

劳动要素数字化与收入分配

——基于机器学习方法的实证研究

刘 波 李芸琪 张 磊*

摘要：近年来,我国出现有酬劳动时间下降与收入水平持续提升的背离现象,成倍增长的上网时间为此提供了新的研究视角。本文从广义与狭义两个层面界定劳动要素数字化程度,运用机器学习方法系统考察数字劳动对收入分配的影响。研究发现,数字劳动能显著提升劳动者收入水平;归因分析表明,有酬数字劳动的重要性得分在现有解释变量中位列第一;不平等分解结果显示,数字劳动差异对劳动者收入差距具有较强解释力,其中有酬数字劳动差异的相对贡献最为突出。进一步分析发现,劳动要素数字化转型存在机会不平等,人口统计学特征和原生家庭环境等外生因素塑造了数字劳动供给差异。本文揭示了劳动要素数字化影响收入分配的路径,为完善数智经济时代的收入分配制度提供了理论依据与经验支持。

关键词：数字经济;数字鸿沟;数字劳动;收入分配;机会不平等

中图分类号：F061.3;F047.1

一、引言

当前,大数据、云计算、人工智能等数字技术正加速创新迭代,并与经济社会各领域深度交融。数据成为新的生产要素,算法演变成劳动分工与价值分配的新逻辑,数字平台发展成要素配置的新中枢。如果不能保障个体平等参与数字劳动和价值创造的权利,可能会加剧贫富分化,阻碍社会进步。经济数字化转型并未自动带来机会共享与收入均衡,反而可能会加剧社会结构分化,使得原本处于边缘位置的群体,更难进入数字价值创造与获取体系(Johnson and Acemoglu, 2023)。从总量上看,数字经济不仅能够提升生产要素的配置效率,还能发挥消费潜力释放、就业创造等效应,为规模经济、范围经济以及长尾效应创造有利条件。然而,从结构上看,由于劳动者的禀赋存在差异,在经济数字化转型的过程中会形成数字鸿沟,产生数字机会不平等问题,从而改变收入分配格局(周文、韩文龙, 2023; 徐怀宁、田

* 刘波,湖南农业大学经济学院,邮政编码:410128,电子邮箱:cq_liubo@163.com;李芸琪(通讯作者),湖南大学经济与贸易学院,邮政编码:410079,电子邮箱:liyunqi0915@126.com;张磊,湘潭大学商学院,邮政编码:411105,电子邮箱:zlpd@xtu.edu.cn。

本文得到国家社会科学基金后期资助重点项目“数字经济时代收入机会不平等的治理研究:理论、实证与对策”(25FJLA002)、国家社会科学基金重大项目“普惠金融促进共同富裕的制度设计与实践路径研究”(25&ZD117)的资助。感谢评审专家和编辑部的宝贵意见,作者文责自负。

亚男,2025)。因此,如何阻断数字机会不平等向收入机会不平等转变,是数字经济时代完善收入分配制度亟待解决的重大现实问题。

马克思(1975)指出,“工作日不是一个不变量,而是一个可变量”。伴随5G等新一代通信技术的普及,以及大数据、云计算、人工智能等数字技术的应用,数字技术的载体——互联网,成为日常工作和生活的必需品,同时也改变了劳动供求形式。随着平台经济、社交媒体和数据驱动型商业模式的兴起,个人在网络空间中的行为逐渐被转化为可量化的数据,并成为新的生产要素,从而产生了一种新的劳动形式——数字劳动(陈梦根、周元任,2022)。数字劳动是一种以互联网、大数据与人工智能等数字技术为依托、以智力创造为核心的劳动形态,主要通过对数据与信息等无形要素的加工以生产数字化产品与服务。数字劳动既涵盖网约车、外卖配送、线上自由职业等有酬的平台劳动与灵活用工形式,也包括用户通过点击、浏览、评价及社交互动所贡献的无偿劳动。Fuchs(2014)以马克思劳动价值论为基础,辩证地分析了数字劳动的两面性。一方面,人工智能技术应用、协作平台构建、开源生态培育和数字公共资源建设等措施,显著提升了劳动生产率与收入水平。另一方面,数字技术淡化了工作与生活的时空边界,工作时间看似下降,实则增加;为数众多的用户为平台无偿贡献了数字足迹,而平台则以广告推送、算法推荐和数据交易的形式获利,形成了隐蔽的数字剥削。更为重要的是,数字劳动往往不由劳动者自主决定,而是被技术架构和平台逻辑所规训,用户被“点赞—分享—沉迷”的机制所驱动,导致劳动与人的分离,有碍于收入分配的公平性。

在数智经济转型持续深化的时代背景下,数字劳动逐步发展成为劳动实践的主要形式,其范畴几乎涵盖数字经济所涉及的全部劳动(杨静、郑嵘,2024)。数字鸿沟已经从最初物理层面的“接入鸿沟”,演进至数字能力和数字劳动层面的“效用鸿沟”,并最终演变为数字劳动的回报差异(Mendonça and Kougiannou, 2023)。鉴于此,本文以数字劳动理论为基础,以上网时间作为劳动要素数字化的代理变量,量化分析居民上网时间与收入水平的内在联系,以探究数字经济时代收入分配的基础逻辑。本文主要的学术贡献包括三个方面:第一,在理论层面,以马克思剩余价值生产、劳动异化与资本积累理论为基础,剖析数字劳动本质,明确劳动要素数字化转型与收入分配的辩证关系,厘清数字经济时代的收入分配逻辑;第二,在测度方法层面,创新性地将基于机器学习的非线性方程引入不平等指标分解中,突破了传统分解方法解释力不足的局限,并揭示了收入函数设定对不平等指标分解结果的影响;第三,在实证层面,明确了劳动要素数字化对收入分配格局的影响,解释了“有酬劳动时间与收入水平倒挂”的异象,并甄别了有酬数字劳动和无酬数字劳动对收入的异质性影响,为数智经济时代的收入分配改革提供经验证据。

本文余下内容包括4个部分:首先,厘清数字劳动的理论内涵,明确上网时间与数字劳动的逻辑关系,并从广义和狭义两个层次界定数字劳动,提出研究假说。其次,根据数字劳动与收入水平的逻辑关系,构建收入方程,设计收入差距的分解方案。第三,以中国家庭追踪调查(CFPS)为样本,实证研究劳动要素数字化对收入分配格局的影响,揭示劳动要素数字化的“有偏”特征,识别影响收入差距的主要因素。最后,总结研究结论,提出相应的政策建议。

二、文献综述

(一) 政治经济学视角下的劳动要素数字化

数字技术革新了劳动要素的供需形式,劳动要素数字化包括劳动配置方式的数字化以及劳动供给形式的数字化。上网时间的增加表面上是个人时间安排的调整,但本质上是生产力发展的结果(Fuchs,2014)。数字劳动并未脱离物质劳动范畴,而是资本运行逻辑在数字技术条件下的延展与重构(庞跃辉、丁乙,2025)。首先,在数字经济中,平台不仅从网约车司机、外卖骑手等提供的有酬劳动中获取剩余价值,还通过用户的点击、浏览、评价与社交互动等行为,将日常生活活动转化为可计量、可交易的数据资源,实现对无偿劳动的隐性占有(黄再胜,2026)。数字劳动使剩余价值生产突破了时间与空间的边界,呈现出“泛在化”与日常化特征。其次,在资本积累方面,数字平台依托数据要素与网络效应,通过持续收集、整合并垄断海量数据资源,实现规模扩张与市场支配,即“数字化原始积累”(欧阳康、杨德科,2023),并通过制度与技术手段将分散的个体数据转化为被资本控制的关键生产要素,进而推动资本积累由传统的物质形态向数据形态升级。总体来看,数字劳动并未改变资本运行的基本逻辑,而是使剩余价值生产更为多元化、劳动异化日益复杂化、资本积累更加集中化,从而改变收入分配格局。

(二) 劳动要素数字化影响收入分配的机制与效应

在历次推动生产力发生重大跃迁的技术革命中,技术进步均呈现出“有偏”特征,对部分群体产生了不利影响(Johnson and Acemoglu,2023)。在数字技术应用普及的过程中,数字鸿沟逐渐形成,其含义并不局限于个体能否使用数字化工具,更为重要的是各群体因数字技术使用能力和利用程度不同所产生的福利差异。在数智经济时代,数字鸿沟将最终演变为收入差距,而数字机会不平等则演变为收入机会不平等。

数字鸿沟的内涵先后经历了接入鸿沟、技能鸿沟和效用鸿沟三个阶段,三者构成递进关系(陈梦根、周元任,2022)。接入鸿沟是指在互联网走入日常生活的初期,欠发达地区、低收入群体难以借助互联网摆脱信息贫瘠的状态,聚焦于数字基础设施对能否使用互联网的影响(Aker and Mbiti,2010)。随着数字基础设施和移动设备的普及,数字鸿沟进一步表现为技能鸿沟,即使用互联网设备的技能差异(van Deursen and van Dijk,2019)。效用鸿沟之所以存在,表面上源于接入鸿沟和技能鸿沟,但本质上是生产要素数字化与数据要素化的结果。然而,并不是所有劳动者都能适应数字化的要素配置方式,数字素养偏低的低收入群体在使用数字技术时,无法充分利用技术改善生活;而数字素养较高的高收入群体则更多地将数字技术应用于学习与工作,由此在两个群体之间产生了数字效用鸿沟(Ramdas and Sungu,2024)。无论是数字鸿沟,还是数字红利,本质上都是与个人禀赋相关的机会不平等(王宁、胡乐明,2022)。社会地位较高的群体因获取信息的成本更小,且能凭借技术优势获得更多的资本增值机会,从而不断强化其既有优势(DiMaggio and Garip,2012)。由此,数字鸿沟通过加剧群体间的数字机会不平等,最终形成收入机会不平等(Robinson et al.,2015;Katz and Gonzalez,2016)。

数字机会不平等的成因是多方面的。物质资本是首要因素,欠发达地区因缺乏数字基础设施投入,数字化程度相对较低(van Deursen et al.,2021)。其次是人力资本,受教育程度是决定数字素养的关键要素,个人受教育程度越高,能够获取的数字红利也就越多;老年人互联网使用率较低,不同年龄组之间存在较为明显的数字鸿沟。三是互联网使用的同群效应,在网民密度相对较高的城市环境中,个体更容易从工作生活中获取数字知识、学习数字技能。最后,国家层面的数字发展战略与地区提升数字化的政策措施,在消弭数字鸿沟、缓解数字机会不平等方面发挥着关键作用(Szeles,2018)。智能化是数字化的必然延伸,随着通用人工智能(Artificial General Intelligence, AGI)适用场景的不断拓展,AGI 可能会导致劳动力的广泛替代、就业的结构性收缩以及劳动报酬的大幅下降,人机矛盾将进一步加深(Acemoglu and Restrepo,2020a)。现有研究表明,AGI 具有明显的偏向性技术进步特征,AGI 的应用主要取代低技能和中等技能劳动者,由此加剧了高技能和中低技能工人之间的机会不平等,进而导致总体收入差距扩大(Acemoglu and Restrepo,2020b)。

综上,随着经济数字化的深入,数字不平等的焦点已从“是否接入互联网”发展到“能否使用互联网”,并最终演进为“能否平等地从互联网中获益”(Van Deursen and Van Dijk, 2019)。鉴于此,探究劳动要素数字化是否显著改变了收入分配格局,劳动要素数字化转型是否存在机会不平等,具有重要理论价值和现实意义。

三、研究设计

(一)典型事实与研究假说

根据第二、三次《全国时间利用调查公报》,2018—2024 年间,中国居民的每日平均有酬劳动时间由 264 分钟下降至 207 分钟,减少 21.59%;互联网的每日平均使用时间从 162 分钟上升至 337 分钟,增长 108.02%。然而,同期全国居民人均可支配收入由 28228 元增至 41314 元,增长 46.36%,有酬劳动时间与居民收入呈现“倒挂”现象。在收入分配政策既定的条件下,收入水平由劳动效率和劳动时间共同决定,当有酬劳动时间下降 21.59%时,仅靠劳动效率的提升难以支撑 46.36%的收入增长幅度。因此,本文认为大幅增长的上网时间在一定程度上弥补了有酬劳动时间的减少。在经济数字化转型持续深化的背景下,数字化正在重塑劳动要素配置方式。劳动从传统物理空间转移到网络平台与虚拟环境中,劳动过程被拆分为更细微的任务单元,劳动者可根据自身时间安排与技能禀赋灵活参与劳动。工作与生活、正式劳动与非正式劳动的界限逐渐淡化,个人上网行为也可能被平台吸纳为价值创造过程。因此,本文将数字劳动划分为广义和狭义两个层次,广义数字劳动包括所有能够产生数字足迹的网络行为,具体包括生产性的有酬数字劳动和休闲性的无酬数字劳动,其中前者被定义为狭义数字劳动。有酬数字劳动的具体形式包括数字化办公、平台劳动,无酬数字劳动则涉及用户在社交媒体、平台中进行的点击、点赞、发布、评论和搜索等活动。随着互联网与数字终端的迭代升级,数字劳动的范畴与形式不断拓展,已不再局限于借助数字终端完成的劳动供给,而是演化为所有劳动者都直接或者间接参与的劳动形态(Fuchs,2014)。因而,有酬劳动时间减少并不意味着有效劳动时间减少,如果将不受时空限制的数字劳动考虑

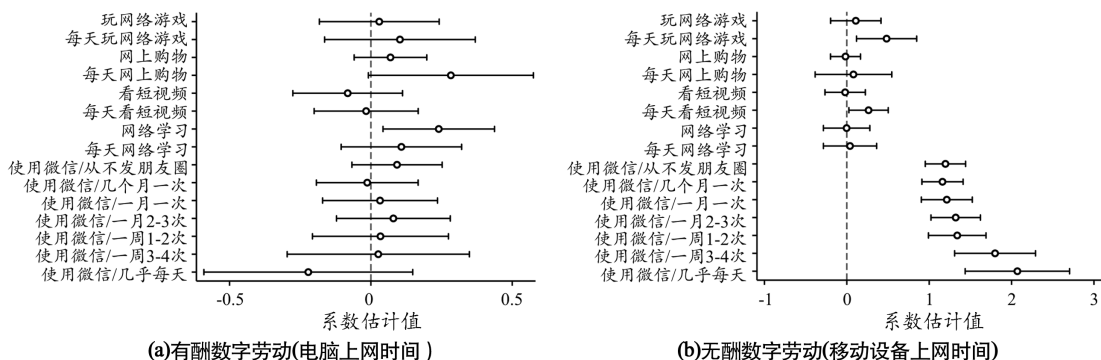
在内,有效劳动时间可能更多。基于上述分析,本文提出第一个研究假说:

H1:数字劳动对收入水平存在显著影响。

然而,由于新技术具有显著的“有偏”特征,并非所有个体都能在互联网环境中获得均等福利。根据《全民数字素养与技能发展水平调查报告(2024)》发布的数据^①,不同群体中的“高级数字素养”占比存在显著差异。在城镇居民中,数字素养达“高级”的比重为19.83%,而在农村居民中仅占9.53%;大学本科及以上的占比为53.49%,高中、初中、小学及以下分别为18.21%、7.62%和5.43%;18-29岁的占比约为60-69岁的4倍;农、林、牧、渔业生产及辅助人员、社会生产服务和生活服务人员、生产制造及有关人员均低于全国平均水平。按照Roemer(1998)的“环境—努力”机会不平等分析框架,能力差异本质上是由外生环境与内生努力共同作用的结果。数字素养既是努力的体现,更受制于环境约束,由环境因素塑造的数字素养差距,会在数字劳动供给中被持续放大,将机会不平等由“起点差异”转化为“结果差异”。一方面,高数字素养群体能够进入高附加值的数字岗位,实现更灵活的劳动供给与更高的收益回报;另一方面,低数字素养群体往往被锁定在低技能、低报酬岗位,甚至面临被边缘化的风险。即便在努力程度相当的情况下,不同个体的劳动供给机会与结果仍呈现系统性分化。因此,数字技术的应用并未自动消解数字机会不平等,反而可能通过能力塑造的偏向性机制,重塑并加深既有的机会不平等。基于此,本文提出第二个研究假说:

H2:数字劳动供给受个人禀赋影响,劳动者的数字化转型存在机会不平等现象。

本文以中国家庭追踪调查(CFPS)为样本数据,将未退出劳动力市场的受访者作为研究对象。对于个人而言,不同的上网设备与上网目的紧密相关,社交娱乐普遍以移动设备为工具,而数据处理、内容创作、学习通常以电脑为工具。图1列示了2020年、2022年的网络足迹(网络游戏、网上购物、短视频、网络学习、微信)与两类上网时间的回归系数^②。



注:系数估计值的置信区间为95%。

图1 网络足迹与上网时间

如图1所示,若以电脑上网时间为被解释变量,在5%的显著性水平上,仅“网络学习”的

①依据数字素养得分,该调查报告将受访者分为“初级”“中级”和“高级”三个等级。

②模型设定为面板数据双向固定效应模型。

系数显著,网络游戏、网上购物、短视频、微信四类社交娱乐活动对电脑上网时间均无显著影响;若以移动设备上网时间为被解释变量,在 5% 的显著性水平上,每天玩网络游戏、每天看短视频、使用微信均对其存在显著影响。鉴于此,本文根据上网设备类型对数字劳动进行分类,即将通过智能手机或者平板电脑等移动设备产生的上网时间界定为无酬数字劳动,而通过电脑产生的上网时间则被定义为有酬数字劳动。考虑到在现有的分配制度中,仅狭义数字劳动(有酬数字劳动)能得到相应的经济回报,因此,本文提出第三个研究假说:

H3:狭义数字劳动对收入水平和收入差距的影响更为突出。

(二)实证模型的构建

如上所述,本文将上网时间作为数字劳动的量化指标,根据数字劳动与收入的逻辑关系,将收入(y)与数字劳动(T)的方程分别设定为:

$$y_i = c_1 + \alpha_1 T_i + \beta_1 X_i + \varepsilon_i \tag{1}$$

$$T_i = c_2 + \beta_2 X_i + \xi_i \tag{2}$$

(1) — (2) 式中: X_i 为控制变量。考虑到数字技术应用具有“有偏”特征,即数字技术的习得受个人难以事先左右的外生因素与个人主观能动性的共同影响,外生因素主要包括人口统计学特征和原生家庭环境变量,而主观能动性则体现为个体的努力程度。需要注意的是,努力程度通常以受教育程度为代理变量,但其本身同样受人口统计学特征和原生家庭环境变量的影响。鉴于此,为避免模型存在多重共线性,本文通过设定辅助方程 $edu_i = c + \beta X_i + \tau_i$,剔除受教育程度中由其他控制变量解释的部分,即以模型的广义残差项作为努力程度的代理变量(刘波等,2020)。根据上网途径,本文进一步将上网时间细分为电脑上网时间(T_{1i})和移动设备上网时间(T_{2i}),由此,基准模型(1) — (2) 式可以改写为:

$$y_i = c_1 + \alpha_2 T_{1i} + \alpha_3 T_{2i} + \beta_1 X_i + \varepsilon_i \tag{3}$$

$$T_{1i} = c_3 + \beta_3 X_i + \xi_{1i} \tag{4}$$

$$T_{2i} = c_4 + \beta_4 X_i + \xi_{2i} \tag{5}$$

考虑到上网时间与收入之间可能存在双向因果关系,本文将采用工具变量法以缓解潜在的内生性问题。本文将上网时间视为劳动技能的体现,而劳动技能的习得与环境、努力等因素密切相关,导致收入方程可能存在混杂结构。为缓解模型设定偏误以及混杂因素带来的潜在偏误,本文进一步在线性模型的基础上引入双重机器学习(Double Machine Learning, DML)法,以验证估计结果的稳健性。本文的研究重点并非识别因果效应,而是借鉴收入机会不平等研究范式,通过基于线性回归方程的夏普利值分解方法,考察劳动要素数字化差异的形成来源,并评估劳动要素数字化差异对收入差距的解释力度。由于线性模型的拟合优度偏低,导致残差项对总体差距的贡献过大。对于残差项应归属于环境因素还是努力程度,现有文献尚未达成一致(Almås et al., 2011; Lefranc et al., 2009)。当残差项的贡献超过 50% 时,无论将之归属于哪一类,均会影响分解结果,其根源在于线性模型的拟合效果较差。相比于线性模型,神经网络、支持向量机、随机森林等机器学习模型在解决拟合问题上具有突出优势,基于此,本文使用非线性的机器学习模型替代线性模型,即将收入函数形式调整为:

$$y_i = F(x_1, \cdots, x_K) + \zeta_i \tag{6}$$

(6)式中： $F(\cdot)$ 为经样本数据训练后的机器学习模型，此时收入函数由非线性部分 $F(x_1, \cdots, x_K)$ 和线性部分 ζ_i 构成。以(6)式为基础，沿用夏普利值分解的思路，可得基于非线性收入方程的夏普利值分解结果。虽然机器学习方法能够有效提升收入函数的拟合优度，但是否会造成收入差距分解结果大幅改变，现有文献尚未给出明确结论。因此，本文在基准分析中采用机器学习模型刻画收入函数，并在稳健性检验中将收入函数调整为线性模型，考察收入差距分解结果是否受收入函数拟合能力的影响。

(三) 变量的选择与赋值

在(1)—(5)式中，变量主要包括三类：收入、数字劳动和控制变量。其中，控制变量的选取主要借鉴机会不平等分析框架，涵盖环境禀赋和个人努力两个维度。环境禀赋主要包括人口统计学特征和原生家庭环境，人口统计学特征包括性别、年龄、民族、城乡属性、户口类型、健康状况和出生地区等变量，原生家庭环境包括父母受教育程度和父母政治面貌等变量；个人努力则通过剥离环境因素的受教育程度加以刻画。考虑到个体后天经历和社会流动特征，进一步控制户口变动、迁徙经历和工作时长等变量。各变量的具体定义与赋值方式如表 1 所示，描述性统计结果见表 2。为保证不同年度的收入具有可比性，本文依据各省城乡居民消费价格指数，将 2020 年和 2022 年的收入水平统一折算至 2010 年的实际价格水平（刘波等，2020）。

表 1 变量的定义与赋值

符号	变量	赋值方式
y	收入	受访者的年收入，单位：万元
T	数字劳动	广义数字劳动以上网时间(T)为量化指标，包括每周的电脑上网时间(T_1)与移动设备上网时间(T_2)，狭义数字劳动为电脑上网时间，单位：小时/周
L	工作时长	每周工作时间，单位：小时
$gender$	性别	男性为 1，女性为 0
age	年龄	受访者在调查年份的年龄
$ethnicity$	民族	汉族为 1，其他为 0
$urban$	城乡	根据国家统计局城乡分类赋值，城镇为 1，乡村为 0
$hukou_{12}$	户口类型	受访者在 12 岁时的户口类型，农业户口为 1，非农户口为 0
$west_birth$ $central_birth$ $east_birth$	出生地区	西部地区： $west_birth$ ，中部地区： $central_birth$ ，东部地区： $east_birth$
$pedu_{14}$	父母受教育程度	受访者 14 岁时，父母受教育程度的最高值；1=文盲/半文盲，2=小学，3=初中，4=高中，5=大专，6=大学本科，7=硕士，8=博士
$pparty_{14}$	父母政治面貌	受访者 14 岁时，父母均为群众时为 0，否则为 1
$hukou_adj$	户口变动	受访者与当前的户口不同时为 1，否则为 0
$migrate$	迁徙	受访者出生时的地区与当前所在地区不同时为 1，否则为 0
edu_res	个人努力程度	教育决定方程的残差项估计值
$health$	健康状况	根据自评健康赋值，1=不健康，2=一般，3=比较健康，4=很健康，5=非常健康
Y_{2022}	年份虚拟变量	调查年份为 2022 年时为 1，否则为 0

表 2 变量的描述性统计

变量	观测值	均值	标准差	最小值	最大值
<i>y</i>	12905	3.8380	4.2008	0.0001	134.5484
<i>T</i>	12905	22.6099	28.8147	0	224
<i>T</i> ₁	12905	5.7498	14.4454	0	168
<i>T</i> ₂	12905	16.8601	19.3895	0	168
<i>L</i>	12905	50.7201	19.9776	0.1000	168
<i>gender</i>	12905	0.5845	0.4928	0	1
<i>age</i>	12905	43.1503	11.5304	16	87
<i>ethnicity</i>	12905	0.9374	0.2423	0	1
<i>urban</i>	12905	0.5796	0.4936	0	1
<i>hukou</i> ₁₂	12905	0.8456	0.3613	0	1
<i>west_birth</i>	12905	0.2497	0.4328	0	1
<i>central_birth</i>	12905	0.3235	0.4678	0	1
<i>east_birth</i>	12905	0.3938	0.4886	0	1
<i>pedu</i> ₁₄	12905	2.4479	1.1440	1	8
<i>pparty</i> ₁₄	12905	0.1791	0.3834	0	1
<i>hukou_adj</i>	12905	0.1888	0.3914	0	1
<i>migrate</i>	12905	0.1372	0.3441	0	1
<i>edu_res</i>	12905	0.0034	0.9597	-3.2063	4.1727
<i>health</i>	12905	3.2258	1.0748	1	5
<i>Y</i> ₂₀₂₂	12905	0.5096	0.4999	0	1

四、实证研究

(一) 数字劳动与收入分配

1. 基于工具变量法的分析

借鉴 Bartik 工具变量构造法 (Bartik, 1991), 按照“基期特征+宏观冲击”的思路, 将 2014 年受访者所在区县的互联网普及率与其所在行业平均上网时间的乘积, 作为总体上网时间的工具变量; 类似地, 电脑上网时间和移动设备上网时间的工具变量, 分别为家庭电脑普及率、手机普及率与其所在行业平均上网时间的乘积^①, 估计结果如表 3 所示。内生性检验的 χ^2 统计量分别为 44.2904、35.4275, 对应的 *P* 值均小于 1%, 表明总体上网时间、电脑上网时间和移动设备上网时间存在内生性; 而弱工具变量检验的 Kleibergen-Paap rk Wald F 统计量分别为 414.6318、272.2021, 排除了弱工具变量问题。对 (1) 式的估计结果表明, 总体上网时间的系数为 0.0902, 在 1% 的水平上显著, 表明上网时间对居民收入水平存在显著的正向影响。进一步区分上网时间类型, 对 (3) 式的估计结果显示, 电脑上网时间的系数为 0.1301, 在 1% 的水平上显著, 而移动设备上网时间的系数不显著, 由此可知电脑上网时间是影响居民收入的主要数字劳动形式。为探究内生性对估计结果的影响, 表 3 列示了 OLS 的估计结果。总体上网时间和电脑上网时间的系数分别为 0.0183、0.0475, 均在 1% 的水平上显著, 而移动设

①以 CFPS(2014 年) 为样本, 接受访者所在区县, 汇总得到互联网普及率、电脑普及率和手机普及率。

备上网时间系数仍不显著。对比发现,虽然 OLS 估计结果在方向和显著性水平上与工具变量法保持一致,但低估了数字劳动对收入分配的作用。综合两组估计结果可知,无论是广义数字劳动,还是有酬数字劳动,两者均对受访者收入水平存在显著的正向影响,由此验证了研究假说 1。

表 3 数字劳动与收入分配的工具变量回归结果

变量	y			
	Bartik IV		OLS	
	(1)	(2)	(3)	(4)
T	0.0902 *** (0.0121)		0.0183 *** (0.0020)	
T ₁		0.1301 *** (0.0182)		0.0475 *** (0.0043)
T ₂		-0.0005 (0.0130)		-0.0012 (0.0024)
控制变量	YES	YES	YES	YES
N	12905	12905	12905	12905
R ²	0.0605	0.1710	0.2248	0.2348
Kleibergen-Paap rk Wald F	414.6318	272.2021		
χ ²	44.2904	35.4275		

注:括号中为稳健标准误;***、**和*分别表示在 1%、5%和 10%的统计水平上显著。下表同。

2. 基于双重机器学习法的分析

为提高模型估计精度,本文放宽对收入函数的线性假说,在工具变量框架下引入 DML 法(Chernozhukov et al.,2018)。表 4 第(1)―(2)列为基于工具变量框架的双重机器学习(DML-Bartik IV)估计结果,总体上网时间、电脑上网时间的系数均为正,且在 1%的水平上显著,证实了劳动要素数字化对居民收入的促进作用,尤其是有酬数字劳动,再次验证了研究假说 1。与此同时,第(3)―(4)列为基于部分线性回归的双重机器学习(DML-PLR)估计结果,对比发现与前文结论一致,在未控制内生性时,劳动要素数字化在收入形成过程中的实际作用被系统性低估。

表 4 双重机器学习法的估计结果

变量	y			
	DML- Bartik IV		DML-PLR	
	(1)	(2)	(3)	(4)
T	0.0789 *** (0.0100)		0.0172 *** (0.0019)	
T ₁		0.0967 *** (0.0165)		0.0372 *** (0.0040)
T ₂		0.0217 * (0.0130)		0.0045 ** (0.0022)
控制变量	YES	YES	YES	YES
N	12905	12905	12905	12905

3.基于 XGBoost+SHAP 的归因分析

在前文结合工具变量法和 DML 法对内生性问题进行识别与校正的基础上,本文进一步采用 XGBoost+SHAP 方法开展归因分析。需要强调的是,该方法不以因果识别为目的,而是从非线性建模和变量重要性角度,检验上网时间在高维环境下对收入的预测贡献。首先,假设收入是数字劳动、环境因素、努力程度的函数。其次,将收入函数设定为决策树模型或者正则化线性模型,并利用样本数据进行拟合;依据拟合值和实际值,可计算得到与线性模型中拟合优度类似的评价指标。再次,鉴于 XGBoost 仅能提供拟合结果,难以揭示输入与输出的“黑箱”,本文进一步引入夏普利值,量化各变量的边际贡献,并据此进行变量重要性排序。

基于上述思路,本文采用 XGBoost 模型拟合收入函数,并以平方误差损失函数作为目标函数,引入 L_1 、 L_2 正则化对模型复杂度进行约束。在参数设定方面,学习率设定为 0.05,以提升模型的泛化能力;最大树深度设为 6,以在模型拟合能力与过拟合风险之间取得平衡;同时将最小子节点权重设置为 5,以限制节点分裂,从而抑制噪声拟合。为进一步增强模型稳健性,引入子样本抽样(Subsample Ratio 设为 0.8)与特征抽样(Column Subsample Ratio 设为 0.8)机制,以降低模型方差。在模型训练过程中,采用 K 折交叉验证($K=5$),并通过“提前停止策略”自动确定最优迭代次数。结果显示,基于广义数字劳动的收入方程,其拟合优度为 0.5269,明显优于线性模型的拟合效果($R^2=0.2248$)。由图 2 变量重要性得分可知,广义数字劳动位列第三,仅次于性别和努力程度,进一步验证了假说 1。细分数字劳动类型发现,非线性收入方程的拟合优度高达 0.5408,而线性模型的拟合优度仅为 0.2348。其中,有酬数字劳动的贡献最高,是决定居民收入水平的首要因素;其次是性别、努力程度、年龄与出生地区,由此验证了假说 3。

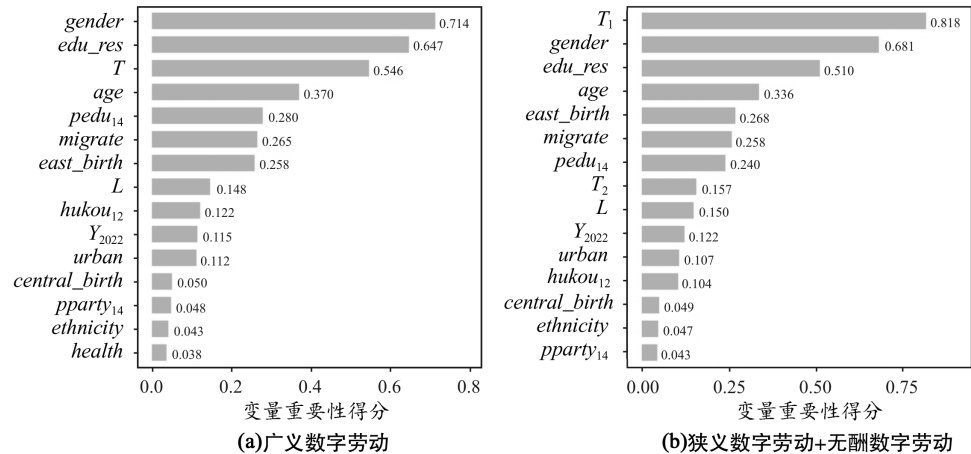


图 2 收入方程的变量重要性得分

(二) 劳动要素数字化转型的“有偏”性检验

为揭示个体禀赋差异对劳动要素数字化转型的影响,本文采用 OLS 方法分别对(2)式及(4)——(5)式进行估计,结果如表 5 所示。第(1)列报告了环境因素和努力程度对广义数字劳动的影响,第(2)——(3)列则分别考察其对狭义数字劳动即有酬数字劳动,以及无酬数字劳动的影响。以广义数字劳动为例,在 1%的水平上,除户口变动外,性别、户口类型、出生

地区、民族、父母受教育程度等外生变量均对广义数字劳动存在显著影响。由此可见,无论是人口统计学特征,抑或是原生家庭环境,均是影响劳动要素数字化的主要因素。有酬数字劳动、无酬数字劳动的估计结果与广义数字劳动基本一致。综合来看,数字技术的普及是有偏的,并不是所有人都拥有同等的劳动要素数字化转型机会,研究假说 2 得证。

表 5 劳动要素数字化的有偏性检验结果

变量	T	T_1	T_2
	(1)	(2)	(3)
<i>gender</i>	-2.3352 *** (0.4412)	-0.7489 *** (0.2395)	-1.5863 *** (0.3085)
<i>age</i>	-0.9451 *** (0.0217)	-0.2998 *** (0.0116)	-0.6453 *** (0.0151)
<i>ethnicity</i>	3.4527 *** (0.8530)	1.4154 *** (0.4463)	2.0373 *** (0.5755)
<i>urban</i>	2.6410 *** (0.4363)	1.0530 *** (0.2315)	1.5879 *** (0.3170)
<i>hukou</i> ₁₂	-12.9219 *** (0.8086)	-5.5936 *** (0.4546)	-7.3283 *** (0.5463)
<i>central_birth</i>	1.5566 *** (0.5452)	-0.0309 (0.2984)	1.5875 *** (0.3793)
<i>east_birth</i>	5.5394 *** (0.5317)	1.5531 *** (0.2901)	3.9863 *** (0.3726)
<i>pedu</i> ₁₄	2.6674 *** (0.2292)	1.1873 *** (0.1214)	1.4801 *** (0.1600)
<i>pparty</i> ₁₄	1.4828 *** (0.5696)	1.2227 *** (0.3171)	0.2601 (0.3989)
<i>hukou_adj</i>	0.7755 (0.6068)	0.1979 (0.3309)	0.5775 (0.4237)
<i>migrate</i>	5.0152 *** (0.7431)	2.2730 *** (0.4266)	2.7422 *** (0.5029)
<i>edu_res</i>	7.2630 *** (0.2314)	4.0486 *** (0.1307)	3.2144 *** (0.1578)
<i>health</i>	-0.7623 *** (0.1805)	-0.3183 *** (0.0946)	-0.4439 *** (0.1280)
Y_{2022}	1.6183 *** (0.4176)	0.0782 (0.2265)	1.5400 *** (0.2934)
常数项	62.2115 *** (1.8680)	18.8142 *** (1.0006)	43.3973 *** (1.2659)
N	12905	12905	12905
R^2	0.3236	0.2129	0.2656

遵循前文的 XGBoost+SHAP 方法设计,本文进一步从非线性视角和变量重要性角度,对人口统计学特征、环境因素与努力程度,在劳动要素数字化转型过程中的作用进行归因分析,估计结果如图 3 所示。在所有解释变量中,年龄、父母受教育程度、户口类型、出生地区以及城乡分类在三个方程中始终位于变量重要性排序前列。归因结果表明,劳动数字化转型高度依赖于外生因素,具有明显的非均衡特征,由此进一步验证了假说 2。

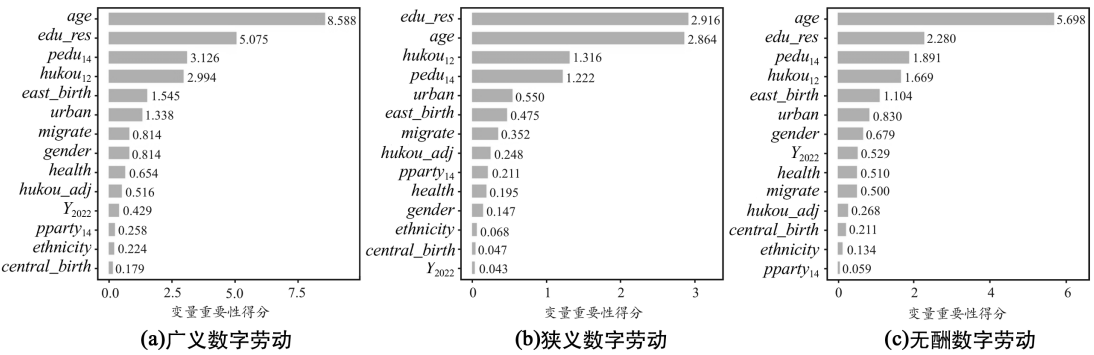


图 3 数字劳动方程的变量重要性得分

(三) 基于机器学习模型的不平等指标分解

1. 收入差距的分解结果分析

在归因分析的基础上,本文进一步基于非线性方程分解收入差距,以量化各变量对收入差距形成的相对贡献,分解结果如表 6 所示。

表 6 基于非线性收入方程的夏普利值分解结果

变量	y			
	绝对值	相对值	绝对值	相对值
	(1)	(2)	(3)	(4)
T	0.0482	6.43%		
T ₁			0.1596	19.19%
T ₂			0.0247	2.97%
L	0.0297	3.96%	0.0244	2.93%
gender	0.1198	15.98%	0.1074	12.91%
age	0.0471	6.28%	0.0362	4.35%
ethnicity	0.0065	0.87%	0.0059	0.71%
urban	0.0210	2.81%	0.0175	2.10%
hukou ₁₂	0.0026	0.35%	0.0026	0.32%
central_birth	0.0038	0.51%	0.0041	0.50%
east_birth	0.0207	2.75%	0.0200	2.40%
pedu ₁₄	0.0122	1.63%	0.0176	2.11%
pparty ₁₄	0.0011	0.15%	0.0018	0.21%
hukou_adj	0.0014	0.18%	0.0009	0.11%
migrate	0.0118	1.58%	0.0118	1.42%
edu_res	0.0306	4.08%	0.0411	4.95%
health	0.0050	0.66%	0.0043	0.52%
Y ₂₀₂₂	0.0257	3.43%	0.0213	2.57%
残差项	0.3627	48.36%	0.3303	39.72%

表 6 第(2)列基于广义数字劳动的分解结果显示,性别对收入差距的相对贡献最高,为 15.98%;其次是广义数字劳动,相对贡献为 6.43%;年龄位列第三,相对贡献为 6.28%;个人努力程度位列第四,相对贡献为 4.08%;每周工作时长位列第五,相对贡献为 3.96%。表 6 第(4)列进一步细分数字劳动类型发现,有酬数字劳动的相对贡献高达 19.19%,在所有变量中

位列首位;其次是性别、努力程度和年龄。值得注意的是,无酬数字劳动的相对贡献仅为 2.97%,表明不同类型数字劳动对收入差距的贡献具有显著异质性。总体而言,劳动要素数字化差异,特别是有酬数字劳动差异,已成为收入不平等的重要来源,由此验证了假说 3。在“环境—努力”框架下,外生环境差异导致的收入差距为收入机会不平等,故将性别、民族、户口类型、父母受教育程度和政治面貌、出生时的经济区域以及城乡的相对贡献加总,可得外生环境差异对收入差距的直接贡献,在广义数字劳动以及细分数字劳动框架下分别为 25.05%、21.26%。

2.数字劳动差距的分解结果分析

对数字劳动方程展开不平等指标分解,结果如表 7 所示。广义数字劳动差距的形成受多重因素的共同作用,年龄和户口类型的相对贡献最高,分别为 8.27%和 5.65%;其次是父母受教育程度、出生地区和城乡属性,相对贡献依次为 5.23%、4.50%和 3.70%。细分数字劳动类型发现,在有酬数字劳动的分解结果中,年龄依然为首要影响因素,相对贡献高达16.69%;其次是个人努力程度,相对贡献为 16.57%;第三至五位分别为父母受教育程度、户口类型和出生地区,相对贡献依次为 11.66%、11.13%和 2.80%。对于无酬数字劳动而言,年龄与个人努力程度仍是主要影响因素,相对贡献分别为 11.20%和 5.67%;其次是年份虚拟变量、出生地区和城乡属性,相对贡献依次为 4.23%、4.12%和 3.78%。由此可见,无论是广义数字劳动,还是有酬数字劳动,抑或是无酬数字劳动,其差距均在很大程度上源于个人禀赋差异。上述分解结果验证了劳动要素数字化转型的“有偏”特征,即数字红利更多地集中于具备较高人力资本和资源禀赋的群体,由此进一步验证了假说 2。

表 7 基于非线性数字劳动方程的夏普利值分解结果

变量	T		T ₁		T ₂	
	绝对值	相对值	绝对值	相对值	绝对值	相对值
	(1)	(2)	(3)	(4)	(5)	(6)
gender	0.0186	1.83%	0.0974	1.65%	0.0170	1.79%
age	0.0841	8.27%	0.9824	16.69%	0.1061	11.20%
ethnicity	0.0075	0.74%	0.0098	0.17%	0.0062	0.65%
urban	0.0376	3.70%	0.1592	2.70%	0.0358	3.78%
hukou ₁₂	0.0575	5.65%	0.6553	11.13%	0.0331	3.50%
central_birth	0.0079	0.78%	0.0214	0.36%	0.0120	1.26%
east_birth	0.0458	4.50%	0.1647	2.80%	0.0390	4.12%
pedu ₁₄	0.0532	5.23%	0.6866	11.66%	0.0259	2.74%
pparty ₁₄	0.0019	0.19%	0.0388	0.66%	0.0010	0.10%
hukou_adj	0.0166	1.63%	0.1530	2.60%	0.0086	0.91%
migrate	0.0130	1.28%	0.0655	1.11%	0.0124	1.30%
edu_res	0.0366	3.60%	0.9753	16.57%	0.0537	5.67%
health	0.0215	2.12%	0.0349	0.59%	0.0233	2.46%
Y ₂₀₂₂	0.0339	3.34%	0.0187	0.32%	0.0401	4.23%
残差项	0.5809	57.13%	1.8240	30.98%	0.5331	56.28%

进一步,将外生环境因素的相对贡献加总,可得数字劳动差异中的机会不平等。但与收入机会不平等不同的是,年龄的相对贡献也应纳入数字机会不平等中。由于出生时代和社

会变迁所造成的结构性差距,限制了老年人在数字经济中实现自身能力与价值的可能,反映的是“机会的初始不平等”,而非“结果的不平等”或个人责任问题(Daniels,2007)。基于此,三组分解结果中的机会不平等指数分别为 30.89%、47.82%和 29.14%。相比于广义数字劳动和无酬数字劳动,有酬数字劳动中的机会不平等程度远高于其他两者,劳动要素数字化转型呈现较为突出的“有偏”特征。

3.收入机会不平等的效应分解

依据收入差距和数字劳动差距的分解结果,可将外生环境差异导致的收入差距分解为直接效应和间接效应,进一步加总即可得到机会不平等总指数,具体如表 8 所示。由效应分解结果可知,若将广义数字劳动作为核心解释变量,收入机会不平等的直接效应、间接效应和总效应分别为 25.04%、1.99%和 27.03%。若将数字劳动细分为有酬数字劳动和无酬数字劳动,三者分别为 21.27%、10.05%和 31.32%;其中,来自有酬数字劳动的间接效应约占总效应的三分之一。

表 8 基于非线性模型的效应分解结果

模型	直接效应	间接效应			总效应
		T	T_1	T_2	
	(1)	(2)	(3)	(4)	(5)
(1)—(2)式	25.04%	1.99%	—	—	27.03%
(3)—(5)式	21.27%	—	9.18%	0.87%	31.32%

(四)基于行业的异质性分析

如上所述,劳动要素数字化水平在不同行业间呈现出明显的结构性差异。鉴于此,依据《职业分类与代码(GB/T6565-2009)》,按行业开展收入差距分解。为避免估计结果不稳定,对样本量较少的行业予以剔除,分解结果如表 9 所示。整体上,分解结果呈现出明显的行业异质性。就广义数字劳动而言,其对服务人员收入差距的相对贡献高达 18.54%,对专业人员与技术人员以及生产、运输设备操作人员及有关人员次之,分别为 10.29%和 7.26%,三者均高于平均相对贡献;对办事人员和有关人员以及农、林、牧、渔、水利业生产人员收入差距的相对贡献较低,分别为 4.92%和 4.73%。细分数字劳动类型发现,有酬数字劳动对各行业收入差距的相对贡献普遍高于无酬数字劳动,占据主导地位。其中,有酬数字劳动差异对服务人员收入差距的相对贡献最高(19.51%),显著高于平均相对贡献;对办事人员和有关人员次之(17.14%),对专业人员与技术人员,以及生产、运输设备操作人员及有关人员的相对贡献分别为 14.97%和 12.98%,对农、林、牧、渔、水利业生产人员相对贡献最低,为 11.79%。

表 9 基于非线性收入方程的分行业夏普利值分解结果

变量	专业人员与技术人员	办事人员和有关人员	服务人员	农、林、牧、渔、水利业生产人员	生产、运输设备操作人员及有关人员
	(1)	(2)	(3)	(4)	(5)
T	10.29%	4.92%	18.54%	4.73%	7.26%
T_1	14.97%	17.14%	19.51%	11.79%	12.98%
T_2	5.14%	2.73%	5.83%	4.34%	4.36%

综合表 6—表 9 的研究结论,本文认为上述结果主要源于数字化转型过程中劳动者禀

赋、就业机会与行业需求的结构性差异。首先,数字劳动具有较高的技能门槛。年龄、教育背景和个人努力程度影响数字技能获取及技术适应能力,使高人力资本劳动者更容易进入数字化程度较高、收入回报较高的岗位,从而形成数字劳动供给的选择性偏向。其次,数字劳动机会存在明显的初始差异。户口、出生地区和家庭教育背景等因素影响数字基础设施、教育资源和就业信息获取,使部分群体在数字化转型中更容易获得有酬数字劳动机会,而弱势群体更多停留于低回报或无酬数字劳动。第三,有酬数字劳动具有更强的收入回报效应。相较于无酬数字劳动,有酬数字劳动能够直接进入劳动市场并转化为工资、奖金等货币收入,其供给差异因而更容易被放大为收入差距。第四,行业数字化需求存在差异。服务业及专业技术岗位对数字技能依赖程度较高,数字劳动能够获得更高的市场溢价,因此数字劳动差异对收入差距的贡献更加突出;而传统生产岗位数字化程度相对较低,影响相对有限。

如前文所述,收入差距的分解结果可能依赖收入函数的设定。因此,本文在稳健性检验中调整收入函数的设定形式和变量数量。稳健性检验结果表明,线性模型的分解结果与非线性方程分解结果基本一致,进一步验证了分解结论的稳健性^①。

五、结论与政策建议

本文将上网时间作为数字劳动的量化指标,评估劳动要素数字化对收入分配的影响,主要结论如下:(1)Bartik 工具变量法与 DML 法的估计结果显示,数字劳动对收入水平具有显著影响;基于 XGBoost+SHAP 方法的归因分析结果表明,广义数字劳动的重要性得分在现有变量中位列第三;若将数字劳动细分为有酬数字劳动和无酬数字劳动,其中有酬数字劳动在现有解释变量中位列第一。(2)数字劳动的参与程度存在显著个体差异。人口统计学特征、原生家庭环境等外生因素与数字劳动参与程度具有较强相关性,说明劳动要素数字化并非随机分布,而是在一定程度上受到个体禀赋的影响,由此形成数字劳动机会获取的不平等现象。(3)基于非线性方程的不平等指标分解结果显示,数字劳动差异对收入差距具有较强解释力,对总体收入差距的贡献为 6.43%,仅次于性别;若细分数字劳动类型,有酬数字劳动的相对贡献值高达 19.19%,在现有解释变量中位列首位。(4)数字劳动差异对收入差距的解释作用存在行业异质性,其中对服务人员收入差距的相对贡献最大。此外,基于线性收入方程的夏普利值分解与基于非线性收入方程的分解结果基本一致,表明实证结论具有较好的稳健性。

以实证结论为基础,本文认为可从以下三个方面规范数字经济时代的收入分配秩序:

第一,提升互联网对低收入劳动者的包容性、降低“数字成本”。对于低收入劳动者而言,能够便捷廉价地使用互联网是其获取数字化转型机会的必要条件,如何提升互联网对低收入劳动者的包容性,是逾越数字鸿沟的必要条件。因此,应持续增加互联网公共基础设施投入,进一步扩大网络覆盖范围、提升网络接入质量,降低劳动者获取和使用数字资源的经济成本与时间成本,增强互联网的普惠性。

第二,提升劳动者的数字素养。劳动者数字化转型的最终目的是利用互联网提升生产

^①线性模型的分解结果参见《经济评论》网站(<http://jer.whu.edu.cn/>)附件。

生活效率,而数字素养是必不可少的条件。从实际情况来看,年龄偏大、受教育水平偏低的劳动者数字素养相对较低,难以适应持续迭代的互联网思维与不断更新的电子设备。对于年轻一代而言,由于数字素养的缺位,电子产品依赖症成为普遍现象,不仅导致虚度光阴,还可能带来价值观念扭曲,影响青少年的健康成长。因此,应将提升数字素养列入国民教育计划,设计数字素养终身教育方案,在提升国民数字技能的同时,塑造正确的数字价值观。

第三,建立健全数字平台的收入分配制度。用户在数字平台上贡献的数字劳动并不都是显性劳动,大量用户贡献的数字足迹并未在其获得的收入中体现,且一般用户并不具有开发数据足迹的机会,由此形成了收入与贡献不匹配、数字平台隐性剥削、平台垄断等问题。因此,有必要围绕数字平台与数字劳动的内在关系,综合运用初次分配和再分配措施,构建面向平台经济的分配制度。一方面,要在初次分配中充分尊重数字劳动者的显性贡献,另一方面,可以考虑使用税收工具,通过再分配政策回馈全社会在平台上的隐性劳动。

参考文献:

1.陈梦根、周元任,2022:《数字不平等研究新进展》,《经济学动态》第4期。
2.黄再胜,2026:《马克思资本循环理论视域下“平台租金说”的问题及其矫正》,《马克思主义研究》第2期。
3.刘波、胡宗义、龚志民,2020:《中国收入差距中的机会不平等再测度——基于“环境-能力-收入”的新思路》,《南开经济研究》第4期。
4.马克思,1975:《资本论(第一卷)》,中译本,人民出版社,第259页。
5.欧阳康、杨德科,2023:《数字劳动的资本积累逻辑探究》,《南京社会科学》第9期。
6.庞跃辉、丁乙,2025:《数字劳动三大前沿问题的哲学思考》,《上海大学学报(社会科学版)》第2期。
7.王宁、胡乐明,2022:《数字经济对收入分配的影响:文献述评与研究展望》,《经济与管理评论》第5期。
8.徐怀宁、田亚男,2025:《数据要素市场化与企业劳动收入份额》,《经济评论》第5期。
9.杨静、郑嵘,2024:《数字劳动的理论化研究新进展》,《经济学动态》第5期。
10.周文、韩文龙,2023:《数字财富的创造、分配与共同富裕》,《中国社会科学》第10期。
11.Acemoglu, D., and P. Restrepo. 2020a. “The Wrong Kind of AI? Artificial Intelligence and the Future of Labour Demand.” *Cambridge Journal of Regions, Economy and Society* 13(1): 25–35.
12.Acemoglu, D., and P. Restrepo. 2020b. “Robots and Jobs: Evidence from US Labor Markets.” *Journal of Political Economy* 128(6): 2188–2244.
13.Aker, J. C., and I. M. Mbiti. 2010. “Mobile Phones and Economic Development in Africa.” *Journal of Economic Perspectives* 24(3): 207–232.
14.Almås, I., A. W. Cappelen, J. T. Lind, E. Ø. Sørensen, and B. Tungodden. 2011. “Measuring Unfair (In) Equality.” *Journal of Public Economics* 95(7–8): 488–499.
15.Bartik, T. J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.
16.Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. 2018. “Double/Debiased Machine Learning for Treatment and Structural Parameters.” *Econometrics Journal* 21(1): C1–C68.
17.Daniels, N. 2007. *Just Health: Meeting Health Needs Fairly*. Cambridge University Press.
18.DiMaggio, P., and F. Garip. 2012. “Network Effects and Social Inequality.” *Annual Review of Sociology* 38(1): 93–118.
19.Fuchs, C. 2014. *Digital Labour and Karl Marx*. New York: Routledge.
20.Johnson, S., and D. Acemoglu. 2023. *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. London: Hachette UK.
21.Katz, V. S., and C. Gonzalez. 2016. “Toward Meaningful Connectivity: Using Multilevel Communication Research to Reframe Digital Inequality.” *Journal of Communication* 66(2): 236–249.

- 22.Lefranc, A., N. Pistolesi, and A. Trannoy. 2009. "Equality of Opportunity and Luck: Definitions and Testable Conditions, with an Application to Income in France." *Journal of Public Economics* 93(11-12): 1189-1207.
- 23.Mendonça, P., and N. K. Kougiannou. 2023. "Disconnecting Labour: The Impact of Intraplatform Algorithmic Changes on the Labour Process and Workers' Capacity to Organise Collectively." *New Technology, Work and Employment* 38(1): 1-20.
- 24.Ramdas, K., and A. Sungu. 2024. "The Digital Lives of the Poor: Entertainment Traps and Information Isolation." *Management Science* 70(12): 9089-9100.
- 25.Robinson, L., S. R. Cotten, H. Ono, and A. Quan-Haase. 2015. "Digital Inequalities and Why They Matter." *Information, Communication & Society* 18(5): 569-582.
- 26.Roemer, J. E. 1998. "Why the Poor Do Not Expropriate the Rich: An Old Argument in New Garb." *Journal of Public Economics* 70(3): 399-424.
- 27.Szeles, M. R. 2018. "New Insights from a Multilevel Approach to the Regional Digital Divide in the European Union." *Telecommunications Policy* 42(6): 452-463.
- 28.van Deursen, A. J. A. M., A. van Der Zeeuw, P. De Boer, G. Jansen, and T. van Rompay. 2021. "Digital Inequalities in the Internet of Things: Differences in Attitudes, Material Access, Skills, and Usage." *Information, Communication & Society* 24(2): 258-276.
- 29.van Deursen, A. J., and J. A. van Dijk. 2019. "The First-Level Digital Divide Shifts from Inequalities in Physical Access to Inequalities in Material Access." *New Media & Society* 21(2): 354-375.

Digitalization of Labor Factors and Income Distribution: An Empirical Study Based on Machine Learning Methods

Liu Bo¹, Li Yunqi² and Zhang Lei³

(1: School of Economics, Hunan Agricultural University;

2: School of Economics and Trade, Hunan University;

3: Business School, Xiangtan University)

Abstract: In recent years, China has witnessed a divergence between declining paid working hours and continuously rising income levels, while the substantial increase in time spent online offers a new perspective for understanding this phenomenon. This study conceptualizes the digitalization of labor factors from both broad and narrow perspectives and employs machine learning methods to systematically examine the impact of digital labor on income distribution. The results show that digital labor significantly increases workers' income levels. Attribution analysis indicates that paid digital labor has the highest feature-importance score among the explanatory variables. The results of inequality decomposition show that disparities in digital labor have substantial explanatory power for income inequality among workers, with differences in paid digital labor making the largest relative contribution. Further analysis reveals that the digital transformation of labor factors is characterized by inequality of opportunity, with demographic characteristics and family background shaping disparities in the supply of digital labor. These findings identify the pathways through which the digitalization of labor factors affects income distribution and provide theoretical foundations and empirical evidence for improving the income distribution system in the era of the digital and intelligent economy.

Keywords: Digital Economy, Digital Divide, Digital Labor, Income Distribution, Inequality of Opportunity

JEL Classification: D31

(责任编辑:陈永清)