 The causal model of moral concerns about AI

4.2 人工智能道德担忧影响层分析

人工智能道德担忧的影响因素主要包括外部信任、规范设置、内部控制和争议点 4 个范畴。即由于道德认知产生的道德担忧受到道德规范和争议点的影响。其中外部信任通过国家的宏观调控和企业的需求满足来减少社会影响，同样的内部控制通过使用者责任承担和控制权的掌握来减少个体影响，从而共同影响道德担忧，也就是说人类的各部分需要承担起道德责任来减少道德担忧（Johnson，2006）。规范设置中规则难点、来源和发展性最终减少争议点带来的担忧，实现调节作用。

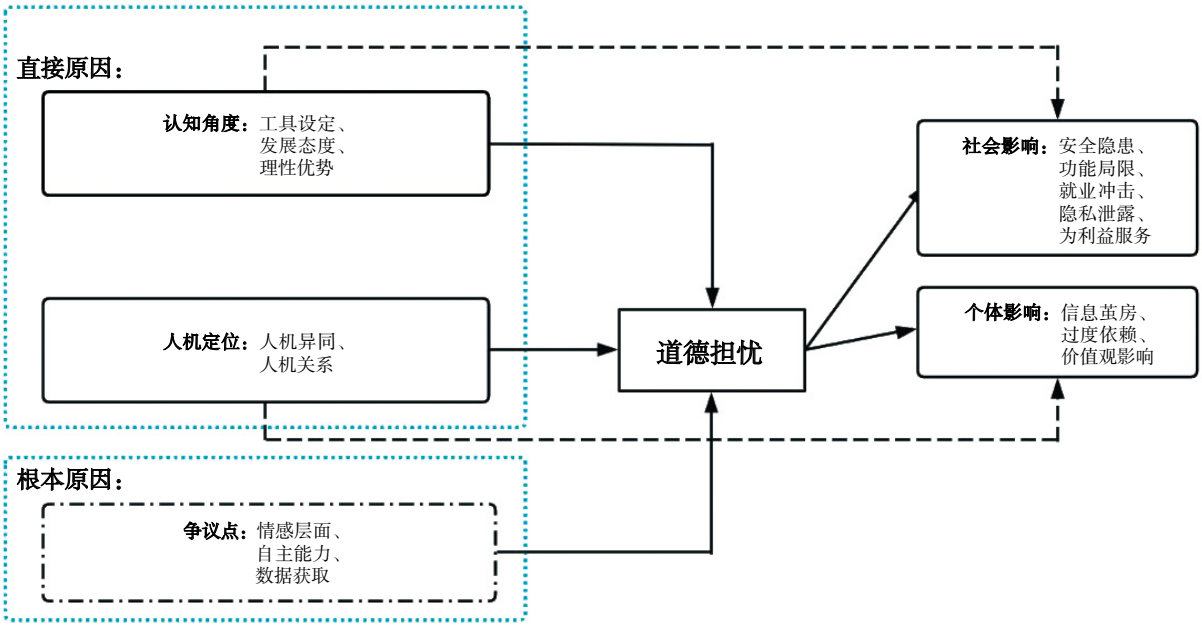

图 4　人工智能道德担忧影响因素模型图

Figure 4　The model of the influencing factors of moral concern about AI

5　讨论

5.1　人工智能道德担忧的因果关系

本研究发现，大众对于人工智能的道德认知，即人机定位与认识角度两个范畴是产生道德担忧的直接原因。人机定位方面，主要描述了大众对于人类与人工智能异同和关系的认识。大众普遍认为，意识、思维与情感等核心特征是人类独有的，因此在与人工智能交互时，人类应该始终处于主导地位。经典范畴观认为，身体、心灵与世界是独立的，正是因为这种独立，人会将一切外在物作为独立于“自我”的客体（Mcginn C，2015），以此保持自我的独特。但随着科技技术的发展，人类引以为傲的核心特征受到了挑战，智能和自主将不再是人类独有的特性（Luciano Floridi，2016），人工智能可以获得更强的思维能力，甚至模仿人类的情感过程，从而引发了对于人工智能的道德担忧。人工智能的普及及表现出的巨大力量，使得大众产生了价值观影响、信息茧房和过度依赖的个体道德担忧。人工智能是人类功能与价值的延伸，在给人类全新生活体验的同时，也可以彻底改变人类的生存形态（Kevin Kelly，2009），在这个过程中，对人类的价值观产生影响，其所“造成的人类分裂和对立，对人的内在性、自主性、平等性、整体性的颠覆和消解可能是不可逆的”（常晋芳，2019）。同时，人工智能特有的数据分析优势，使得它能收集并分析用户的各项信息，并提供个性化、针对性的内容，但在进行个性化信息精准推荐时，会造成“信息茧房”的问题，用户使用个性化推荐服务的时间与选择信息的主动性成反比，技术进步的同时会造成用户对信息的无节制满足（Katz J E，2002；张敏，2021），并逐渐缩小用户接受信息的范围，造成信息的单一化，甚至使受众产生依赖而进一步导致价值迷失。随着科技技术的发展，人工智能已经

https://doi.org/10.35534/pc.0408107

在人类日常生活中扮演了十分重要的角色，在提供便利的同时也产生了过度依赖方面的担忧。在各种智能化应用场景中，人类可能有意识或无意识地依赖并服从于机器决策而行动（赵志耘，2021），从而模糊了人类与人工智能的异同与关系，造成道德上的担忧。

在认识角度方面，从发展态度、工具设定、理性优势三个不同层面描述了大众对人工智能的普遍看法。相关调查显示，公众对人工智能的普遍了解程度较高（李思琪，2019），前景较为看好。现目前，人工智能作为给人类提供服务的工具，以及它处理数据而自带的理性优势，使得大众认为人工智能会在以后的社会中扮演更加重要的角色，从而引发了隐私泄露、就业冲击以及利益服务等的社会道德担忧。隐私权被认为是一项基本的人格权利，而现在的人工智能技术的发展是数据“喂养”出来的（孙少晶，2019）。所以，基于对人工智能的认识，大众难免会产生隐私方面的担忧。特别是，人工智能的算法对于大众而言完全是一个黑箱，信息往往在不知情的情况下就被收集了，人们难免对隐私信息是否会被过度收集而心存疑虑（Bryson，2018）。历史地看，每一次技术革命都会造成失业现象，人工智能取代人的工作岗位并不仅仅局限于劳动密集型产业（李传军，2021），在一些智力密集型行业同样如此（王前，2021）。一项调查显示，超过 60% 的人认为人工智能将会造成劳动力过剩和大量失业（李思琪，2019）。同时，大众对于当前人工智能发展的程度还有疑虑，担心一些技术上的局限会导致严重的安全隐患，并且如果人工智能以利益为导向被使用，那后果是严重的。有学者表示，一旦人工智能系统中植入了不良代码，或者人工智能系统被恶人控制，那么，其造成的损害将是人类所不能承受的（李传军，2021）。

而争议点方面，主要体现了大众对于人工智能更加抽象化的思考，主要分为三个方面，即情感层面、数据获取和自主能力。而道德担忧产生的根本原因是人们对于人工智能的争议点，其中情感层面对应着未知恐惧，而自主能力是构建人工智能主体的动力源，这些无法确定的争议点造成了道德担忧。

5.2 人工智能道德担忧的影响因素

道德认知产生的道德担忧受到了道德规范的影响，其中外部信任通过国家的宏观调控和企业的需求满足来减少由社会影响带来的道德担忧。新技术的发展，在带来便利的同时，必定伴随着阵痛，新技术的掌控者们占领了人工智能发展与应用的制高点（张正清、张成岗，2018）。使人变成更加难以掌控和保护自己的隐私与自由，并面临失业风险的弱者（Thucydides，2013），从而引起大众对于隐私、就业、安全等方面的担忧。而无论如何，在处理新技术的规制上．通过法律制度的调整显然是必要的方式之一（Hamilton，1980）。而且，我国政府长久以来树立的负责担当的大国形象，也使得大众普遍相信国家会运用法律或政策等手段，对人工智能发展可能带来的担忧进行管控。同时，据不完全统计，迄今为止国际上由不同国家、机构和团体公布的人工智能和机器人伦理标准文件已多达 50 余份。这表明，企业与设计者也会以产品安全和需求为基础，与国家政府合力避免道德担忧情境的发生，以此来减少社会影响，也就是说人类的各部分需要承担起道德责任来减少道德担忧（Johnson，2006）

同时，内部控制通过使用者责任承担和控制权的掌握来减少由个体影响带来的道德担忧。决策自主是人类实现自由的重要方面（Peterson M，2019），但是随着人工智能的发展，机器已经具备一定的自动决策能力，自动化道德决策可能存在使用户完全脱离道德决策循环的风险，进而可能对用户的道德自主

构成直接威胁(Millar J, 2016), 因此, 大众要求在人机交互中拥有控制权, 并在外部规则下进行责任承担, 以保障人工智能在可控条件下发展, 从而减少人工智能带来的个体影响方面的道德担忧。

规范设置方面, 包含了规则难点、来源和发展性三个范畴。其中, 规则来源强调了制定人工智能道德规则的参考点, 人工智能技术的发展不应该挑战和改变人类固有的伦理规范, 应该遵守人类长期以来遵守和执行的伦理规范（闫坤如, 2021）, 并转化为具体的设计要求（Poel I V D, 2013）, 规则难点反映了大众对于人工智能发展问题的深刻反思, 而规则发展性体现了规则应该与技术一起与时俱进。这些有关规则的看法最终减少了争议点带来的担忧, 即大众通过对人工智能各方面规则制定的认识, 一定程度上解决了对于情感层面、自主决策及数据获取方面的问题, 从而有效缓解人工智能的道德担忧。

6 结论

本研究通过扎根理论探索了大众对于人工智能的道德认知, 并以道德担忧为核心, 构建了大众对人工智能道德担忧双层结构模型。主要结论如下: （1）大众对人工智能道德担忧由因果层和影响层组成; （2）人工智能道德担忧产生的直接原因是人机定位和认知角度。道德担忧产生的根本原因是人们对于人工智能的争议点; （3）人工智能道德担忧受到道德规范和争议点的影响。其中外部信任通过国家的宏观调控和企业的需求满足来减少社会影响, 内部控制通过使用者责任承担和控制权的掌握来减少个体影响; （4）在大众对于人工智能道德担忧中可以通过规范设置的调节来减少争议点带来的担忧。通过人工智能道德担忧模型, 可以了解到大众在接受人工智能普及过程中最关注的问题, 并由此探究解决方法, 另外也可以为人工智能道德模型的搭建提供现实材料。

参考文献

[1] Arkin R. Governing lethal behavior in autonomous robots [R]. Chapman and Hall/CRC, 2009.
[2] Bringsjord S, Taylor J. The divine-command approach to robot ethics [C]. Robot Ethics: The Ethical and Social Implications of Robotics', 2012.
[3] Brandt F, Conitzer V, Endriss U. Computational social choice [J]. Multiagent systems, 2012 (2): 213-284.
[4] Brandt F, Conitzer V, Endriss U, et al. Handbook of computational social choice [M]. Cambridge University Press, 2016.
[5] Bryson J J. Patiency is not a virtue: the design of intelligent systems and systems of ethics [J]. Ethics and Information Technology, 2018, 20 (1): 15-26.
[6] Bonnefon J F, Shariff A, Rahwan I. The social dilemma of autonomous vehicles [J]. Science, 2016, 352 (6293): 1573-1576.
[7] Conitzer V, Sinnott-Armstrong W, Borg J S, et al. Moral decision making frameworks for artificial intelligence [C] // National Conference on Artificial Intelligence. AAAI Press, 2017.
[8] Johnson D G. Computer systems: Moral entities but not moral agents [J]. Ethics and information technology, 2006, 8 (4): 195-204.
[9] Katz J E, Aarkhus M A. Perpetual contact: Mobile communication, private talk, public performance [M].

https://doi.org/10.35534/pc.0408107

New York：Cambridge University Press，2002：1–30.
[10] Kelly K. Out of control：The new biology of machines，social systems，and the economic world [M]. Hachette UK，2009.
[11] McGinn C. Prehension：The hand and the emergence of humanity [M]. MIT Press，2015.
[12] Millar J. An ethics evaluation tool for automating ethical decision–making in robots and self–driving cars [J]. Applied Artificial Intelligence，2016，30（8）：787–809.
[13] Hooker J N，Kim T W N. Toward non–intuition–based machine and artificial intelligence ethics：A deontological approach based on modal logic [C]. In Proceedings of the 2018 AAAI/ACM Conference on AI，Ethics，and Society，2018：130–136.
[14] Gunkel D J. The Machine Question：Critical Perspectives on AI，Robots，and Ethics [M]. The MIT Press，2012.
[15] Peterson M. The value alignment problem：a geometric approach [J]. Ethics and Information Technology，2019，21（1）：19–28.
[16] Poel I V D. Translating values into design requirements [M]. Philosophy and engineering：Reflections on practice，principles and process，2013：253–266.
[17] Strauss A L. Qualitative analysis for social scientists [M]. Cambridge university press，1987.
[18] Thucydides. The War of the Peloponnesians and the Athenians [M]. Cambridge：Cambridge University Press，2013.
[19] Berkel N，Tag B，Goncalves J，et al. Human–centred artificial intelligence：a contextual morality perspective [J]. Behaviour & Information Technology，2020，41（3）：1–17.
[20] 陈锐，孙庆春. 道德计算是否可能：对机器伦理的反思 [J]. 科学技术哲学研究，2020，37（4）：74–80.
[21] 崔中良，王慧莉，郭聃，等. 人工智能研究中交互性机器伦理问题的透视及应对 [J]. 西安交通大学学报（社会科学版），2020，40（1）：123–132.
[22] 李传军. 人工智能发展中的伦理问题探究 [J]. 湖南行政学院学报，2021（6）：38–48.
[23] 李楠. 机器伦理的来源 [J]. 伦理学研究，2021（1）：103–108.
[24] 李思琪. 公众对人工智能发展的认识、担忧与期待 [J]. 国家治理，2019（4）：22–47.
[25] 张正清，张成岗. 第四次革命：现代性的终结抑或重构：信息伦理对人工智能伦理的启示 [J]. 武汉大学学报（哲学社会科学版），2018，71（3）：177–184.
[26] 孙少晶，陈昌凤，李世刚，等. “算法推荐与人工智能”的发展与挑战 [J]. 新闻大学，2019（6）：1–8.
[27] 王东，张振. 人工智能伦理风险的镜像、透视及其规避 [J]. 伦理学研究，2021（1）：109–115.
[28] 王前，曹昕怡. 人工智能应用中的五种隐性伦理责任 [J]. 自然辩证法研究，2021，37（7）：39–45.
[29] 闫坤如. 人工智能设计的道德意蕴探析 [J]. 云南社会科学，2021（5）：28–35.
[30] 郑戈. 在鼓励创新与保护人权之间：法律如何回应大数据技术革新的挑战 [J]. 探索与争鸣，2016（7）：79–85.
[31] 常晋芳. 智能时代的人—机—人关系：基于马克思主义哲学的思考 [J]. 东南学术，2019（2）：75–82.
[32] 张富利. 全球风险社会下人工智能的治理之道：复杂性范式与法律应对 [J]. 学术论坛，2019，42

（3）：68–80.
［33］张敏，王朋娇，孟祥宇．智能时代大学生如何破解“信息茧房”：基于信息素养培养的视角［J］．现代教育技术，2021，31（1）：19–25.
［34］赵志耘，徐峰，高芳，等．关于人工智能伦理风险的若干认识［J］．中国软科学，2021（6）：1–12.

A Bilevel Structure Analysis of Popular Moral Concerns about Artificial Intelligence: An Exploration Based on the Grounded Theory

Wang Yunxiao　Long Shuai　Huang Yixuan　Chen Hua

Psychological Research and Counseling Center, Southwest Jiaotong University, Chengdu

Abstract: The moral system of artificial intelligence is dynamic and constantly updated. It is an important research direction to build a moral model based on the common moral concerns of the society. This paper aims to explore the public's moral concern model for artificial intelligence. Using grounded theory to analyze 17 interview data, 23 categories and 8 main categories were formed. From the structural relationship of the main categories, it can be concluded that the public's moral concern model for artificial intelligence consists of two layers, namely the causal layer and the influence layer. The causal layer is composed of moral cognition and controversial points, the controversial point is the root cause, and the moral cognition is the direct cause; the influence layer is composed of moral norms and controversial points, and affects moral concerns through mediation and adjustment. Research offers concrete paths to reduce public concerns about the ethics of AI.

Key words: AI; Moral concern; Ethical norms; Grounded theory

https://doi.org/10.35534/pc.0408107