Advances in Social Sciences 社会科学前沿, 2024, 13(6), 332-337
Published Online June 2024 in Hans. https://www.hanspub.org/journal/ass
https://doi.org/10.12677/ass.2024.136504

Hans汉斯

人工智能产业中数据标注众包劳动的法律检视

——基于国内典型平台的分析

魏　凤，徐钰涵，张天启，王　澈

天津工业大学法学院，天津

收稿日期：2024年4月17日；录用日期：2024年6月18日；发布日期：2024年6月27日

摘　要

随着人工智能产业的高速发展，数据标注众包劳动成为平台零工经济下新的密集型就业途径，但这一业态也成为法律规制的薄弱地带。立足于国家人工智能发展战略，通过对典型数据标注平台的追踪调研，总结其众包劳动的运行模式并进行实效评估，在检视问题与完善策略的基础上，提出细化的体系架构，形成工作派单、风险评估和劳动奖励的内部循环系统，并通过保护、促进和规划的方式形成外部运行机制，最终促进人工智能产业健康发展。

关键词

人工智能，数据标注，众包劳动，劳动关系

Legal Review of Crowdsourcing Labor for Data Annotation in Artificial Intelligence Industry

—Analysis Based on Typical Domestic Platforms

Feng Wei, Yuhan Xu, Tianqi Zhang, Che Wang

Law School, Tianjin University of Technology, Tianjin

Received: Apr. 17th, 2024; accepted: Jun. 18th, 2024; published: Jun. 27th, 2024

Abstract

With the rapid development of the artificial intelligence industry, data annotation crowdsourcing

文章引用: 魏凤, 徐钰涵, 张天启, 王澈. 人工智能产业中数据标注众包劳动的法律检视[J]. 社会科学前沿, 2024, 13(6): 332-337. DOI: 10.12677/ass.2024.136504

labor has become a new intensive employment path in the platform gig economy, but this industry has also become a weak area of legal regulation. Based on the national artificial intelligence development strategy, through tracking and researching typical data annotation platforms, summarizing their crowdsourcing labor operation modes and conducting effectiveness evaluations, on the basis of examining problems and improving strategies, proposing a refined system architecture, forming an internal circulation system for work dispatching, risk assessment, and labor rewards, and forming an external operating mechanism through protection, promotion, and planning, ultimately promoting the healthy development of the artificial intelligence industry.

Keywords

Artificial Intelligence, Data Annotation, Crowdsourcing Labor, Labor Relations

1. 行业发展背景及意义

ChatGPT (Chat Generative Pre-trained Transformer)的发布与爆火，把人工智能行业推入了一个新的风口。全球不少研究机构和媒体以及企业领导发布了关于人工智能大规模发展的预测，认为接下来即将进入生产式人工智能生产及应用的爆发时代。

瑞银集团曾在报告中表示，截至 2027 年，人工智能在各经济体的广泛应用预计将使其成为一个规模达 2250 亿美元的市场。与 2022 年的 22 亿美元相比，这将是一个巨大的飞跃，标志着约 152%的复合年增长率。人工智能行业的收入将随之增长 15 倍，从 2022 年的 180 亿美元增至 2027 年的 4200 亿美元，相较瑞银此前预期的上调了 40% [1]。由此可见人工智能在科技水平持续不断发展的当下，会逐步成为新时代的核心与焦点，而人工智能的不断发展对于算力的要求也在逐步增高，算力是将人工智能领进实际应用的主要载体。经过测算显示，当计算力指数平均每提高 1 点时，国家的数字经济和 GDP 将分别增长 3.6%和 1.7% [2]。算力提升的投入要远远超过对于数据人为精确的投入，并且就当下而言对于算力、算法的不断提优，远远没有对数据不断提优所获得的效益大。因此便衍生出了对于数据不断精确的“人工”，数据标注师。

于此同时，数据标注师也是人工智能产业发展过程中催生出的“新型流水线工人”，在我们日常生活中所能接触到的语音助手、智能音箱、线上客服、人脸识别、智能家电等一系列人工智能产品都离不开数据标注师的功劳，我们的手机软件、系统、智能家居等一系列产品的更新也依仗于数据标注师对数据进行的不断优化与精确。但是其时薪大致在 10 元至 30 元之间，可以说是处于中等偏下的水平，而预估需要人工标记的图片高达 12 万张，平均每张图片有 44 个图层，每个图层的标注时间需要 14 秒。总体计算之下，人工标注的时间需要 20,533 个小时。如果聘用一百人团队每天工作 9 个小时，这需要 23 天[3]。这一系列数据表明，在人工智能市场中对于数据标注师的需求是巨大的，然而将其称之为“新型流水线工人”，正因其数据标注工作是重复且乏味的，每天要面对同样一台设备做着重复的动作。有时甚至可能要面对一些最原始的暴露且暴力的数据，长时间的低薪、枯燥、伏案式的工作无疑对数据标注师的身心都造成了一定程度的伤害。

因此在市场需求量巨大与收入微薄的两相比较之下，数据标注行业存在着极大的问题的同时也在不

断朝着规范化与规模化发展，外加伴随着越来越多的软件及日常生活用品人工智能化，对于数据标注的需求量呈直线上升趋势，目前数据标注产业仍在持续不断高速发展，在未来的五年当中年复合增长率在27%左右，产业规模有可能达到百亿[4]。由此看来此行业具有一定的研究价值。

2. 分析调研

2.1. 实践调研

网上的数据标注众包平台众多，从中选取了 4 个较为有名的数据标注众包平台进行实践与分析，分别为京东微工、百度众测、腾讯搜活帮、有道众包。

1) 京东微工

其推广为企业客户、高校及科研院校提供人工智能大数据采集标注服务，其主攻数据采集、数据标注与数据清洗三个方面。而通过我实践所得，京东微工中发布的占绝大多数的是大学生和自由职业者在线求职的信息，几乎寻找不到企业所发布的任务，并且每笔订单并没有风险评估程序，劳动者需要自己抢单且求职困难，并未真正做到数据的大量收集，保证所需数据量。

2) 百度众测

百度众测中主打连接众测测友、服务商和数据需求方，其中平台发布的任务少且单一，大部分是通过工会将劳动力集结起来，数据需求方向工会发布任务，平台进行质检来完成数据标注的工作。经过实践后，发现在工会中全职答题人员平均每小时可以答 200~300 道数据标注题，每道题 8 分钱，时薪大约在 20~30 元之间，印证了 21 世纪经济报道中的观点。此外百度众测对整体任务的质检要扣除总任务费用的 10%，实际到劳动者手上的钱只在 90%，经过计算全职劳动者在一天之中工作 9 个小时且周末不休息的情况下，月薪可达到 8000~9000 元，但是这对劳动者的身体伤害是巨大的，并且劳动者不存在任何保障。而对于兼职人员来说，数据标注众包的任务量大、单一且每笔订单价格低廉，这无一不是造成他们退出的因素。

3) 腾讯搜活帮

腾讯搜活帮主要推广用闲暇时间，在线赚奖励，是一个线上兼职平台，其增添了一些入门以及专业教学课程，方便新手更快上手。于此同时在任务页面也属于抢单模式，需要劳动者去随时关注着任务界面，在任务开始之前有相关方面的测验测试任务者是否具备进行数据标注的能力。但是任务量很匮乏且大多数任务为无报酬任务，且平台对于数据的审核时间过长，导致数据标注师花费大量时间却无法得到相应的劳动报酬。每个任务收益只在 0.2~1 元之间不等，收益排行当中每周最高收益者只能勉强达到 1000 元的收入，每个月最多只可获得 3000~4000 元的收入，并且工作时间远远超过了兼职甚至是全职会花费的时间。

4) 有道众包

有道众包类似于一个线上公司，对于劳动者分为实习、兼职与全职三部分进行管理，对于劳动者的使用有一定的考核与标准，不具备手机软件，因此电脑对于做数据标注任务是必要的。其中不仅仅包含数据任务，还包含着文章转写、翻译、采集及选择判断的任务。但是每个月报酬最高者仅为 1600 元不到，这大大降低了兼职与全职数据标注师的劳动积极性。

2.2. 社会调研

通过观察社会调研所获得的 200 份有效调查问卷，其中 98.45%的人觉得数据标注众包劳动具有工作量大、收入低、相关方面法律保护不完善的问题，56.1%的人觉得数据标注众包劳动具有随时被逃单的风险，并且对新人很不友好。但是仍然吸引着他们去劳动的原因一大部分是数据标注众包劳动的工作时间

自由、工作量自定、适合当代社恐人群，而少部分的人则是想要拥有一份兼职和学习一些新的技能。

在这其中他们也给出了相应的结论 75.06%的人群认为，平台应当对数据标注众包的任务进行一定的风险评估，89.22%的人群希望能在工作量达到一定数额时获得一些额外的奖励费用。最后询问他们是否担心数据标注众包工作最终有一天会被机器所替代时，他们始终坚信着人工智能起源于人类、发展依靠于人类，人类对于数据标注众包工作在某种层次上而言无法被替代。

实际上，当前的智能体往往都是使用数据分析人员提供的数据，智能体自身的数据处理能力还是比较薄弱的，而且智能体对于数据价值的判断能力也比较弱，这些都需要数据分析人员来进行标注和处理。从当前人工智能技术的应用情况和创新方式来看，未来较长一段时间内，数据分析领域不仅不能被人工智能技术所取代，反而需要大量掌握大数据和人工智能技术的专业性人才。

3. 当前行业困境

3.1. 劳动关系难以认定

首先是数据标注众包劳动的劳动关系难以确定，且社会保障欠缺。平台企业通常将数据标注师定义为独立承包商。《关于确立劳动关系有关事项的通知》(劳社部发[2005] 12 号)第一条规定，(二)用人单位依法制定的各项劳动规章制度适用于劳动者，劳动者受用人单位的管理，从事用人单位安排的有报酬的劳动[5]。数据标注师的灵活性工作特点使其不全部受单位的约束管理，而是凭借个人的技能能力自行承担经营风险，因此认定不存在劳动关系。因此，依据目前的劳动关系认定标准，对众包平台新业态的用工关系缺乏足够的解释力。一些平台企业为了避免与从业人员建立直接的劳动关系，刻意规避劳动法律规定，借助平台的强势地位，采取了各种办法。

因此，无法以职工身份参加医疗和养老保险，保险费用全由个人承担，导致经济负担较重。同时，劳动者对众包平台依赖程度越高，社会保障享受越少。

3.2. 劳动报酬水平不容乐观

数据标注众包劳动的收入从实证得知是很不可观的，大部分是低于居住地最低工资水平。平台众包零工新业态适应高速发展的互联网时代，对数据标注师的需求不断增长，接单也是大多数劳动者的主要收入来源。据调查，2018 年以来，随着众包接单者的增加，尤其是层层转包对利润空间的畸形挤压，原本不高的数据标注接单收入持续下降。以科大讯飞旗下“爱标客”众包平台为例，它目前大部分项目时薪最多只有 10 到 15 元，有时可能连 10 元都不到。造成工作总量和劳动报酬不对等的现象，数据标注师更期盼公平的薪酬。

3.3. 众包平台存在技术风险

在相关访谈中，有人表示，平均每 8 到 10 次就有 1 次，工作半途出问题，如网络中止、网页打不开。众包平台的技术问题所带来的风险常常导致数据标注师得不到劳动报酬，甚至要重新完成工作，大大增加了工作量，然而众包平台并不能对此作出合理的解决措施，数据标注师受到众包平台和发单者的双重压力。

发单企业或发单者个人虚假支付，具有信用风险。在众包平台上，数据标注师所完成的任务经常会受到任务发布者的不合理拒绝。调查显示，“近 90%的受访者报告他们提交的任务曾被拒绝支付”，“仅有 12%的数据标注师认为自己经历的所有拒绝支付都是合理的”。此外，平台的用户对于劳动者的不良评价将影响劳动者获得新工作的可能性，而用户并不需要为作出不良评价而给出具体理由，平台也未给提交不良评价的内容是否真实进行核实，并且没有给数据标注师提供任何有效抗议或申诉的渠道。

4. 解决方法与措施

针对研究所发现的问题，我们逐一进行思考并做出了相应的解决机制：

1) 多方位促进数据标注师的劳动关系认定

在法律保护方面，司法实践中应当扩宽对数据标注师的劳动关系的法律标准认定范围，调整“有劳动关系才社保”的政策思路，突破现行将劳动关系作为社保门槛的传统理念，例如，平台对全职数据标注师提供工伤保险和生育保险等，从而明确平台零工劳动中的法律主体关系、厘清主体责任，规范众包平台和数据标注师的行为，为数据标注众包行业提供新型劳动关系下的法律保护。

在市场促进方面，与数据标注师连接最为紧密的就是众包平台，这就需要市场的促进。对于众包平台，应当鼓励平台采取家政业的“员工制”用工模式，明确平台用工法定保障义务，把工伤保险参保等内容逐步纳入平台用工的必要条件，才能充分释放平台零工劳动的潜力，为数据标注师的劳动权益提供最优的保障。

在产业规划方面，数据标注众包平台需要一个合理的产业规划和行业准则。数据标注行业的规范化，可以促使层层分包的现象减少，数据标注师作为人工智能发展下的大趋势，产业规划对于选择投入该行业的人来说是一个至关重要的参考点。众包平台的产业规划应该进一步规避市场竞争下，不良业态给数据标注师带来的劳动权益风险。

2) 内外结合实现数据标注师劳动所得提升

从平台内部而言，将“抢单”模式转化为“派单”模式，对于数据标注众包平台常见的抢单模式，十分考验数据标注师的网速与手速，而错过订单后又要开始新一轮漫长的等待接收订单的过程。时常导致一部分数据标注师无单可做，而一部分订单过多堆积使得数据标注师疲劳且订单完成效率低下。

因此，为了优化数据标注众包模式下的劳动时间，实现真正的自主性。在平台设置中，数据标注师可以针对自己的具体情况(工作领域、工作时长等)选择自己的工作状态及工作范围，平台根据大数据统计对数据标注师进行工作分配，有效减少无效的工作时长。平台应当为数据标注师设立三种工作状态：上线服务中、上线空闲中和下线。当数据标注师登录平台时，平台能及时为数据标注众包人推送订单消息；当数据标注师接受订单时，平台应立即显示其为上线服务中状态，且在此时间段内不再为该数据标注师分配工作任务；订单结束后，平台接受到其为上线空闲中状态，则继续为该数据标注师分配工作任务。这样一个机制就能最大程度地保障有效工作时长，同时也能平衡数据标注师的工作与生活。在派单的择优机制下，数据标注师就能够接收到渡河自己工作领域，复合自己工作时间的订单，实现平台与数据标注师“双赢”的模式。

从外部劳动者所得而言，平台可以建立相应的奖励机制。在传统雇佣关系中，劳动者只需遵循固定的工作时间，在工作场所按照雇主的指令从事有偿的劳动，并接受雇主的监督和激励。但在平台经济中，数字化管理可以将数据标注师的工作时间精准化，众包平台作为发单者和接单者的中间人，当然希望数据标注师可以高时、高效、高质量的完成发单者的订单任务，进而得到发单者的信任，形成良好的消费形态，众包平台才可以获得更多的利润。因此，通过平台累计一天工作时长，工作完成效率及质量的数据分析，平台可以对高时并且高效完成数据标注任务的数据标注师提供一些奖励，例如连续一周完成订单数达到一定数额，可以获取 200~500 不等的红包奖励。此机制可以鼓励数据标注师塑造“准时”“快速”的劳动时间感，同时也在一定程度上提高了数据标注师的收入水平。此外，众包平台也可以根据用户评价为数据标注众包人提供奖励，提高数据标注师的信用水平以及工作水平。

3) 建立发单者信用和风险评估体系促进安全性有效提升

为解决因发单者的信用问题，避免虚假支付，我们提出在平台内设置发单者的信用级别机制，例如，

发单者信用可设置为三级，级别越高表明发单者信用越好。当数据标注师遇到发单人虚假支付、骗取标注成果的行为时，数据标注师有权通过举报机制，对发单人的行为进行举报，经平台监察核实确认后，平台应对发单人信用级别采取降级制度公开并要求发单人支付报酬。当发单人再次发布数据标注时，数据标注师的接单平台就会显示出该订单的发单人信用级别，对信用极差的发单者，平台会作出警告提示，进而自主选择是否接受订单，是否承担风险。

除此之外，众包平台应当设有自己独立的风险承担责任机制，且应公开在平台登录时的用户须知中。对发单者虚假支付等行为，平台无法为数据标注师追回劳动报酬的，众包平台应当自行承担赔偿责任，以此保障数据标注师的劳动报酬。该机制下不仅可以加强众包平台对低信用度的发单者的监管，降低管理风险，还能更好的保障数据标注师的劳动权益。

5. 结论

在这个科技高速发展的风口，数据标注师的行业前景拥有着较好的趋势，于此同时也推动着数据标注师的收入与法律保障发展至与之相匹配的高度。该研究针对数据标注众包劳动所提出的解决方式，具有一定的合理性与可实施性，但是真正的落实该问题的解决不能只限于纸上谈兵，要多方合作、共同促进该行业的可持续性发展与建设更加美好的就业前景。面临诸多难题，如何实现内部运行机制与外部管理机制协调统一，共同促进数据标注行业的良性发展，是该研究的最终目标。

参考文献

[1] 张薇. 关于人工智能的 60 条趋势预测[EB/OL]. http://www.cbdio.com/BigData/2024-01/08/content_6176209.htm, 2024-01-17.

[2] 金观平. 夯实算力高质量发展基础[N]. 经济日报, 2023-10-17(001).

[3] 江月. AI 催生“数据标注员”需求数据质量或释放更多价值[N]. 21 世纪经济报道, 2023-04-14(012).

[4] 黄爱林. 未来五年数据标注行业人才缺口或达百万[N]. 四川日报, 2023-12-04(010).

[5] 中华人民共和国人力资源和社会保障部. 关于确立劳动关系有关事项的通知[EB/OL]. http://www.mohrss.gov.cn/ldgxs/LDGXzhengcefagui/LDGXzyzc/201107/t20110728_86296.html, 2022-07-20.