



人机共同体伦理对齐设计初探

阮凡¹, 潘恩荣^{2*}

1. 浙江工业大学 马克思主义学院, 杭州 310023;

2. 中山大学 哲学系(珠海), 珠海 519082

摘要: 伦理先行是数字技术发展的必要条件, 也是数字经济新质生产力发展的新方向。作为新兴数字经济业态, 直播经济形成了一个独特的“网络主播+数字技术平台”模式, 构成一种新的数字经济“人机共同体”。“伦理对齐设计”指的是网络主播从销售者转向生产者, 基于自身的知识技能和伦理素养赋予商品新的伦理价值。“人工智能+生产性劳动者+伦理对齐设计”是新兴数字技术和数字经济的伦理先行发展方向。“人机共同体伦理对齐设计”为未来人工智能对齐研究提供了中国式现代化的视角和进路。

关键词: 伦理先行; 人机共同体; 伦理对齐设计; 新质生产力; 董宇辉现象

中图分类号: N03; B829

文献标识码: A

文章编号: 1674-4969(2025)02-0222-12

引言

伦理先行是国家人工智能战略的核心思想之一。随着大数据技术和人工智能技术融入数字技术体系和数字经济业态, 伦理先行同样成为了数字经济驱动新质生产力发展的核心理念之一。从数字经济角度看, 数字技术平台应该发挥其数字技术优势, 坚持伦理先行, 进而“积极培育新能源、新材料、先进制造、电

子信息等战略性新兴产业, 积极培育未来产业, 加快形成新质生产力, 增强发展新动能”^[1]。那么, 数字经济如何坚持伦理先行发展新兴产业的新质生产力? 本研究引入了“伦理对齐设计”(ethical aligned design)的理念, 以数字经济中的“董宇辉现象”为具体案例, 探究网络主播是否能够成为新质劳动者, 并为商品赋予新的伦理价值。同时分析网络主播如何凭借伦理先行的思维, 运用劳动资料(即数字技术平

收稿日期: 2024-11-27; **修回日期:** 2025-03-26

基金项目: 国家社会科学基金一般项目: 技术哲学社会人物研究(19BZX027); 浙江省社会科学规划项目后期资助重点项目: 马克思主义批判思想及其在智能时代的应用研究(SKY-ZX-20230076); 浙江省软科学研究计划项目: 以人工智能为代表的新一代信息技术科研机构科技伦理治理研究(2024C35099); 中央高校基本科研业务费

作者简介: 阮凡(1990—), 女, 博士研究生, 讲师。研究方向为数字经济与技术伦理。

*潘恩荣(1979—), 男, 博士研究生, 教授, 博士生导师。研究方向为工程设计哲学与设计伦理。E-mail: pan.en.rong@qq.com (通讯作者)

引用格式: 阮凡, 潘恩荣. 人机共同体伦理对齐设计初探[J]. 工程研究——跨学科视野中的工程, 2025, 17(2): 222-233. DOI: 10.3724/j.issn.1674-4969.20250031. CSTR: 32282.14.JES.20250031

Ruan F, Pan E R. Preliminary Study on Ethical Alignment Design of Man-machine Community[J]. Journal of Engineering Studies, 2025, 17(2): 222-233. DOI: 10.3724/j.issn.1674-4969.20250031. CSTR: 32282.14.JES.20250031

台), 改造劳动对象(如橱窗商品), 从而实现劳动要素的优化组合与质变。本文将通过深入探讨伦理要素在数字经济催生新质生产力过程中的作用机制, 归纳数字经济通过“伦理对齐设计”推动发展新质生产力发展的策略。

1 伦理先行: 网络主播伦理赋能的新质生产力效应

伦理先行是数字经济大规模使用大数据技术和人工智能技术的必要条件, 进而“伦理要素”是数字经济发展新质生产力的新要素。自习近平总书记提出“新质生产力”^[1]概念以来, 学界对其进行了多角度、深层次的研究, 其中一个关注点便是“新”和“质”如何体现。韩文龙^[2]认为新质生产力是数字经济时代先进生产力的具体实现形式, 新质生产力的“新”体现为要素构成新和具体表现新, “质”体现为新本质、高质量、高品质、“质”优势。张林忆和黄志高^[3]、张晋铭和徐艳玲^[4]、李政和廖晓东^[5]认为数据和数字技术是新质生产力的要素。但是, 当前的分析中鲜少讨论“伦理”是否能作为生产力新要素并组合出新质产品或服务。

在人工智能时代, 笔者认为, 新质生产力的“新”体现之一是“人机共同体”, 后者是推动生产力高质量发展的新要素; “质”不仅体现为传统生产的“高质量”产品或服务, 还体现为人工智能生产体系的“高伦理”产品或服务, 以应对以人工智能为代表的新一轮科技革命和产业变革引发的伦理挑战。

“伦理对齐设计”思想为数字经济伦理先行发展提供了一种分析框架。“伦理对齐设计”是由电气与电子工程师协会(Institute of Electrical and Electronics Engineers, IEEE)提出的一种技术开发(尤其是人工智能和自主系统)方法论, 强调将伦理原则和价值观系统性地“嵌入”设计、开发及部署全过程, 确保技术发展符合人类利益、社会价值及道德规范, 避免潜在风险与危害^[6]。在数字经济中, 数字平台企业成为新质生产力形成的关键创新主体^[7], 劳动者则在新质生产力发展过程中处于首要

地位^[2], 通过劳动者赋能各要素实现结构性升级和质的提升。后文将以数字经济中“网络主播”与直播经济的案例, 分析“伦理先行”对直播经济效果的影响; 然后, 分析网络主播将伦理价值“嵌入”商品的过程与影响。

网络主播是“通过互联网提供网络表演、视听节目服务的主播人员, 包括在网络平台直播、与用户进行实时交流互动、以上传音视频节目形式发声出镜的人员”^[8]。据《中国网络视听发展研究报告(2024)》^[9], 截至2023年底, 我国职业网络主播数量已达到1508万人。短视频直播对消费的拉动效应显著, 超过七成的用户因观看短视频或直播而购买商品, 超过四成的用户更是认为短视频或直播已成为其主要消费渠道。网络视听市场规模已超过1.15万亿元, 逐渐成为激活数字经济新质生产力的关键引擎。在此背景下, 网络主播在数字经济中扮演的角色愈发重要。

关于数字技术平台和网络主播对新质生产力的影响, 从传统经济角度看, 网络主播是商品流通阶段的“售货员”而非生产阶段的“劳动者”, 他们只是转移商品的价值而不是创造新价值, 因而并不属于生产力范围。但是, 从新质生产力发展角度看, 有必要在数字经济领域培养具有“新素质”的新型劳动者。“要坚持以自主创新为经济发展‘新引擎’, 以智能数字为经济发展‘新介质’, 以新兴产业未来产业为‘新阵地’, 培养创新型人才急需的‘新素质’。”^[10]钱贵明等^[7]研究了数字平台视角下新质生产力的形成机制与推进策略, 但他们却把数字平台上最为典型的网络主播群体排除在劳动者视域之外。他们认为“数字平台视角下, 数字化劳动者包括新型的平台建构者、双边用户、资源互补商等类型。”尽管网络主播在数字经济领域发挥关键性角色, 但是在理论研究者眼里却“消失”了。

然而, 在数字经济发展过程中, 公众关于网络主播的身份解读引发了社会争论。例如, 东方甄选“小作文”事件及其背后的“董宇辉现象”显示, 网络主播除了“售货员”身份外, 还有一种“具有伦理关怀的知识生产劳动者”身份。

东方甄选“小作文”事件, 指的是数字经济技术

平台（后文简称“直播平台”）东方甄选与网络主播董宇辉之间，就“吉林之行”宣传文案的创作归属问题所产生的争论^[1]。2023年12月5日，董宇辉朗读了一段约700字的文案（网络称之为“小作文”），为12月8日至10日东方甄选“吉林之行”专场直播做宣传，获得网友广泛赞誉。次日，东方甄选内容团队在其抖音评论区置顶了一条评论，称直播现场用于开场或转场的“小作文”文案由主播、文案团队、拍摄团队、剪辑团队共同完成。这条被特意置顶的评论激起了董宇辉“粉丝”的强烈不满，他们视其为直播平台对主播董宇辉的“背刺”，认为东方甄选此举意在“卸磨杀驴”，削弱董宇辉的影响力。由此，网络空间内“小作文”风波骤起，关于其撰写者的猜测和争论四起，由此引发了一系列连锁反应：董宇辉的“粉丝”转而前往东方甄选的竞争对手直播间消费，以此表达不满；同时，直播经济圈内的各路主播和网络投资人，无论明里暗里，均纷纷卷入这场争论。这一系列事件，最终导致东方甄选的“粉丝”数量锐减，股价也大幅下挫。面对此境，东方甄选迅速采取行动，包括处置相关人员、安排俞敏洪与董宇辉共同直播以回应网友关切、晋升董宇辉为东方甄选的高级合伙人，并为其另辟独立工作室及直播间等，逐步平息了此次风波。

“小作文”事件显示，传统经济思维已经难以解释和应对直播平台及其优秀的网络主播（网络称之为“网红主播”）现象。从传统经济角度看，网红主播只是商品“售货员”而非“生产者”。一方面，直播平台享受由网红主播带来的经济效益；另一方面，直播平台又竭力避免对特定主播的过度依赖。根据澎湃新闻报道，东方甄选实控人俞敏洪表示，如果只是依赖董宇辉一个人，那东方甄选就不是一个正常的公司^[12]。换言之，东方甄选应选择产品路线而非网红路线。这一抉择关乎其在资本市场的估值水平等关键因素。倘若采取网红路线，东方甄选虽作为电子商务平台，本质属商贸公司范畴，资本周转效率虽高，但长期成长空间受限，其资本市场估值或将停留于十亿量级；反之，若走产品路线，东方甄选则定位为数字科技平台公司，展现出高成长潜力、高投入需求及伴随的高风险特征，其资本市场的估值将是百亿量级。

因此，对于上市公司东方甄选而言，避免过度依赖网红主播是必然的选择。网络争论也盛传，如果不是大老板默许，东方甄选不会出现在公开场合以小作文“打压”网红主播董宇辉的事情。“打压”的目的是解构网红主播个体与直播平台销售额之间的高度相关性，把网红主播降级为可批量培训的普通网络主播，进而降低直播平台对特定主播的依赖。

然而，“粉丝”对网络主播董宇辉的认知超越了普通网红主播的范畴。“小作文”由谁撰写的争论之所以能引发如此大的影响，是因为传统的商品销售者定位与“粉丝”对网络主播董宇辉的定位之间存在巨大差距，且前者严重违背“粉丝”的意愿。当时的网络盛传，“董宇辉”已经是“现象级”形象，是百万粉丝级网红主播中的一股“清流”。一方面，董宇辉的直播具有明显的“伦理先行”性质。相比较于普通主播各种嘶吼、各种煽动性“买它”“上一号链接”等话术，董宇辉的出口成章、中英文知识讲解和“谷贱伤农”等富有人文气质的“现场解说”更容易与消费者达成购买共识。因此，董宇辉的“现场解说”是一种伦理赋能行动，赋予商品新的伦理价值，在商业买卖过程中触动了消费者心中的柔软之处。《21世纪经济报道》认为，董宇辉“表现出的腹有诗书气自华，是娓娓道来的精彩讲述能力，是真诚到阿姨们不由自主把他当女婿看待。”^[13]另一方面，董宇辉的直播具有“伦理关怀”正外部性效应。相比较于一般网红主播掩盖不住的“铜臭味”，董宇辉的“现场解说”更倾向于介绍商品相关知识和发扬传统文化。例如，董宇辉与诺贝尔文学奖得主阿卜杜勒拉克·古尔纳（Abdulrazak Gurnah）的文学对话，以及他针对地方特色所做的文化旅游描绘，不仅实现了商业营销上的成功，更意味着数字时代优秀传统文化的成功传播。

从现象层面看，消费者“买单付费”的行为明面上是指向董宇辉本人，即传统意义上的“网红主播”。但从新质生产力层面看，根据“粉丝”关于“为知识付费”等热门评论，笔者认为，上述行为最终指向董宇辉在直播间里生产的数字商品——具有伦理关怀性质的“小作文”。因此，在“粉丝”眼里，董宇辉就是基于数字平台生产数字商品的新质劳动者，他赋予橱窗商品“更高的价值”。

综上所述,“粉丝”对董宇辉的身份认知已经超越了传统经济的商品“售货员”,而是新兴产业直播经济领域从事“更高价值商品”生产的新质劳动者。由此,可以进一步以“董宇辉现象”为例探讨“伦理对齐设计”的路径。

2 新质劳动者:将伦理价值嵌入商品

直播经济能否孕育发展新质生产力的核心,在于探讨“网络主播是否属于生产‘更高价值商品’的新质劳动者”。那么,何谓“新质劳动者”?这需从新质生产力的定义中寻根溯源。“新质生产力是由技术革命性突破、生产要素创新性配置、产业深度转型升级而催生的当代先进生产力,它以劳动者、劳动资料、劳动对象及其优化组合的质变为基本内涵,以全要素生产率提升为核心标志。”^[14]一般认为,劳动者、劳动资料、劳动对象是马克思主义政治经济学生产力理论的三要素。在数字经济背景下,直播经济新质生产力指的是劳动者基于新型数字化等劳动资料作用于直播商品,通过重新“设计”劳动者、劳动资料和劳动对象之间的关系提升全要素生产率,增加单位直播商品的价值量或使用价值。

“新质劳动者”就是直播经济中能够赋予商品“更高价值”的劳动者,他们的“劳动”能够创造直播商品的新价值或新使用价值。于是,“是否创造新价值和(或)新使用价值”将成为董宇辉与一般网红主播的实质性区别。那么,作为新质劳动者,网络主播创造什么“新价值”或“新使用价值”?

从传统经济角度看,网络主播“董宇辉们”与一般网红主播的身份一样,他们都是商业劳动者,即直播经济“售货员”。网络主播的本职工作是促成买卖双方之间的交易,通过“营销行为”吸引消费者注意进而推销商品。董宇辉在东方甄选直播平台的首要活动是向观看者展示并推销橱窗商品,争取实现橱窗商品由商品(W)到货币(G)的形态转变。“形态变化W-G和G-W,是买者和卖者之间进行的交易……状态的变化花费时间和劳动力,但不是为了创造价值,而是为了使价值由一种形式转化为另一种形式。”^[15]因此,在传统经济思维中,售货员是商业劳

动者,但其与生产商品的劳动者之间存在着本质区别,即后者有能力创造商品的新价值,而前者并不创造新价值只是转移商品原有的价值。董宇辉虽然有出色的销售能力和流量优势完成了大额交易量,但这并不改变其售货员的身份。在资本和平台公司面前,他只是从事“不创造新价值”的售货员。

然而,从新质生产力的视角来看,网络主播董宇辉还兼具另一重身份——生产数字商品的新质劳动者。“新质生产力形成和发展于生产力要素的系统性质变,在理论上可归纳为劳动者在科学技术驱动下更大规模、更高效地结合起来认识、利用、改造自然进而创造满足人民美好生活需要的物质、精神财富的能力”^[16]。董宇辉在直播间生产的数字商品是“现场解说”的“小作文”,它契合了人们对美好精神生活的向往。需着重指出的是,核心所在并非单纯的“小作文”,而是董宇辉本人的“现场解说”。即便是相同的“小作文”,经由一般网红主播的阐述,亦难以企及董宇辉解说所带来的同等魅力。更进一步说,即便利用现今先进的大模型技术塑造另一个“数字人”或“智能体”(AI Agent)版的董宇辉,也未必能再现董宇辉本人解说的独特韵味。这构成了董宇辉与一般网红主播之间的本质分野,也是董宇辉作为网红主播界“清流”的标志。

作为新质劳动者,董宇辉的“现场解说”具有创造“伦理价值”的劳动性质。在《政治经济学批判(1861—1863年手稿)》中,马克思曾对生产劳动的特征如此描述:“从资本主义生产的意义上说,生产劳动是雇佣劳动,它同资本的可变部分(花在工资上的那部分资本)相交换,不仅把这部分资本(也就是自己劳动能力的价值)再生产出来,而且,除此之外,还为资本家生产剩余价值。”^{[17]213}。在《资本论》第一卷资本的生产过程中,马克思用劳动二重性分析商品,即劳动具有创造价值和使用价值的双重属性。“一切劳动,一方面是人类劳动力在生理学意义上的耗费;就相同的或抽象的人类劳动这个属性来说,它形成商品价值。一切劳动,另一方面是人类劳动力在特殊的有一定目的的形式上的耗费;就具体的有用的劳动这个属性来说,它生产使用价值。”^{[18]60}概括起来,生产性劳动具有三个特征,即雇佣关系、

创造新价值和创造新使用价值。董宇辉的“现场解说”满足了以上三个特征。

首先,董宇辉与东方甄选直播平台之间存在雇佣关系。雇佣关系的核心在于劳动力是否出卖自身以获取生活所必需的工资。在《雇佣劳动与资本》中,马克思指出,“劳动力并不向来就是商品。劳动并不向来就是雇佣劳动”^{[19]716},只有当“工人是以出卖劳动力为其收入的唯一来源”^{[19]717}时,劳动力才成为雇佣劳动者。董宇辉从新东方的英语教师,转型为东方甄选直播平台的网络主播,再到新东方文旅集团副总裁,一直是隶属于新东方集团的员工。他在这些不同岗位上,通过付出个人劳动力获取报酬,这表明了其与企业存在的雇佣关系。

其次,董宇辉的“现场解说”创造橱窗商品的新价值。传统经济认为,董宇辉与一般网络主播在直播期间的工作属于商业劳动,不创造新价值但进行价值形式转化^{[11]147}。那么,董宇辉是否能够创造橱窗商品的新价值?马克思强调,劳动力能够作为商品的价值就在于它独有的创造商品价值的价值^{[17]213}。原因在于,劳动力具有“专属于人的那种劳动形式”。

如果董宇辉的“现场解说”满足“专属于人的那种劳动形式”,那么,董宇辉可以作为能够创造商品价值的劳动力,而非商业“售货员”的劳动力。在《资本论》中,马克思详细论述了这种专属于人的劳动形式。“我们要考察的是专属于人的那种形式的劳动。……最整脚的建筑师从一开始就比最灵巧的蜜蜂高明的地方,是他在用蜂蜡建筑蜂房以前,已经在自己的头脑中把它建成了。……这个目的是他所知道的,是作为规律决定着他的活动的方式和方法的,他必须使他的意志服从这个目的。但是这种服从不是孤立的行为。除了从事劳动的那些器官紧张之外,在整个劳动时间内还需要有作为注意力表现出来的有目的的意志。”^{[18]208}

“专属于人的那种劳动形式”有三个性质:目的导向、认知转换和注意力聚焦。首先,目的导向指的是劳动主体为达到某种使用价值而进行的有意识地创造活动。董宇辉是劳动主体,他“现场解说”的“小作文”是在直播带货目的下进行的有意识的创造活动,富有“伦理关怀”风格,而且“这个目的是他所

知道的,是作为规律决定着他的活动的方式和方法的”^{[18]208}。其次,认知转换指的是劳动主体作用于对象客体以形成劳动产品的自然力消耗过程,使自然物的形式发生变化的过程。董宇辉的劳动——现场解说——以直播带货过程中的橱窗商品为作用对象,使得观看者获得关于商品的新的理解——即关于商品的认知出现了新的形式。当直播观看者选择购买橱窗商品成为消费者,董宇辉在商品的新认知形式中实现了自己的目的。最后,注意力聚焦指的是劳动主体的实践以有目的的意志为聚集指向。董宇辉在直播间的镜头下保持高度注意力,才思敏捷、出口成章。同时,他的现象级“清流”状态彰显了他有别于普通网红主播的有目的的意志,流露出深厚的伦理关怀。换言之,董宇辉的“现场解说”为商品赋予了“内在”的伦理价值,这与马克思所述的手工劳动者创造价值的劳动形式相契合,从而使得董宇辉能够被视为创造新伦理价值的劳动力。

接着探讨董宇辉的“现场解说”如何创造橱窗商品的新伦理价值。在《资本论》中,马克思认为,“价值”是在商品生产上耗费的人类劳动力及积累的人类劳动^{[18]51}。商品的价值量可以用“社会必要劳动时间”计量“凝结在商品上的无差别的人类劳动”。社会必要劳动时间是“在现有的社会正常的生产条件下,在社会平均的劳动熟练程度和劳动强度下制造某种使用价值所需要的劳动时间”^{[18]52}。从劳动力消耗角度看,董宇辉创作的数字商品——“现场解说”——凝聚了董宇辉劳动,具备商品的价值属性。从社会必要劳动时间角度看,董宇辉所创作的数字商品,凝聚了他过往受教育与生活经历中积累的知识技能及伦理素养所耗费的劳动时间。换言之,董宇辉的“现场解说”创造了橱窗商品的新价值。

最后,董宇辉的“现场解说”创造了商品的新使用价值。使用价值是商品的另一重属性,即物的有用性。“物的有用性使物成为使用价值……使用价值只是在使用或消费中得到实现……使用价值同时又是交换价值的物质承担者”^{[18]48-49}。董宇辉的“现场解说”为观看者带来了信息传递、知识普及、情感共鸣、文化享受等体验,这是对橱窗商品的深度生产,使消费者体验到原来橱窗商品更多元的使用价值。以

带货华为产品为例,董宇辉援引了中华民族的神话传说,生动诠释了华为的一系列产品:鸿蒙系统、鲲鹏计算平台、昆仑玻璃以及玄武架构,使消费者对这些原本冷冰冰的科技产品产生了自然而然的文化认同感。这正是董宇辉的“现场解说”为原有商品增加的新使用价值。

综上所述,董宇辉在直播带货平台上的“现场解说”创造了橱窗商品的新伦理价值和新使用价值。相比于传统经济思维中的商业劳动者——“售货员”,或者产业劳动者——工人,董宇辉是“新质劳动者”,且是具有“伦理关怀”风格的劳动者。他的“现场解说”是借助劳动资料——东方甄选数字技术直播平台、自身的知识技能和伦理素养,针对劳动对象——直播间售卖的橱窗商品——进行的二次劳动。这提升了橱窗商品的价值,消费者也获得了商品额外的使用价值——背景知识、传统文化和伦理关怀等。

3 伦理要素与新质生产力生成机制

新质生产力以劳动者、劳动资料、劳动对象及其优化组合的质变为基本内涵^[10]。那么,新质生产力的生成机制是董宇辉通过优化组合劳动资料并创造劳动对象(即橱窗商品)来赋予其新伦理价值或新使用价值的过程。

一般认为,劳动者优化组合劳动资料的过程属于“创新”。“创新之父”熊彼特(Joseph Alois Schumpeter)强调,他的创新研究只是马克思研究领域的一小部分^[20]。因此,回到马克思《资本论》关于手工业、工场手工业和机器大工业的论述,作为数字技术的一种新兴产业形态,直播经济的新质生产力的特征逐渐清晰。

第一,直播经济处于手工业阶段,新质生产力发展主要依靠网络主播的个人能力,存在初步的协作分工。回顾整个“小作文”事件,东方甄选直播平台在很大程度上依赖于董宇辉个人能力所带来的网红效应。董宇辉的“现场解说”中的“小作文”主要通过三种方式创作而成:一是完全由董宇辉本人独立创作并决定采用;二是董宇辉提供整体的思路和创意,再

由团队协作共同完成;三是小编团队自主进行“小作文”的创作,最终由董宇辉决定选用^[7]。上述三种方式具有典型的手工业风格。一方面,董宇辉犹如一位行会师傅,拥有独立完成“小作文”及其“现场解说”的全套技能;另一方面,文案团队、拍摄团队、剪辑团队等构成了他的坚实后盾,为其提供了必要的辅助。通过“小作文”事件,董宇辉从一名“学徒”晋升为“行会师傅”,实现了从东方甄选平台的普通网络主播到高级合伙人的晋升,并独立开设一个直播间——“与辉同行”,一个新“作坊”。

第二,直播经济的新质劳动者是数字手工业者,生产数字商品主要依靠个人伦理素养和知识技能。“与辉同行”已经超越个人作坊,开始吸纳新的网络主播加盟,类似于直播经济的工场手工业形态。基于新质生产力的精神,理想的直播经济只有是机器大工业的形态时,才能满足“技术革命性突破、生产要素创新性配置、产业深度转型升级”的特征。在《资本论》中,马克思把工场手工业作为机器大工业的初始阶段进行研究。产生工场手工业的一种方式“不同种的独立手工业的工人在同一个资本家的指挥下联合在一个工场里,产品必须经过这些工人之手才能最后制成。”^{[18]390}这一时期的手工业者与机器大工业时代的劳动者不同之处在于,前者凭借个人技艺使用劳动资料独立完成一件完整的产品,后者则作为分工体系中的机器助手从事某个生产片段。从工场手工业到机器大工业,工业生产力完成了从初级到高级的转变。董宇辉创作数字商品的过程依赖的主要劳动资料是其自身知识技能,具有初步的协作分工。在这过程中,董宇辉是“数字手工业者”,充分发挥了他的知识储备、选择、转化、运用能力,针对消费者群体有目的有选择地转化,进而形成具有新使用价值的数字商品,满足观看者的消费需求。

第三,直播经济的产业形态尚处于产业资本阶段,特征是产销共同体。作为直播领域的新质手艺人,除了依靠自身技能完成产品生产,董宇辉与传统手艺人的区别在于他在生产数字商品的同时完成了数字商品与橱窗商品的销售,即数字商品与橱窗商品的“产销共同体”。董宇辉如同工场手工业时代的手艺人,依靠个人技艺生产数字商品,对带货商品进行进

一步加工从而形成新质商品。同时董宇辉还是数字经济时代直播间网络主播的典型代表,即销售者。在数字商品创作和商品销售的过程中,董宇辉完成了“产销者(生产者和销售者)”的统一,也完成了“伦理嵌入”的机制。

第四,直播经济的新质生产力的生产资料主要是新质劳动者本身的知识体系和伦理素养,其他包括直播的数字技术平台。东方甄选直播平台,作为一个初级的数字技术平台,其功能类似于手工业时代的作坊,扮演着劳动资料的角色。但是,单纯的数字技术平台不足以构成直播经济的全部劳动资料,网络主播的个人能力及其相应的知识体系和伦理素养也是劳动资料。一方面,新质劳动者通过其手艺创作出新质数字商品获取网络的注意力资源。例如,董宇辉通过自嘲“撞脸兵马俑”、双语带货以及精彩的“小作文”现场解说等视频片段,迅速在网络上走红,并吸引了大量关注,由此产生了可观的网络流量。在逐渐展示其创作的数字商品的过程中,董宇辉让观众获得了极高的伦理关怀和精神享受。这个过程是免费享受董宇辉的数字商品的过程,也是快速“涨粉”获得消费者注意力的过程。另一方面,数字商品的价值兑现离不开数字平台。“新质劳动者”的产品通过融合流量实现价值的形式转化。董宇辉的流量变现通过消费者购买直播间商品得以实现,而具体商品的销售基本形成了“数字商品A免费使用+橱窗商品B展示”的公式(图1)。以在短视频平台播放的直播片段为例,在介绍商品前通常会呈现一段“小作文”或故事情节等,以吸引和激活“粉丝”的实时注意力,然后再推出销售的商品,并附上购物链接,引导消费者完成购买。

新质生产力的生成机制是新质劳动者的个人技艺、伦理素养和数字技术平台结合为劳动资料作用于劳动对象实现“质变”的过程。“质变”包括两个方面:数字商品生产与销售。一方面是新质劳动者董宇辉团队完成数字商品的生产和橱窗商品的加工。观看者在直播间面对的是两个产品:数字商品A和橱窗商品B。董宇辉在直播间的直播过程是现场生产数字商品A,并将其与橱窗商品B建立强相关性。这是创造数字商品A的价值,增加橱窗商品B的使用价值的过程(见图1)。整个过程中,董宇辉发挥着独立且完整的作用,表现出生产过程的全面性。这与工场手工业时期的手艺人工作方式具有一致性。

另一方面,作为新质手艺人的董宇辉在直播带货过程中同时融合了生产和销售环节。在传统经济活动中,生产和销售环节是相互独立的过程。在马克思主义政治经济学理论中,“生产-分配-交换-消费”等环节是一个有序流转的过程。这个过程在常规直播带货领域体现为:生产环节既是商品生产的物质环节,也是价值创造的唯一环节;销售环节则是价值形式转换的环节。但是,董宇辉基于“伦理先行”思维对橱窗商品B进行二次“生产”得到数字产品A,同时融合了价值生产的劳动过程和价值形式转换的销售过程,且这两个原本独立的环节由单个主体在同一时间完成,如图1所示。

从价值实现的角度看,董宇辉创作了数字商品A的价值,尤其是伦理价值,“嵌入”到直播间的实体橱窗商品B+中,而使用价值以免费的形式提前完成了与消费者的交换。在实体商品中,使用价值和价值以及主播创作的伦理价值凝结在同一个商品身上,在时间和物质载体上具有不可分割性。但是,数字商品

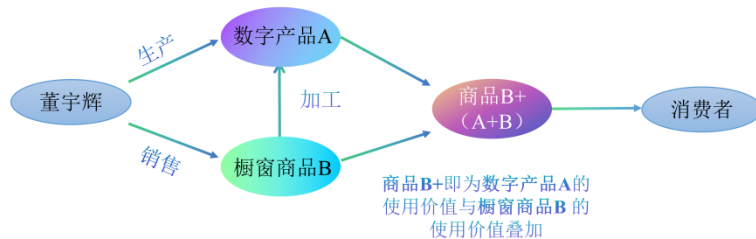


图1 董宇辉数字商品“伦理先行”生产过程

Figure 1 Dong Yuhui's "ethical first" production process for digital products

A作为与橱窗商品B共生的数字商品,它的使用价值让渡、价值与伦理价值转移在时间上具有先后性,在载体上具有独立性。经过董宇辉对数字商品A的“解说”,原来的橱窗商品B额外获得了一道“加工”,这个过程是劳动力的生产过程,凝聚了“网络主播”的脑力劳动。在这个过程中,消费者购买的橱窗商品B已经不再是原来的橱窗商品,它的价值发生了变化,实现了橱窗商品B到商品B+的转变(见图2)。

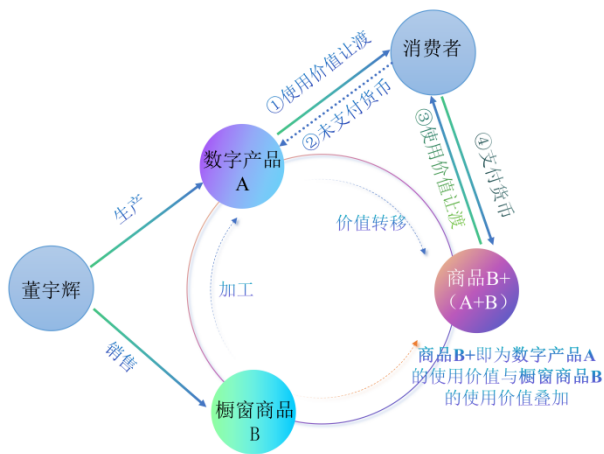


图2 橱窗商品价值增加与兑现图

Figure 2 Value increase and redemption of showcase products

董宇辉数字商品“伦理先行”生产过程能够引领生产方式变革。关于生产方式变革,马克思认为有两条途径^[21]。“一条是生产者成为商人或资本家,指的是技术和资本共同支配生产;另一条是商人直接支配生产,指的是资本完全支配生产。”^[22]前一条方式包含两层含义,即一方面资本是生产方式变革的决定性因素,但不是唯一因素;另一方面,“生产者”的注意力也是决定性因素之一。“伦理先行”思维要求网络主播不同于商人的追求。后者追逐的是最大利润,但不能引领精神产品和服务的生产方式变革;前者则引入“伦理要素”,与其他生产要素组合探索新“质”生产方式。

4 结论

在数字经济蓬勃发展的今天,伦理先行已成为推动新质生产力发展的关键力量。作为新兴数字经济业

态,直播经济形成了一个独特的“网络主播+数字技术平台”模式,构成一种新的数字经济“人机共同体”。“伦理对齐设计”指的是网络主播从销售者转向生产者,基于自身的知识技能和伦理素养赋予商品新的伦理价值。本文引入“伦理对齐设计”理念,揭示了其在提升商品价值、促进高质量消费以及引领生产方式变革中的重要作用。

首先,伦理先行是数字技术发展的必要条件,也是数字经济新质生产力发展的新方向。随着大数据和人工智能技术的深度融合,伦理先行不再仅仅是道德层面的要求,而是成为驱动数字经济创新发展的核心理念。它要求我们在技术发展的同时,充分考虑伦理原则和价值观,确保技术应用的正当性和合理性。

其次,“伦理对齐设计”为数字经济中的新质劳动者提供了行动框架。网络主播作为新质劳动者的代表,通过其自身的知识技能和伦理素养,赋予商品新的伦理价值。这种价值不仅体现在商品本身的物质属性上,更体现在其背后的文化、情感和社会责任上。董宇辉现象就是一个生动的例证,他通过富有文化底蕴和人文气质的“现场解说”,创造了商品的新伦理价值,实现了商业营销和文化传播的双重成功。

进一步地,本文强调了新质劳动者在数字经济中的首要地位。这些新质劳动者具备生产性劳动能力和“伦理对齐设计”能力,能够创造商品的新伦理价值。他们的劳动不仅提升了商品的市场价值,更推动了生产方式的变革。通过引入“伦理要素”,新质劳动者引领了生产过程的革新,为数字经济的未来发展开辟了新的道路。

最后,本文展望了数字经济的未来发展方向。随着人工智能和大模型技术的不断进步,“人工智能+生产性劳动者+伦理对齐设计”将成为新兴数字技术和数字经济业态伦理先行的发展方向。这一方向不仅强调了技术的创新应用,更强调了伦理原则和价值观的引领作用。它要求我们在享受技术带来的便利的同时,不忘伦理责任和社会担当。

综上所述,伦理先行在数字经济中具有不可替代的作用。通过“伦理对齐设计”,可以提升商品的价值,促进高质量消费,引领生产方式变革,并为数字经济的未来发展指明方向。

“人机共同体伦理对齐设计”为未来人工智能对齐 (AI alignment) 研究提供了中国式现代化的视角和进路。“生成式人工智能提出了自启蒙运动以来从未经历过的哲学挑战和实践挑战”^[23]。人工智能对齐是应对上述挑战的重要视角和进路。OpenAI 提出, 对齐研究旨在提高人工智能系统从人类反馈中学习的能力, 并帮助人类评估人工智能, 使通用人工智能 (AGI) 与人类价值观保持一致, 并遵循人类的意图^[24]。那么, 在人工智能研发过程中, “对齐”^[25-26]指的是“将伦理价值嵌入人工智能或数字化技术系统”。然而, 这种对齐思路困难重重, 核心挑战在于“人类群体没有统一的伦理价值”^[27]以至于难以在深

度神经网络中嵌入明确的伦理价值原则。与之不同, 基于数字经济“董宇辉现象”的分析, 可以看到“新质劳动者”作为主体将伦理价值嵌入“商品”生产过程, 即“将伦理价值借道人工智能或数字化技术系统嵌入产品及其生产过程”。因此, 关于未来人工智能对齐研究, 人机共同体伦理对齐设计实践提出了一种新的中国式现代化视角与进路。这是一种面向实践的“知识转向”研究, 有助于进一步探讨“AI Agent”作为主体“如何接受并重现人类能力的知识”^[28], 将伦理价值嵌入“商品”生产过程, 从而能够回应人工智能对齐研究的核心挑战。

参考文献

- [1] 习近平. 发展新质生产力是推动高质量发展的内在要求和重要着力点[J/OL]. 求是, 2024(11): 1. [2024-05-31]. http://www.qsttheory.cn/dukan/qs/2024-05/31/c_1130154174.htm.
Xi J P. Developing new forms of productive forces is the internal requirement and key focus for promoting high-quality development [J/OL]. Qiushi, 2024(11): 1. [2024-05-31]. http://www.qsttheory.cn/dukan/qs/2024-05/31/c_1130154174.htm.
- [2] 韩文龙. 新质生产力的政治经济学阐释[J]. 马克思主义研究, 2024(3): 100-115.
Han W L. Political economy interpretation of new quality productive forces[J]. Studies on Marxism, 2024(3): 100-115.
- [3] 张林忆, 黄志高. 技术、空间与生态: 数字经济赋能新质生产力的逻辑探析[J]. 经济学家, 2024(8): 35-44.
Zhang L Y, Huang Z G. Technology, space and ecology: Exploring the logic of the digital economy to enable new quality productivity[J]. Economist, 2024(8): 35-44.
- [4] 张晋铭, 徐艳玲. 数字经济时代发展新质生产力的内在逻辑与现实启示: 基于马克思生产力二维理论的分析[J]. 兰州学刊, 2024(9): 74-86.
Zhang J M, Xu Y L. The inner logic and realistic inspiration of developing new qualitative productivity in the digital economy era: Analysis based on Marx's two-dimensional theory of productive forces[J]. Lanzhou Academic Journal, 2024(9): 74-86.
- [5] 李政, 廖晓东. 新质生产力理论的生成逻辑、原创价值与实践路径[J]. 江海学刊, 2023(6): 91-98.
Li Z, Liao X D. Generative logic, original value and practical path of the theory of new qualitative productivity[J]. Jianghai Academic Journal, 2023(6): 91-98.
- [6] IEEE. Ethically Aligned Design: A vision for prioritizing human well-being with autonomous and intelligent systems[R]. The IEEE Global Initiative, 2019.
- [7] 钱贵明, 阳镇, 陈劲. 数字平台视角下新质生产力的形成机制与推进策略[J]. 西安交通大学学报(社会科学版), 2024, 44(5): 15-26.
Qian G M, Yang Z, Chen J. The formation mechanism and promotion strategy of new quality productive forces from the perspective of digital platforms[J]. Journal of Xi'an Jiaotong University (Social Sciences), 2024, 44(5): 15-26.
- [8] 网络主播行为规范(广电发〔2022〕36号)[R]. 北京: 国家广播电视总局、文化和旅游部, 2022.
Code of conduct of online anchors (Radio and Television [2022] No. 36) [R]. Beijing: State Administration of Radio and Television, Ministry of Culture and Tourism, 2022.
- [9] 《中国网络视听发展研究报告(2024)》发布 我国网络视听用户规模达 10.74 亿(文化市场新观察)[N/OL]. 人民日报, 2024-03-28(12). http://paper.people.com.cn/rmrb/html/2024-03/28/nw.D110000renmrb_20240328_1-12.htm.
"Research Report on the Development of Internet Audio-Visual in China (2024)" released. The scale of Internet Audio-Visual in my country reached 1.074 billion (New Observation on the Cultural Market)[N/OL]. People's Daily, 2024-03-28(12). <http://paper.people.com.cn/>

- rmrb/html/2024-03/28/nw.D110000renmrb_20240328_1-12.htm.
- [10] 蒲清平, 黄媛媛. 习近平总书记关于新质生产力重要论述的生成逻辑、理论创新与时代价值[J]. 西南大学学报(社会科学版), 2023, 49(6): 1-11.
- Pu Q P, Huang Y Y. Generation logic, theoretical innovation and time value of general secretary xi Jinping's important exposition on new quality productivity[J]. Journal of Southwest University (Social Sciences Edition), 2023, 49(6): 1-11.
- [11] 12·5 东方甄选"小作文"事件[EB/OL]. [2025-03-24]. https://www.baike.com/wikiid/7312764246588194835?share_token=d8031968-1a40-41fe-bae0-1e1554992f62&tab=doc.
- East Buy Holding Limited "Small Essay" event[EB/OL]. [2025-03-24]. https://www.baike.com/wikiid/7312764246588194835?share_token=d8031968-1a40-41fe-bae0-1e1554992f62&tab=doc.
- [12] 澎湃新闻. 东方甄选净利下滑,薪酬开支增近1.7倍!俞敏洪:只依赖董宇辉就不是正常公司[EB/OL]. (2024-01-26)[2025-03-24]. https://www.thepaper.cn/newsDetail_forward_26151301?commTag=true.
- The Paper. Net profit of East Buy Holding Limited declined, salary and expenditure increased by nearly 1.7 times! Yu Minhong: A company that only relies on Dong Yuhui is not a normal company[EB/OL]. (2024-01-26) [2025-03-24]. https://www.thepaper.cn/newsDetail_forward_26151301?commTag=true.
- [13] 21 世纪经济报道. "董宇辉现象"的启示[EB/OL]. (2023-12-19) [2025-03-24]. <https://www.21jingji.com/article/20231219/herald/786781a28a27fc39a49ee43973003fd1.html>.
- 21st Century Economic Report. The enlightenment of the Dong Yuhui Phenomenon[EB/OL]. (2023-12-19) [2025-03-24]. <https://www.21jingji.com/article/20231219/herald/786781a28a27fc39a49ee43973003fd1.html>.
- [14] 人民网. 什么是新质生产力?中央财办最新解读[EB/OL]. (2023-12-18)[2025-03-24]. <http://finance.people.com.cn/n1/2023/1218/c1004-40141533.html>.
- People. cn. What is the New Qualitative Productivity? The latest interpretation of the Central Finance Office[EB/OL]. (2023-12-18)[2025-03-24]. <http://finance.people.com.cn/n1/2023/1218/c1004-40141533.html>.
- [15] 马克思恩格斯文集(第六卷)[M]. 中共中央马克思恩格斯列宁斯大林著作编译局, 编译. 北京: 人民出版社, 2009: 147.
- Collected Works of Marx and Engels (Volume VI) [M]. Central Compilation and Translation Bureau for Marx, Engels, Lenin, and Stalin Works of the CPC Central Committee, Compiled. Beijing: People's Publishing House, 2009: 147.
- [16] 胡莹, 刘铿. 新质生产力推动经济高质量发展的内在机制研究: 基于马克思生产力理论的视角[J]. 经济学家, 2024(5): 5-14.
- Hu Y, Liu K. Research on the internal mechanism of new quality productivity promoting high quality economic development: From the perspective of Marx's productivity theory[J]. Economist, 2024(5): 5-14.
- [17] 马克思恩格斯文集(第八卷)[M]. 中共中央马克思恩格斯列宁斯大林著作编译局, 编译. 北京: 人民出版社, 2009.
- Collected Works of Marx and Engels (Volume VIII) [M]. Central Compilation and Translation Bureau for Marx, Engels, Lenin, and Stalin Works of the CPC Central Committee, Compiled. Beijing: People's Publishing House, 2009.
- [18] 马克思恩格斯文集(第五卷)[M]. 中共中央马克思恩格斯列宁斯大林著作编译局, 编译. 北京: 人民出版社, 2009: 147.
- Collected Works of Marx and Engels (Volume V) [M]. Central Compilation and Translation Bureau for Marx, Engels, Lenin, and Stalin Works of the CPC Central Committee, Compiled. Beijing: People's Publishing House, 2009.
- [19] 马克思恩格斯文集(第一卷)[M]. 中共中央马克思恩格斯列宁斯大林著作编译局, 编译. 北京: 人民出版社, 2009.
- Collected Works of Marx and Engels (Volume I) [M]. Central Compilation and Translation Bureau for Marx, Engels, Lenin, and Stalin Works of the CPC Central Committee, Compiled. Beijing: People's Publishing House, 2009.
- [20] 熊彼特. 经济发展理论: 对于利润、资本、信贷、利息和经济周期的考察[M]. 何畏, 易家祥, 等. 译. 北京: 商务印书馆, 2000: 68.
- Schumpeter J A. The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle[M]. Cambridge: Harvard University Press, 1934.
- [21] 马克思恩格斯文集(第七卷)[M]. 中共中央马克思恩格斯列宁斯大林著作编译局, 编译. 北京: 人民出版社, 2009: 373.
- The Collected Works of Marx and Engels (Volume VII) [M]. Central Compilation and Translation Bureau for Marx, Engels, Lenin, and Stalin Works of the CPC Central Committee, Compiled. Beijing: People's Publishing House, 2009: 373.
- [22] 潘恩荣. 《资本论》研究需要引入"技术逻辑"[J]. 哲学研究, 2015(9): 103-107.
- Pan E R. The study of *Das Kapital* needs to introduce "technical logic"[J]. Philosophical Researches, 2015(9): 103-107.
- [23] Kissinger H A, Schmidt E, Huttenlocher D. ChatGPT heralds an intellectual revolution[N]. The Wall Street Journal, 2023-02-24.

- [24] OpenAI. Our approach to alignment research[EB/OL]. (2022-08-24) [2025-03-24]. <https://openai.com/index/our-approach-to-alignment-research/>.
- [25] Christian B. The alignment problem: Machine learning and human values[J]. Perspectives on Science and Christian Faith, 2021, 73(4): 245-247.
- [26] Ji J M, Qiu T Y, Chen B Y, et al. AI alignment: A comprehensive survey[EB/OL]. 2023: 2310.19852. <https://arxiv.org/abs/2310.19852v5>.
- [27] Awad E, Dsouza S, Kim R, et al. The moral machine experiment[J]. Nature, 2018, 563(7729): 59-64.
- [28] 潘恩荣, 孙宗岭. 知识转向: 人工智能时代技术哲学的分裂风险与应对思路[J]. 自然辩证法研究, 2023, 39(2): 81-87.
- Pan E R, Sun Z L. The know turn: The risk of division and countermeasures in the philosophy of technology for the era of artificial intelligence[J]. Studies in Dialectics of Nature, 2023, 39(2): 81-87.

Preliminary Study on Ethical Alignment Design of Man-machine Community

Ruan Fan¹, Pan Enrong^{2*}

1. School of Marxism, Zhejiang University of Technology, Hangzhou 310028, China;

2. Department of Philosophy (Zhuhai), Sun Yat-sen University, Zhuhai 519082, China

Highlights

Ethical precedence is core to the development of new quality productivity in the digital economy.

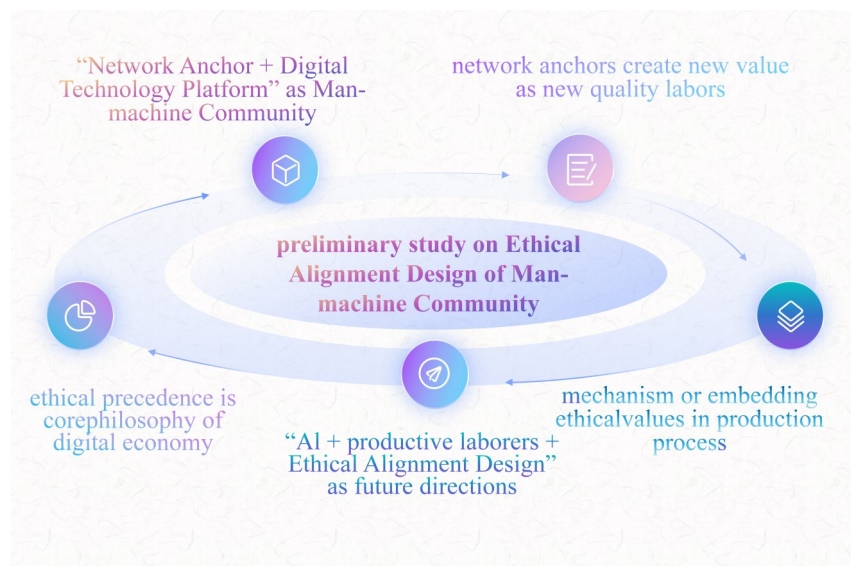
Introducing “ethical alignment design” as a framework for new quality laborers in the digital economy.

The Dong Yuhui case demonstrates ethical precedence’s practical effects in live streaming.

Highlighting the crucial role of new quality laborers with “ethical alignment design” skills in the digital economy.

Future directions include “AI + productive laborers + ethical alignment design” for sustainable development of digital economy.

Graphical Abstract



Abstract: In today's thriving digital economy, ethical prioritization has emerged as a critical force in advancing the development of new quality productivity. This paper introduces the concept of Ethically Aligned Design and reveals its important role in enhancing commodity value, promoting high-quality consumption and leading the transformation of production mode. As an emerging form of digital economy, live broadcast economy has formed a unique model of "Network Anchor + Digital Technology Platform", constituting a new Man-machine Community of digital economy. In the era of artificial intelligence, this paper believes that one of the New embodiment of the new quality productivity is the community of Man-machine, which is a new factor to promote the high-quality development of productivity, and the other of the quality is not only reflected in the traditional production of high-quality products or services, but also reflected in the artificial intelligence production system of

high ethical products or services, in order to cope with the new round of scientific and technological revolution and industrial change represented by artificial intelligence caused by ethical challenges. Ethically Aligned Design refers to the shift of network anchors from sellers to producers, giving new ethical values to commodities based on their own knowledge, skills and ethical literacy. This paper takes Dong Yuhui Phenomenon of digital economy as a specific case to explore how network anchors can become new quality laborers to embed new ethical value to commodities. At the same time, it analyzes how network anchors use labor materials (that is, digital technology platform) to transform labor objects (such as window goods) by the thinking of ethical prioritization, so as to realize the optimal combination and qualitative change of labor elements. Dong Yuhui's Live Commentary on the live delivery platform created new ethical value and new use value of window goods. Compared with the traditional economic thinking of commercial salesman or industrial workers, Dong Yuhui is a new quality laborer, and is a ethical care style of labor. His live commentary is to carry out secondary labor for the object of labor-the window goods sold in the broadcast room-with the help of labor materials-digital technology live broadcast platform of East Buy Holding Limited, and his own knowledge and skills and ethical literacy. This increases the value of the goods in the window, and the consumer also gains the additional use value of the goods-background knowledge, traditional culture and ethical care. The conclusions of this paper are as follows. First of all, ethical prioritization is a necessary condition for the development of digital technology, and it is also a new direction for the development of new quality productivity in the digital economy. As big data and artificial intelligence (AI) technologies converge, ethical considerations transcend mere compliance, *i.e.* they become a core philosophy driving innovation. This approach demands that technological advancements align with ethical principles and societal values, ensuring the legitimacy and rationality of their applications. Second, the framework of Ethical Alignment Design provides actionable guidance for new quality laborers in the digital economy. Network anchors, as representatives of this emerging workforce, leverage their expertise and ethical awareness to imbue commodities with novel ethical value. Such value extends beyond material attributes to encompass cultural resonance, emotional connection, and social responsibility. A prime example is the Dong Yuhui Phenomenon that, through culturally rich and humanistic live commentary, he redefined the ethical value of products, achieving dual success in commercial marketing and cultural dissemination. Further, the paper emphasizes the primacy of new quality laborers in the digital economy. These new quality laborers have the ability of productive labor and Ethically Aligned Design, which can create the new ethical value of commodities. Their labor not only increased the market value of goods, but also promoted the transformation of production methods. Through the introduction of ethical elements, new quality laborers lead the innovation of the production process, opening a new path for the future development of the digital economy. Finally, this paper looks forward to the future development direction of digital economy. As AI and large language models advance, the integration of "AI + Productive Labor + Ethical Alignment Design" will define the ethical prioritization of emerging digital technologies and economic models. This paradigm emphasizes not only technological innovation but also the guiding role of ethics and values. It calls for balancing technological convenience with ethical accountability and social stewardship. To sum up, ethical prioritization is indispensable to the digital economy. Through Ethical Alignment Design, we can enhance commodity value, promote high-quality consumption, revolutionize production systems, and chart a principled path for the future of digital innovation. This approach ensures that technological progress remains anchored in human dignity and societal well-being. Ethical alignment design of Man-machine Community provides a Chinese-style modernized perspective and approach for future AI alignment research. Western ethical alignment design advocates "ethical values embedded in artificial intelligence or digital technology systems", while Chinese ethical alignment design of Man-machine Community practices "ethical values embedded in products and their production processes through artificial intelligence or digital technology systems".

Keywords: ethical prioritization; Man-machine Community; Ethical Alignment Design; New Quality Productivity; Dong Yuhui Phenomenon