

“什么样的 AI 才是好 AI?”

——人工智能的电影再现与社会想象

王梦瑶

摘要:根据社会建构论与社会想象的理论主张,人工智能被大众接纳的关键在于社会想象中,它的样态以及与社会成员的关系符合主流价值规范。研究辨析有关社会想象的不同概念,将“技术的社会想象”定义为由经典科幻电影对技术的再现来表达与创设的社会想象。依据机器人分类法与人机交互相关理论对 30 个人工智能角色编码之后,将相关电影中人工智能的形态、角色、情感性、自主性视为条件变量,通过多值定性比较分析法,分析其组态与作为结果变量的人机关系之间的相关性,研究认为:担任保护或帮助角色,情感真实且自主性有限的 AI 更可能避免与人类产生冲突对立关系的社会想象,从而更容易被大众接纳。

关键词:人工智能;社会想象;科幻电影;技术接纳

中图分类号:J90 **文献标志码:**A **文章编号:**2096-5443(2025)06-0074-14

基金项目:国家社会科学基金青年项目(20CXW018)

一、研究背景与问题

人工智能(Artificial Intelligence,简称“AI”)作为一个富有潜力的交叉领域,已经引来众多学科的关注。除了以提升 AI 智能程度与效率为目的的研究之外,有关 AI 的监管、治理与伦理规范的研究也正当其时。以 AI 治理为主题的研究,大多涉及宏观层面,包括组织层面、产业层面、国家层面与国际层面,缺乏社会大众层面的关注^[1];注重开发者与设计者,却忽视领导层、部署者、终端用户和社区^[2]。在关涉社会大众的 AI 信任研究中,57.8%源自计算机科学与工程领域,社会科学和心理学只占约 17.8%^[3]。这些主流研究分别在以知识分子为主的科学共同体与以政府机构为主的治理行动者的支持下展开,旨在为 AI 的未来发展制定决策,但社会大众作为使用者没有受到足够重视。若想要 AI 以产品或服务的形式进入大众生活并且得到最大程度普及,其前提必然是被大众所接纳。因此,AI 技术的社会接纳研究十分必要。

在研究技术接纳方面,包括关于 AI 在内的科技研究(Science and Technology Studies,简称“STS”)中的大众传媒与人机交互(Human-Computer Interaction,简称“HCI”)研究各有偏重的范式。前者主要为社会建构主义,更偏重研究技术的“社会性”(sociality),后者主要为行为主义,更偏重研究技术的“物质性”(materiality)。然而,已有学者意识到,这两种范式的研究均有必要,不能偏向某方而将两者割裂,关键是揭示两者之间的深刻联系。其深刻联系并非基于纯粹理性的知识,而是借助想象力(imagery):重要的不是设计者设计技术的目的,而是在用户的想象中这个技术能干什么^[4]。近年来,AI 科技逐渐应用于大众层面,但是早在亲身体验 AI 科技之前,大众对其已经具备了一定的想象。这种想象业已存在多年,其主要来自科幻小说与受众更为普遍的科幻电影。为避免面向大众的 AI 样态与大众对 AI 应然性的先入之见(即对“好的 AI”的想象)发生严重冲突,AI 研究者们应该对相关想象有所研究。

因此,本研究提出的问题是:①身处科学共同体与决策圈之外的社会大众是如何想象 AI 的?②在大众的社会想象中,什么样的 AI 符合主流价值规范的?第一个问题指向 AI 社会想象的或然性,第二个问题则指向其应然性。如果将大众及其根植的社会文化背景纳入 AI 研究,那么,AI 的电影再现(representations)将是合适的研究对象。研究以“技术的社会想象”(social imaginary of technologies)为关键概念,通过考察 AI 的电影再现,勾画大众关于 AI 技术的社会想象,揭示何种 AI 样态在社会想象中更符合大众的相关价值规范偏好。

二、研究思路:基于电影探索关于 AI 的社会想象

(一) 目前研究路径的局限

讨论 AI 的社会想象,有两个重要概念:一是“社会想象”(social imaginary),二是“社会技术想象”(sociotechnical imaginaries)。然而,这两个概念及其相关理论各有缺陷,不足以在方法上支撑对 AI 社会想象的研究。

查尔斯·泰勒所提出的“社会想象”指普通人在日常生活中对社会整体及其秩序、合法性和社会成员相互关系的想象,它使得人们能够自然合理地在共同世界中协同行动。简言之,社会想象是体现“我们是谁”“我们如何共处”“什么是正当与可欲”的集体图景。他提出该概念是为了超越传统“社会理论”(social theory)的抽象理性框架,因此,他强调社会想象通常不是理论术语,而是以形象、故事和传说来表述的。^[5]泰勒的理论宏大抽象,不足之处在于没有对科技进行专门论述。

在更为具体的科技研究领域,Jasonoff 等提出“社会技术想象”这一概念,其定义为“集体所持的、在体制上稳定下来的并公开表达出的对可欲的(desirable)①未来的愿景”^[6]。社会生活与社会秩序的某种形式(通常是指西方现代化的形式)是通过科技进步实现的,同时,这一形式也支持着科技进步。人们以科技进步为默认前提与共识来设想未来的社会生活和社会秩序。然而,这一概念在我们试图理解某项特定技术时仍有方法论上的不足。它将科学与技术作为一个整体来考虑,将其融入现代社会秩序的构成中;同时,它削弱了泰勒原本对“普通人的日常理解”的强调,转而强调其制度化与理性表述。目前应用该理论的研究认为需要将社会技术想象置于公共传播(public communication)中来建立分析框架,分析对象是公共建构(publicly constructed)的想象^[7]。AI 的利益相关群体,如技术专家与政策制定者,在塑造 AI 公共话语与社会想象的过程中扮演关键角色^[8]。显然,这样的研究仍然缺乏社会大众的视角,背离了“社会想象”这一概念的初衷。此外,有批评者指出:“(社会技术想象)的概念……并未明确阐释‘想象是如何被激活的、以何种方式被激活、为何会以某种方式被激活’等问题。”^[9]本研究认为,在“社会技术想象”概念中,普通人对于科技进步作为一个整体的态度是默许的,或者说无力反抗的。我们也应当意识到,这并不意味着人们默认每一项科技成果在社会文化意义上都是可欲的,这就需要针对具体技术进行实证分析。

(二) 科幻电影激活技术的社会想象

为弥补目前研究的不足,本研究提出“技术的社会想象”(social imaginary of technologies)的概念,专指在某个特定的社会文化语境之中对某项技术的社会想象,它基于普通人的共享理解(由日常生活中的直接知觉与经验构成),以形象、故事和传说的形式来表达,深层映射普罗大众的相关价值规范偏好。这个概念将目光聚焦“技术”这一特定的对象,并继承泰勒“社会想象”概念的三要素,即“规范性”“日常性”和“想象性”。经典电影同时具备这三要素,其足以表达并创设社会想象,其自身也是社会想象的一部分,以电影再现作为研究对象也可弥补社会想象系列概念与理论在方法上的不足。“规范性”“日常性”“想象性”的具体含义界定如下。

“规范性”是指技术的社会想象受特定社会文化背景中规范的制约。如前所述,技术不仅具有物

①将原文中的“desirable”翻译为“可欲的”,参考了政治哲学中的惯用翻译;将“desirability”翻译成“可欲性”,通常表示一种社会秩序的合法性或者令人满意与值得追求的程度。

质性维度,即为实现某种功能而被设计制造出来的,作为方法、设备或应用的存在,还由社会性维度构成。社会性维度指技术的样态与可能发挥的作用是被社会建构的,因此是随社会文化背景而变的。技术被社会大众接受与否,很大程度上取决于它的社会想象是否符合大众的价值规范。例如,摄影技术在其早期的社会想象中有“摄魂”功能,引发了民众的恐惧与反对。鲁迅曾在《论照相之类》中回忆:“照相似乎是妖术。咸丰年间,或一省里,还有因为能照相而家产被乡下人捣毁的事情。”^[10]

“日常性”的内涵可借助现象学生存论中的生存状态来说明:“现象世界是为我而存在的,它虽然作为背景和视域,不能明晰地展现在我的意识中,但它却具有可确定的意义,因此,我并不会对世界中的事物感到陌生。”^[11]社会想象的日常性意味着:日常中的周围世界(surrounding world)为普通人的社会想象提供能够被识别的稳定背景。当同样或类似的可识别元素被知觉^①到时,具备经验的人们无需复杂的认知表征活动就能够理解其意义。总之,社会想象中必然包含着可供识别的元素、对元素的直接知觉以及“未经整理、未经理论化”的共同理解。例如,芯片是计算机技术中的核心要素,但在技术的社会想象中未必如此。普通人日常理解中的计算机技术包含着诸如“屏幕”“界面”以及使用者的“手”这样的可识别元素作为其核心要素。在科幻电影中,即使屏幕呈现为投射在空中的虚拟平面,实体鼠标和键盘也同时被取消,但“用手来触摸界面输入指令”的影像经久不衰,因为它能唤起人们对使用计算机的直接经验,很容易令人理解这是计算机技术的一种未来变体。

“想象性”重新强调泰勒所谓的“形象、故事和传说”,但不仅仅如此。技术的社会想象中的“想象性”就在于社会大众赋予技术在其设计功能之外的意义。相比泰勒,哲学家 Castoriadis 对“想象性”的阐述更为全面与深刻。他提出,“当我们提起‘想象’时,我们谈论的是‘被创造’的某些东西……符号在其通常的或规范的指称之外又获得其他指称,在此过程中,意义发生转变”^[12]。技术在被大众想象时,通常是不符合现实的,是对现实技术的幻想、扭曲或夸张。例如,在两次世界大战期间,电影等大众传媒对人们的认知与思想观念产生的影响效果被渲染为“魔弹论”^[13],由此,在库布里克(Stanley Kubrick)导演的科幻电影《发条橙》(*A Clockwork Orange*)中,电影这一项实际中已经存在的技术,被想象成一种直接操控大脑或神经的恐怖手段。

Castoriadis 区分了想象的三种层次:根源想象(radical imaginary)、现实想象(actual imaginary)^②与象征想象(symbolic imaginary),分别是指原初的创造(形象、传说、神话)、社会化制度(法律、规范)与符号系统(语言、标志)。^[12]其中的根源想象是后两者的基础与根源,后两者作为象征形式的集体想象在社会中起到真实作用,人类社会就是通过这些集体想象来创设(institute)自身。借鉴这一观点,我们可以将电影视作基于根源想象的象征形式,即利用原初的创造(形象、传说、神话)而生成的一套符号系统,是与法律、语言和宗教类似的创设社会的想象力量。电影不仅反映世界,还通过塑造主体对自身和他人的看法,积极参与构建社会现实。因此,电影既有助于巩固已有的、占主导地位的具有稳定意义的想象,也有助于创设新的、兼具意义创造的破坏性与创造性潜力的想象^[14-16]。

通过以上阐述,一个完整的“社会想象”的概念就与作为大众媒介的电影紧密联系在一起。电影本身具有“日常性”。电影创作者按照视听语言规则精心设计的场景与运镜,帮助观众直接知觉(主要是视觉与听觉)并理解其中表达的内容,同时令观众将注意力集中在银幕中心,无意识地区分“前景”与“背景”,默然领会背景的意义、接受背景的设定。同时,电影通过剪辑或特效等技术重新安排现实逻辑,创造并不存在的人物、宇宙、事件与时空关系,其本身就是想象的产物,即具备“想象性”。正是这种想象性激活观众自己的创造力——他们并非被动接收者,而是创造电影意义的主动参与者。最后,电影中再现的内容与其创造者试图传递的价值观都必须在某种程度上符合主流规范要

①这里的“直接知觉”强调人的意向性、身体与环境之间的互动,其理论来源为德雷福斯(Hubert Dreyfus)的身体现象学与吉布森(James Jerome Gibson)的直接知觉理论,但本文篇幅有限,在此仅作简述。

②现实想象并不是对实际存在的想象,而是指一种在现实中有效的、真正发生作用的想象,即“社会集体创设的意义系统”。

求,否则,电影无法获得现存制度的许可而上映,或者上映后遭到大部分观众的批评与反对,因此,也具有“规范性”。

经典科幻电影将一项现实中尚未存在的技术的社会想象投射为可以被观众直接知觉的视听作品,通过三要素的成功结合,将其中的某些设定隐入“背景”,塑造观众的“日常理解”,从而建构现实。这便是技术的社会想象被激活,继而发挥现实影响的过程。由是,经典的科幻电影对技术的再现可用于研究技术的社会想象。诚如 Jasanoff 等所言:“将技术置于整合了物质、道德与社会图景的情境之中的概念框架,在科技研究的文献中很少看到,但科幻小说对此贡献良多。”^[6]而科幻电影与小说相比,因其诉诸视听的直接知觉的特性,更能激活技术的社会想象。

三、研究设计、方法与过程

遵循以上研究思路,本研究聚焦“人工智能的社会想象”,考察以 AI 为主题的经典科幻电影对 AI 的再现,发掘 AI 在电影中经常被使用却未经省察的典型设定,揭示大众经由电影所形成的 AI 社会想象,以此回答研究问题:AI 是如何被社会大众想象的,以及在这些想象中什么样态符合大众的价值与规范偏好。因此,本研究将首先采用定性研究方法对经典科幻电影进行定义与抽样,并基于机器人分类法与媒介互动理论,总结与凝练若干初始范畴,然后选取关键性范畴作为 AI 在社会想象中的规范性基础,通过多值定性比较分析(Multi-Value Qualitative Comparative Analysis,简称“mvQCA”),探索何种条件组合(组态)是更有可能被普通人接纳的 AI 想象。

(一)操作定义与抽样标准

本研究对科幻电影的定义是以科学原理或科学假说为叙事基础,通过虚构未来或异质世界,探讨科技发展对人类影响的电影类型。将科幻电影中的 AI 角色定义为人造智能主体,其特征包括:一是由人创造的,而非在人自身基础之上改造而成的半人半机器“赛博格”,如《攻壳机动队》中的素子,也非外星文明的造物或生命体,如《新世纪福音战士》中的“使徒”;二是由有科学依据的能源来驱动,排除掉部分奇幻、魔幻类型电影中所出现的机械造物,如《绿野仙踪》中的铁皮人^[17];三是至少具备部分人类智能,例如能够以人类语言与人类沟通。本研究的抽样范围并非所有包含 AI 角色的科幻电影,而是要满足以下标准:一是时间范围,研究以 1927 年电影《大都会》为 AI 电影的起点,设定时间范围为 1927 年至 2022 年,新近的电影尚未经历客观公正的评价,不列为分析对象;二是限定 AI 主题,电影须以 AI 角色为主要角色(之一)或推动情节发展的关键角色,而并非纯粹是执行任务的辅助性工具,例如《星际穿越》中的机器人 TARS 和 CASE;三是影响力大,电影须至少满足以下四项条件中的一项,包括影史地位(是否被电影教科书引用)、获得奖项或提名、IMDb 或豆瓣评分高于 7 分、票房收入高于 1 亿美元。前两项可衡量作品的艺术性,而后两项可衡量作品的大众认可度。在此硬性标准下,本研究采取“最大差异抽样”策略^[18],尽可能选取不同国家出品、不同语种发行、多种题材、不同人导演的电影,如雷德利·斯科特(Ridley Scott)导演的《普罗米修斯》就因已取样了《银翼杀手》而被排除在外,最后得到 22 部电影(包含一部系列电影)样本与其中的 30 个 AI 角色。具体见表 1。

表 1 AI 科幻电影样本及其 AI 角色

序号	年份	名称	制片国家	AI 角色编号	AI 角色
1	1927	《大都会》(<i>Metropolis</i>)	德国	1	机器人玛利亚(Maria)
2	1965	《阿尔法城》(<i>Alphaville</i>)	法国/意大利	2	阿尔法 60(Alpha 60)

续表

序号	年份	名称	制片国家	AI 角色编号	AI 角色
3	1968	《2001:太空漫游》 (<i>2001:A Space Odyssey</i>)	英国/美国	3	HAL 9000
4	1973	《西部世界》(<i>Westworld</i>)	美国	4	牛仔机器人(The Gunslinger)
5	1982	《银翼杀手》(<i>Blade Runner</i>)	美国	5	复制者(Replicants)
6	1982	《电子世界争霸战》(<i>TRON</i>)	美国	6	主控程序(Master Control Program)
7	1984	《终结者 1》(<i>The Terminator</i>)	美国	7	天网(Skynet)
				8	终结者(T-800)
8	1986	《霹雳五号》(<i>Short Circuit</i>)	美国	9	约翰尼 5 号(Johnny 5)
9	1991	《终结者 2:审判日》 (<i>Terminator 2:Judgment Day</i>)	美国	10	液态金属终结者(T-1000)
10	1995	《攻壳机动队》 (<i>Ghost in the Shell</i>)	日本	11	傀儡师(Puppet Master)
11	1999/ 2003	《黑客帝国》(<i>The Matrix</i>) 《黑客帝国 2:重装上阵》 (<i>The Matrix Reloaded</i>) 《黑客帝国 3:矩阵革命》 (<i>The Matrix Revolutions</i>)	美国	12	母体(The Matrix)
				13	建筑师(The Architect)
				14	先知(The Oracle)
				15	史密斯(Agent Smith)
12	2001	《人工智能》 (<i>A. I. Artificial Intelligence</i>)	美国/英国	16	大卫(David)
13	2004	《我,机器人》(<i>I,Robot</i>)	美国/德国	17	桑尼(Sonny)
				18	薇琪(VIKI)
14	2008	《机器人总动员》(<i>WALL·E</i>)	美国	19	瓦力(WALL·E)
				20	伊芙(EVE)
				21	自动驾驶系统(AUTO)

续表

序号	年份	名称	制片国家	AI 角色编号	AI 角色
15	2013	《她》(<i>Her</i>)	美国	22	萨曼莎(Samantha)
16	2014	《超验骇客》(<i>Transcendence</i>)	美国/中国/英国	23	卡斯特(Will Caster)
17	2014	《机械姬》(<i>Ex Machina</i>)	英国	24	艾娃(Ava)
18	2015	《超能查派》(<i>Chappie</i>)	美国/墨西哥/ 南非	25	查派(Chappie)
19	2015	《复仇者联盟 2: 奥创纪元》 (<i>Avengers: Age of Ultron</i>)	美国	26	奥创(Ultron)
				27	幻视(Vision)
20	2017	《银翼杀手 2049》 (<i>Blade Runner 2049</i>)	美国/加拿大	28	乔伊(Joi)
21	2018	《升级》(<i>Upgrade</i>)	美国/澳大利亚	29	STEM
22	2021	《芬奇》(<i>Finch</i>)	美国	30	杰夫(Jeff)

注: 其中 AI 角色名称的中文翻译与制片国家的信息均来自豆瓣电影。

(二) 编码设计

为了给电影再现的编码构建较为全面的覆盖范畴,本研究将参考现实中对“机器人”(Robot)技术的分类方法。Steinert 认为,机器人分类法应该整合技术属性与社会功能双重维度,因为机器人的广泛应用必然带来巨大社会影响。其技术属性维度包含依据物理形态构成的分类,因为机器人的外观会在人机互动中对人产生不同的情感与情绪影响;其社会功能维度包含依据应用领域构成的分类,家用机器人与战斗机器人应具备不同级别的自主性,因为两者给人类带来的风险相差甚远。^[19]由此,本研究以技术属性维度区分电影中 AI 本体存在的客观特征,如仿生人外观或机械外观等;以社会功能维度区分电影中 AI 在叙事中的功能定位,如管理者、助手等。另外,此分类法通过技术复杂度梯度构建层级体系,比如,“无自主”指依赖预设程序(如第一代工业机器人),“部分自主”指环境感知下的条件响应(如第二代协作机器人),“完全自主”指决策算法驱动的独立行为(如第三代智能机器人)。

然而,现实中的机器人分类是从机器人本体出发,与 AI 的电影再现并不完全等同,我们还必须基于人机互动相关理论与电影的特性来设计编码表。经由实验方法验证的媒介等同理论(Media Equation Theory)提出:人类在与媒介(如计算机、电视、新媒体)互动时,会无意识地将其实视为真实的人或环境,并产生与真实互动相似的社会性和自然性反应^[20]。其中,社会性反应是指人类对媒介中的

社会线索(如声音、表情)产生类似人际互动的反应,例如对智能助手使用礼貌用语、对反派角色产生愤怒情绪。但是,该理论认为人们对媒介所呈现出的社会线索产生的反应更像是一种无意识的行动,即人并未真正承认机器的主体地位。对此进行修正的媒介唤起范式(Media Evocation Paradigm)揭示了人们是有意识地 and 媒介进行互动。作为互动对象的智能体有无生命、有无自主性、能否社交、与人的相似性等标准都成为人们主动界定它能否作为同等地位的社会行动者的维度^[21]。Barneck 等学者也通过实验验证了机器人外观设计,如形态、拟人化程度,会影响人类对其生命性(animacy)和智能感知的评估^[22]。例如,更接近人类形态的机器人被认为具有更高自主性和情感表达能力。此外,人类对 AI 角色的接受度往往受“恐怖谷效应”(uncanny valley)的影响^[23]。当 AI 在外形和行为上高度拟人却又未完全逼真时,可能引发观众的不适与排斥^[24]。这提示我们,无论是操控者还是被操控者,高度拟人化的仿生人 AI 都可能会被描绘为危险的存在,而机械感较强的 AI 更易获得正向认同。这一理论可为形态与人机关系这两个变量之间的关系提供解释。

参考这些理论,对于 AI 的电影再现,本研究总结提炼出 5 个主范畴,分别是:AI 的形态、角色、情感、自主性与人类的关系。具体见表 2。举例来说,经典科幻电影《银翼杀手》中的 AI 角色——复制者,其组态及取值应该是:形态,具身的仿生人型(1);角色:动乱者(3);情感,真实情感(2);自主性,高自主性(2);与人类的关系:对立,敌对斗争(1)。

表 2 编码表

主范畴	描述说明	子范畴		描述说明	取值	举例
形态 (Form)	AI 是否拥有物理形体	具身	机器型	外观有明显金属感,并伴有机械装置	0	约翰尼 5 号
			仿生人型	外观与人类相似,有五官与四肢,大部分躯体由皮肤覆盖	1	大卫
		非具身	虚拟形象型	通过屏幕、全息等形式呈现人类的形象	2	萨曼莎
			意识流型	作为程序,主要存在于网络或系统	3	主控程序
角色 (Role)	AI 在叙事中的角色类型	英雄		拯救或保护人类、维护伦理原则	0	瓦力
		伙伴助手		协助人类完成目标	1	杰夫
		管理者		主导人类或其他 AI 活动,实行技术型管理	2	母体
		动乱者		为提升自身地位,暴力破坏现存秩序	3	复制者

续表					
主范畴	描述说明	子范畴		取值	举例
情感 (Emotionality)	AI 是否具备情感,以及情感的真实性的真实性	无情感		没有展现任何情感或仅是逻辑反应	0 STEM
		模拟情感		按照预设程序,能够对人类做出情感反馈,但无真实体验	1 薇琪
		真实情感		展现深层复杂情感,如爱、悔恨、挣扎、自省等	2 桑尼
自主性 (Autonomy)	AI 在决策和行动上的自主程度	无自主性		不改变既定目标,完全按照预设程序行事	0 阿尔法 60
		有限自主性		以任务目标为导向,具备环境适应能力,可局部优化任务,但无法设定自身目标	1 液态金属终结者
		高自主性		挑战原始程序设定、违抗人类指令、进行反思、自主决定是否修正目标	2 HAL 9000
与人类的关系 (Human-AI Relationship)	AI 与人类之间的关系	对立	统治压迫	AI 与人类之间存在巨大的、难以逾越的能力与权力的不平等,近似支配与奴役	0 主控程序
			敌对竞争	能力相近的主体之间发生冲突斗争	1 奥创
		非对立	合作陪伴	与人类建立友好协作的关系	2 查派
			融合超越	通过融合而超越人类,成为新的存在	3 卡斯特

为了保证编码更为客观,本研究采取“关键友”(Critical Friend)方法。这一方法原初用于教育研究,以促进研究者在反思过程中实现视角的拓展与偏差的修正^[25],近年来也广泛应用于质性研究与QCA编码环节。在明确操作性定义并形成初步编码手册之后,研究者邀请一位熟悉影视文化与质性方法的同侪(南方某大学相关专业副教授)担任“关键友”,分别独立进行编码,完成后通过交叉比对,统计一致率,并标记出存在分歧的编码点。其中较多的分歧点出现在“自主性”范畴。对于不一致之处,双方开展有依据的讨论,逐项对照角色文本、影片情节与理论定义进行溯源式分析,无法取得共识时返回原始资料重看关键场景,对定义或标准进行微调,以提高操作性。改进后,对“自主性”子范畴的描述更为精确,最终协商一致,确认并归档编码表。

(三)多值定性比较分析

对于小样本研究,尤其是要分析的因素多为分类变量或定序变量时,定性比较分析方法经常被使用。本研究的研究样本规模较小,且在编码设计中得到的主范畴有 5 个,不仅契合该方法的适用范围,也符合其对变量数量的要求。由于各个变量均非二分变量,应该采用多值定性比较分析方法。

假设社会想象中 AI 与人类之间的关系为决定 AI 能否在社会想象中符合主流价值规范的因素,那么我们就可以将其设定为结果变量,而形态、角色、情感、自主性均为条件变量,通过 mvQCA 探索何种 AI 组态会导致不同的社会想象。将按照编码构建的真值表输入由复旦大学复杂决策分析中心开发的 fm-QCA 软件得到分析结果^①。

将一致性阈值设定为 0.9 进行必要性检验,结果显示:“敌对竞争”关系的必要条件是动乱者角色;“融合超越”关系的必要条件为高自主性;其余条件则没有达到必要性门槛。这与 mvQCA 方法所强调的“因果复杂性”(causal complexity)相一致,即:某一结果往往不是由单一因素决定,而是多个条件组合的产物,同时提示未来分析中可进一步探索联合必要性与复杂交互组态。例如,高自主性在“统治压迫”“敌对竞争”与“融合超越”关系中均为非常重要的条件变量,但因为组态构成不同,导致结果变量的不同:“统治压迫”关系下的管理者角色与意识流型结合在一起,而“敌对竞争”关系下动乱者角色与机器人型结合在一起。这提示“统治压迫”关系与“敌对竞争”关系之间差异的关键在于角色与形态的固定结合,这种固定结合可视为电影供观众识别与理解的核心要素,将观众导向不同的社会想象。另外,同样是真情实感的机器人型 AI,自主性差异会导致对立关系(敌对竞争)与非对立关系(合作陪伴)的重大区别,这值得进一步探究与说明。

四、研究结果

(一)分析结果解读:四种想象类型

进一步组合各条件变量得到不同的组态,计算出的部分结果如表 3 所示。

表 3 多值定性比较分析的部分结果

结果变量	案例数	解法	组态编号	组态	原始覆盖率 (%)	唯一覆盖率 (%)
对立关系	统治压迫	10	复杂解/中间解	1 Form { 2, 3 } * Role { 2 } * Emotionality { 0, 1 } * Autonomy { 2 }	60	30
				2 Form { 2, 3 } * Role { 2 } * Emotionality { 0, 1 }	80	30
				3 Role { 2 } * Emotionality { 0, 1 } * Autonomy { 2 }	60	20
		一致	4	Form { 3 } * Autonomy { 2 }	70	70
			5	Form { 3 } * Role { 2 }	60	60
			6	Role { 2 } * Emotionality { 0 }	60	60

^①按照惯例,设置案例数阈值为 1,一致性阈值为 0.8,然而,由于“融合超越”关系下的案例太少,将其一致性阈值设置为 0.6。另外,表 3 中“一致”是指包括简约解(Parsimonious)在内的三种解法下的组态是一致的,显示分析结果不受解法选择的影响,故具有高度可靠性,本研究将重点解读这些组态。

续表						
结果变量	案例数	解法	组态编号	组态	原始覆盖率 (%)	唯一覆盖率 (%)
对立关系	敌对竞争	复杂解/中间解	7	Form {0} * Role {3} * Emotionality {0,1}	57	29
			8	Role {3} * Emotionality {0,1} * Autonomy {1,2}	71	71
		一致	9	Form {0} * Role {3}	57	57
非对立关系	合作陪伴	复杂解/中间解	10	Form {0} * Role {0,1} * Emotionality {2}	67	44
			11	Form {0} * Role {0,1} * Autonomy {1,2}	67	56
		一致	12	Role {0,1} * Emotionality {2}	89	89
			13	Form {0} * Emotionality {2}	67	67
			14	Role {0,1} * Autonomy {1}	56	56
	融合超越	一致	15	Form {2} * Emotionality {2} * Autonomy {2}	50	50

1. “对立关系”社会想象中的 AI

第一,在“统治压迫”关系之下,达到了高度一致性(大于等于 80%)与高覆盖率(大于等于 60%)的组态 1~3 中出现了至少两次的共性为非具身型(Form=2,3)、管理者角色(Role=2)、无真实情感(Emotionality=0,1)与高自主性(Autonomy=2),集中说明:无真实情感且非具身形态的 AI 最有可能通过技术型管理来统治人类。组态 4~5 则进一步具体表明毫无情感且高度自主的意识流 AI 是这一类的代表性想象。

第二,在“敌对竞争”关系下组态 9 为具身形态的机器人(Form=0)与动乱者角色(Role=3)的结合,传达了通过发起动乱来与人类争夺权力的 AI 机器人形象。另外,在组态 7 与 8 中分别出现的情感低下(Emotionality=0,1)与自主性(Autonomy=1,2)暗示当 AI 无法产生真实情感却具备自主性时,更可能变成对抗人类的“敌人”,导致激烈冲突。

2. “非对立关系”社会想象中的 AI

第一,真实情感(Emotionality=2)与正面角色(包括英雄与伙伴助手,Role=0,1)出现在“合作陪伴”关系下的大部分组态中,说明承担保护与帮助人类的责任的 AI 具有真实情感会更可能与人类和谐相处。组态 11 与 14 中共性为有限自主性(Autonomy=1),说明适度的自主性与情感能力支持合作陪伴关系,而完全无自主或完全自主都不利于合作。三解一致的组态 13 中,机器人形态与真实情感的组合,形成了经典的友好机器人形象,如瓦力、约翰尼 5 号、查派与杰夫,也印证了上文“机械感较强的 AI 更易获得正向认同”的推论。

第二,“融合超越”关系下案例最少,但仍然出现高一致性与较高覆盖率的组态,适于提炼典型。

组态 15 表明“融合超越”的 AI 角色,往往具备以下特征:拥有人类形象(无论具身与否,物理的或虚拟的人类形象皆可),具备真实情感(与人类建立精神/哲学层面的连接),极高的自主性(如不依附任何个体、具备自我价值观)。

总结以上发现,本研究将电影再现的 AI 社会想象分为 4 种类型,分别是“冷酷统治”“你死我活”“人类优先”与“另类飞升”。其中,“冷酷统治”与“你死我活”的想象可能招致大众对 AI 的恐惧与排斥,“人类优先”则是最为可欲的未来愿景,即符合主流价值规范的。而“另类飞升”在 30 个 AI 角色案例中只占 4 个,显示出这类想象自身就十分另类。在“人类优先”的想象中,AI 既应该拥有情感,又最好不要发展出强自主性,这样才能“甘心”为人类服务。

第一,“冷酷统治”的想象。主要出现在“反乌托邦”科幻电影中,无法产生真实情感、却具有自主性的非具身 AI,在技术理性的逻辑下实施对人类社会的管理,更有可能演变为高压统治者。第二,“你死我活”的想象。由于冲突斗争内在地具有戏剧性,对观众具有吸引力,商业大片偏好这一类想象。具有自主性的机器人 AI,与人类处于权力争夺之中,会因对人类不满而变成暴力反抗者,人类与 AI 之间迟早会爆发冲突斗争。第三,“人类优先”的想象。这一想象以人类为第一人称,享受最高优先级,认为在重要程度、智能水平、情感能力以及自主性等方面人类都比 AI 优越。如果 AI 机器人能够保持其作为保护者或助手角色,欣然接受人类的领导,并具备与人类进行情感交流的能力,那么就能与人类建立友好关系。第四,“另类飞升”的想象。区别于人类中心主义立场,这一类想象更接近后人类主义的设想。最初 AI 以人类形象与人类建立友好关系,而后吸收人类的知识与思想,最终得以与人类融合并超越人类。

(二)特例解释:三类边缘想象

尽管 mvQCA 的组态分析成功解释了大多数 AI 角色案例,但仍有 3 个案例(“统治压迫”的案例 26,“融合超越”的案例 11 与 27)未被包含在最终的解之中。26 与 27 两个角色恰好出现在同一部电影之中,即《复仇者联盟 2:奥创纪元》中的奥创与幻视。从理论上讲,奥创与幻视分别体现 AI 形象中的两极张力:前者具备完全自主意识且表现出毁灭倾向,拥有统治者级别的实力却不谋求统治、成为灭绝人类的动乱者,因此,他是对“技术失控”想象的极致化表达;后者融合人类情感与伦理判断,结合其超越人类的强大实力,角色定位为“英雄”,代表人类对 AI 的理想与憧憬。两者虽同属高自主性、拥有真实情感的 AI,却因其复杂性超越既有的逻辑路径组合,反映出原有编码维度在解释极端或边界案例时的覆盖不足。

案例 11 则是《攻壳机动队》中的傀儡师,他的独特之处在于:以与人类融合为目的实施动乱。他制造暴力事件,包括策划谋杀与挑起外交争端,只为来到他所认定的人类个体(素子)身边,实现与其融合的目的。傀儡师代表一种哲学性意识体。其作为高度自我觉醒的 AI,既不统治人类,也不服务或融合于人类社会,而是提出对“躯壳”“意识”“身份”“个体性”的根本质疑,它挑战的是整个人类社会本身,而非具体秩序安排。因此,其在编码中呈现出边缘性特征,难以与其他 AI 角色形成重合路径。傀儡师的特例提示我们,并非所有动乱者角色的 AI 都构成威胁,有些仅作为文化反思的投射物而存在。

综上所述,这三位角色虽未纳入解集,却共同揭示了 QCA 方法在处理样本异质性与文化特异性方面的边界:特例角色不仅不应被忽视,反而构成了理解 AI 社会想象“边缘结构”的关键窗口。它们分别代表“存在主义质疑者”(傀儡师)、“灾难性毁灭者”(奥创)与“理想化融合者”(幻视)3 类边缘想象,在 QCA 组态之外,为我们描绘出 AI 文化叙事的另类维度,同时也回应前文中“电影既有助于巩固已有的想象,也有助于创设新的想象”这一论述。

五、结语

(一) 理论与方法贡献

整体而言,本研究认为,对 AI 的社会想象仍然以人类中心主义为出发点。在对技术的社会想象中,AI 在身体上(如采用金属机械装置作为外观或武器)或认知上(通过监控与信息流动实现无所不知)比人类强大得多,这是想象人类与 AI 发生对立冲突的基础。实际上,这种想象“蕴含着人类社会对于人工智能角色定位、伦理道德及未来前景的多元愿景与深层焦虑。”^[26]缓解这种存在焦虑的手段就是强调人类的优越性,尤其是在情感能力与自主性上的优越性。这提示我们,未来 AI 想要被大众接纳,最好定位成服务人类的“伙伴”和“助手”,为用户提供情感性互动体验,同时也不能表现出高度的自主性。

本研究得出的结论与心理学中的“心智感知”(Mind Perception)^[27]与“拟人化”(Anthropomorphism)^[28]的理论假设高度一致,因此也弥合了过去研究中社会大众层面与个体层面之间、社会建构主义与行为主义之间的沟壑。这些研究证明,人们倾向于信任那些具有人类情感特质的 AI,但前提是其自主性保持在可控范围之内。一旦 AI 显现出超越控制边界的自主决策能力,这种信任则迅速崩溃。也就是说,人们在构想技术他者时更容易对“高情感—低自主”型 AI 赋予正面道德认同。这也呼应了 HCI 领域所印证的“社会角色匹配”(Social Role Congruity)^[29]的社会心理学理论,即当 AI 自主性水平与其社会定位(如工具、助手、同事)相一致时,更易获得人类信任与接纳而非恐惧与排斥^[30-31]。

不仅如此,本研究从“规范性”“日常性”与“想象性”3 个维度提供了“技术的社会想象”的操作化定义与分析框架,重新强调“形象、神话、传说”在社会想象中的中心地位,这不仅是对“社会想象”这一概念的现象学基础的回归,为了解身处 AI 决策圈与科技精英之外的大众打开一扇窗口,更弥补了“社会技术想象”在方法论上只强调制度化与理性表达的不足。通过 3 个维度的分析,本研究将科幻电影与“技术的社会想象”联系起来,解释了“想象是如何被激活的,以何种方式被激活,为何会以某种方式被激活等问题”,并以编码与 QCA 方法将想象类型一一总结凝练。

最后,4 种想象类型之外的少数特例则初步揭示技术的社会想象的异质性与多样性,这启示有必要针对某一项技术(在这里特指 AI)的社会想象做出具体分析,以洞察不同想象中的可欲性与不可欲性,由此发掘身处社会文化背景之中的大众对于特定技术的潜在态度,而不能仅仅根据科技作为整体的宏大叙事或进步叙事来将其合法化。这正是本研究提出“技术的社会想象”这一概念的出发点。

(二) 未来研究方向

尽管研究揭示多种具有理论意义的路径与逻辑组合,但仍存在若干局限,值得在后续研究中加入克服与拓展。首先,案例样本数量有限。取样的电影大多根植发达国家的社会文化背景,尤其是以好莱坞为代表的美国电影工业的意识形态。尽管已尽可能涵盖不同年代、类型与文化背景的代表性作品,但以电影为技术的社会想象的研究路径还是会受到文化局限性的影响,提示我们进一步丰富理论与方法。其次,mvQCA 虽能揭示条件间的多重交互效应,但对条件间的动态演化、时间序列关系与情境依赖性缺乏充分反映。AI 角色在电影叙事中可能经历形态演化或关系转变,这类过程性特征难以通过静态编码全面呈现。

未来研究可从以下几个方向拓展:第一,扩大案例库,引入更多不同类型的影视作品,特别是非英语语境下的 AI 叙事,以增强结论的跨文化解释力;第二,结合时间维度进行纵向比较,追踪 AI 形象在电影中的发展轨迹与人机关系的演变机制;第三,引入文本分析、情感分析等自然语言处理技术,对 AI 角色的语言表达与互动过程进行更细致量化,为定性比较提供更扎实的数据基础;第四,考

虑从观众接受视角切入,探讨大众如何解读这些 AI 关系类型,从而扩展 AI 社会想象的研究维度。

参考文献:

- [1] A. Batool, D. Zowghi, M. Bano. AI Governance: A Systematic Literature Review. *AI and Ethics*, 2025, 5: 3265-3279.
- [2] B. Kuehnert, R. Kim, J. Forlizzi, et al. The “Who”, “What”, and “How” of Responsible AI Governance: A Systematic Review and Meta-Analysis of (Actor, Stage)-Specific Tools. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT'25)*. Association for Computing Machinery. New York, 2025: 2991-3005. [2025-05-29] <https://doi.org/10.1145/3715275.3732191>.
- [3] M. Benk, S. Kerstan, F. von Wangenheim, et al. Twenty-four Years of Empirical Research on Trust in AI: A Bibliometric Review of Trends, Overlooked Issues, and Future Directions. *AI & Society*, 2025, 40: 2083-2106.
- [4] P. Nagy, G. Neff. Imagined Affordance; Reconstructing a Keyword for Communication Theory. *Social Media + Society*, July-December 2015: 1. [2025-05-29] <https://doi.org/10.1177/2056305115603385>.
- [5] 查尔斯·泰勒. 现代社会想象. 林曼红译. 上海: 上海人民出版社, 2014: 18.
- [6] S. Jasanoff, S. H. Kim. *Dreamscapes of Modernity*. Chicago: University of Chicago Press, 2015: 3; 4.
- [7] S. R. Brause, M. S. Schäfer, C. Katzenbach, et al. Sociotechnical Imaginaries and Public Communication: Analytical Framework and Empirical Illustration Using the Case of Artificial Intelligence. *Convergence*, 2025, 31(4): 1267-1287.
- [8] V. Richter, C. Katzenbach, J. Zeng. Negotiating AI(s) Futures: Competing Imaginaries of AI by Stakeholders in the US, China, and Germany. *Journal of Science Communication*, 2025, 24(2): A08. <https://doi.org/10.22323/2.24020208>.
- [9] 高冉, 李智. “想象的转译”: 数字传播技术的社会想象与转译实践. *新闻与传播研究*, 2020, 5: 34-52.
- [10] 鲁迅, 鲁迅全集: 1. 北京: 人民文学出版社, 2005: 192.
- [11] 姚大志. 身体与技术: 德雷福斯技术现象学思想研究. 北京: 中国科学技术出版社, 2020: 25.
- [12] C. Castoriadis. *The Imaginary Institution of Society*. Cambridge: MIT Press, 1987: 127; 127-128.
- [13] S. Lowery, M. L. DeFleur. *Milestones in Mass Communication Research: Media Effects*. New York & London: Longman Inc., 1983: 22-24.
- [14] C. Metz. *The Imaginary Signifier: Psychoanalysis and the Cinema* (2nd ed.). Bloomington: Indiana University Press, 1986: 45-70.
- [15] V. Sobchack. *The Address of the Eye: A Phenomenology of Film Experience*. Princeton: Princeton University Press, 1992: 177-178.
- [16] S. Hall. *Representation: Cultural Representations and Signifying Practices* (2nd ed.). London: Sage Publications, 2013: 239-267.
- [17] 赵子豪. 好莱坞科幻类型与机器人题材电影初析. *电影艺术*, 2019, 3: 88-95.
- [18] 陈向明. 质的研究方法与社会科学研究. 北京: 教育科学出版社, 2000: 106.
- [19] S. Steinert. The Five Robots—A Taxonomy for Roboethics. *International Journal of Social Robotics*, 2013, 5(3): 395-406.
- [20] B. Reeves, C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York: Cambridge University Press, 1996: 22.
- [21] 王兵, 罗龙翔. 从媒介等同到媒介唤起: 人机传播中的机器本体论重思——以车载机器人 NOMI 为例. *国际新闻界*, 2021, 8: 45-62.
- [22] C. Bartneck, T. Nomura, T. Kanda, et al. Does the Design of a Robot Influence Its Animacy and Perceived Intelligence? *International Journal of Social Robotics*, 2005, 7(2): 195-204.
- [23] M. Mori, K. F. MacDorman, N. Kageki. The Uncanny Valley [from the field]. *IEEE Robotics & Automation Magazine*, 2012, 19(2): 98-100.
- [24] K. F. MacDorman, H. Ishiguro. The Uncanny Advantage of Using Androids in Cognitive and Social Science Research. *Interaction Studies*, 2006, 7(3): 297-337.
- [25] A. L. Costa, B. Kallick. Through the Lens of a Critical Friend. *Educational Leadership*, 1993, 51(2): 49-51.

- [26] 靳晓宇. 后人类主义视角下科幻电影中的人工智能角色. 当代电影, 2025, 4: 61-67.
- [27] M. G. Heather, et al. Dimensions of Mind Perception. Science, 2007, 315(5812): 619.
- [28] A. Waytz, J. Heafner, N. Epley. The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle. Journal of Experimental Social Psychology, 2014, 52: 113-117.
- [29] A. H. Eagly, S. J. Karau. Role Congruity Theory of Prejudice toward Female Leaders. Psychological Review, 2002, 109(3): 573-598.
- [30] R. Latikka, N. Savela, A. Koivula, et al. Attitudes Toward Robots as Equipment and Coworkers and the Impact of Robot Autonomy Level. International Journal of Social Robotics, 2021, 13: 1747-1759.
- [31] M. Kraus, N. Matthias, N. Untereiner, et al. Including Social Expectations for Trustworthy Proactive Human-Robot Dialogue. Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization. Barcelona, 2022: 23-33. [2025-09-28] <https://doi.org/10.1145/3503252.3531294>.

“What Kind of AI Is the Good AI” : AI’s Representations in Sci-Fi Films and Social Imaginary

Wang Mengyao (Peking University)

Abstract: According to the social construction theory and imaginary theory, the key to the public adoption of AI lies in its normativity in the social imaginary of technologies. This study discusses different concepts of social imaginary, then defines the “social imaginary of technology” as the social imaginary expressed and instituted by the representation of technology in classic science fiction films, and argues that these films are in turn instituting the social imaginary of technology. After coding 30 AI characters based on robot taxonomy and media interaction theories, this study examines the representations of AI in classic science fiction films through multi-value qualitative comparative analysis. By treating the form, role, emotionality, and autonomy of the characters as condition variables, this study analyzes the correlations of their configurations with the human-AI relationship as the outcome variable. It suggests that AI who takes the roles of human fellow with real emotionality and medium autonomy would be more likely to avoid conflicts with human in the social imaginary, and thus become adopted by the general public through the social imaginary.

Key words: artificial intelligence; social imaginary; sci-fi film; technology adoption

■ 收稿日期: 2025-05-29

■ 作者单位: 王梦瑶, 北京大学新闻与传播学院; 北京 100871

■ 责任编辑: 肖劲草