MCE BUSINESS RESOURCES

The Hidden Workforce

Big Tech's Debt to Data Labelers

BY MATTHEW MCMULLEN

Even as AI takes industries by storm, the workers whose labor provides the data remain in the shadows.

Every day, headlines trumpet new and astonishing developments in artificial intelligence (AI). These achievements include digital voice assistants such as Alexa and Siri and the latest sensation, ChatGPT-4. However, as we applaud this meteoric rise of AI, we overlook the unsung heroes fueling this revolution—the vast workforce tirelessly annotating the data that ensures our algorithms can think, learn, and adapt.

The process of data annotation involves transforming our rich and contextual understanding of everyday items into numerous illustrative examples so that AI can grasp the abstract concepts that seem intuitive to us. Simplifying our complex world for AI consumption is no easy feat. A "shirt" to us is laden with context and cultural understanding. For AI, it's an abstract concept that requires countless examples to decipher.

Moreover, the granularity of instructions needed often borders on the surreal. A layperson might laugh at the thought of a 43-page guide on labeling shirts. But in the world of data annotation, such depths are imperative. Everyday quandaries become monumental tasks, from identifying a shirt's primary hue to differentiating between a bowl's utility and aesthetic value.

DIVING INTO THE SHADOW WORLD OF DATA VENDORS

When we think of AI, the tech giants immediately come to mind. But behind these luminaries are lesser-known entities powering the AI engine. Data vendors operate in places such as Kenya, Nepal, and India. They function like a traditional call center, managing complex projects with teams of hired workers. On the other hand, platforms act as digital bazaars, where gig workers peddle their data-labeling prowess.

Often, successful AI vendors shield their operations behind a cloak of secrecy, because the methods of annotating data reveal too much about their treatment of workers. And, as we marvel at the success stories of the tech giants, we must also reckon with the veiled nature of the AI industry. While the opacity safeguards proprietary AI initiatives, it can also mask a questionable objective, such as exploiting annotators by offering

This article was originally published in the MCE Quarterly!

subpar compensation for detailed, rigorous work, showcasing a more unscrupulous pursuit of a competitive advantage within the AI world. But whispered industry insights reveal a staggering fact: Potentially millions of workers labor in the shadows, perhaps a billion, annotating our AI futures.

Data labeling extends far beyond generative AI such as ChatGPT. It underpins various AI applications, fueling the global data collection and labeling market, projected to reach $47 billion by 2030, according to a November 2023 report by KBV Research ("Global Data Labeling Solution and Services Market Size, Share & Industry Trends Analysis Report By Type, By Labeling Type, By Sourcing Type, By Vertical, By Regional Outlook and Forecast, 2023 – 2030"). From training autonomous vehicles to navigating robot vacuums safely, human expertise remains indispensable for AI's progress.

BEYOND THE GLITZ AND GLAMOUR OF AI

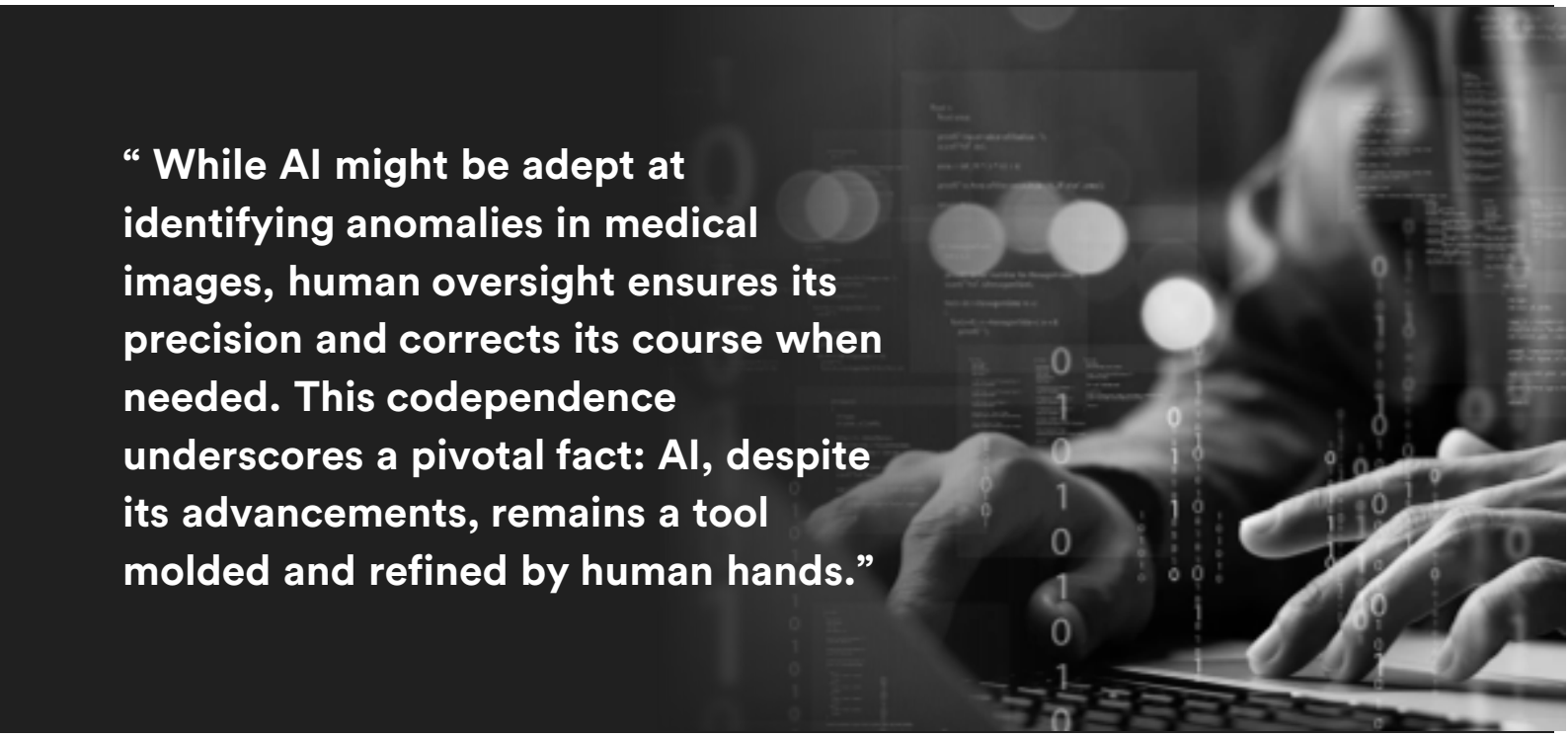
The AI industry, marked by its inviting digital voice assistants and staggering valuations, owes its inception to the tireless labor of human data labelers. Despite being foundational to AI, many developers see data annotation as a fleeting necessity: Gather labeled data, perfect the model, and then move on. Yet, beneath the surface, there's a vast, often overlooked world where data labeling, frequently outsourced to places such as Kenya, employs workers for wages as low as $2 an hour. These data workers' pivotal role in shaping AI cannot be understated. But tragically, the industry often marginalizes and undervalues them.

Recent scrutiny over data labeling and annotation practices has drawn attention to ethical vulnerabilities, especially concerning the exploitation of low-wage workers in developing countries. These workers, often serving as the backbone for data labeling, may lack a comprehensive understanding of their rights and the implications of their contributions.

ADDRESSING THE ILLUSION OF AI'S BRILLIANCE

As we marvel at AI's breathtaking capabilities, we must acknowledge its inherent limitations. While AI can imitate human discernment, thinking, and processes, its performance hinges on the quality of the data it receives. Let's dispel the illusion of AI's brilliance—it leans heavily on human-crafted data. The old adage "garbage in, garbage out" holds true for AI applications, as its output reflects the quality of the data it processes. Large language models (LLMs) may generate seemingly genuine information, but this façade can crumble when faced with deceptive inputs, demonstrating AI's lack of true cognition and emphasizing the importance of human intervention in its training.

Data labelers and content moderators aren't just contributors, and to undervalue their role isn't merely an oversight—it's an affront to fair labor and a blind spot in our grasp of AI ethics and constraints. We must not only acknowledge but fervently champion their indispensable contributions.



“ While AI might be adept at identifying anomalies in medical images, human oversight ensures its precision and corrects its course when needed. This codependence underscores a pivotal fact: AI, despite its advancements, remains a tool molded and refined by human hands.”

DECODING THE HUMAN-AI TANDEM

Times have changed. Not long ago, professions such as radiology trembled at the thought of AI usurping human roles. The reality today paints a more collaborative picture. While an AI might be adept at identifying anomalies in medical images, human oversight ensures its precision and corrects its course when needed. This codependence underscores a pivotal fact: AI, despite its advancements, remains a tool molded and refined by human hands.

Even the pandemic has been a testament to AI's inherent limitations, with machine-learning models, initially perceived as self-sufficient entities, showing cracks under the unanticipated surge in divergent human activities and preferences, according to Will Douglas Heaven in the May 2020 *MIT Technology Review* (“Our Weird Behavior During the Pandemic Is Messing with AI Models”). These unforeseen circumstances have emphasized the crucial role humans play in fine-tuning and recalibrating these models, ensuring their continued relevance and accuracy. It is through this mutualistic tandem that the full potential of AI can be realized, allowing for a future where technology and humanity co-evolve, leveraging each other's strengths to navigate through uncertainties and complex times.

THE ROCKY TERRAIN OF COMPENSATION

Given the intricate nature of the work of data labelers and the skill it takes to perform such tasks, we encounter another pressing concern: the precarious nature of data labelers' earnings. Their tasks vary wildly, from minutes-long assignments to hours of meticulous work. The nebulous nature of compensation, exacerbated by global disparities and shifting demand, casts a shadow over the profession.

International guidelines and emerging standards such as the International Labour Organization (ILO) and the ILO Tripartite Declaration have made strides in advocating for worker rights, fair wages, and benefits. These guidelines call for businesses, including data vendors, to implement ethical practices that ensure workers are compensated fairly, have the freedom to associate and collectively bargain, and are provided with social security and a safe working environment.

However, a glaring discrepancy exists between these well-intentioned guidelines and the on-the-ground reality faced by laborers. When workers' efforts to secure better working conditions were not only ignored but actively suppressed, it exposed the limitations of existing international standards. According to a July 2023 article in *Time*, “Gig Workers Behind AI Face 'Unfair Working Conditions,' Oxford Report Finds,” rather than engaging in dialogue when faced with a potential strike, certain companies chose to exercise their power to quell dissent. High-level executives were flown in to deal with the situation, and the leader of the strike was summarily fired for allegedly putting business relationships at “great risk.”

The message is clear: The laborers are expendable, and their demands for fair treatment are secondary to any company's business interests. This high-handed approach not only undermines intentional guidelines but also sends a discouraging message to employees everywhere, questioning the efficacy of collective bargaining and the quest for better working conditions.

While international guidelines are a step in the right direction, enforcement and accountability remain significant hurdles. Regulatory bodies, clients, and the public must ensure that companies don't just pay lip service to these standards but

This article was originally published in the MCE Quarterly!

implement them in spirit and action. Our collective goal should be an industry where international standards aren't merely checkboxes to tick. Rather, we are all holding companies accountable for lapses.

BOUND FROM EXISTENCE

Data vendors maintain strict control over workers' ratings, reviews, and feedback, according to Seats2Meet.com's Martijn Arets in "Research on Platform Based Reputation Scores Contributes to an Inclusive Labor Market." Such monopolized ownership of information not only deprives workers of their professional reputation but also can hinder their mobility and career advancement. If workers decide to transition away from their current employment, they would effectively be starting from scratch, losing all the reputation and credibility they built over time.

The same laborers who are advancing AI technologies often find themselves caught in contracts that virtually bind them out of professional existence. The crux of the issue lies in the strict contractual agreements and terms of service that these workers are compelled to sign. The contracts often grant the data vendor exclusive ownership over critical components of a worker's professional identity. The loss of control over this professional data handicaps these workers, leaving them unable to negotiate better terms, seek higher-paying opportunities, or even defend against sudden account closures. This creates a perilous cycle where laborers have limited career mobility and advancement prospects, confining them to a data vendor's reputation and ability to provide work. These contractual constraints thwart immediate career advancement. In an age where personal branding and reputation are increasingly vital, these workers face the unnerving reality of existing as mere cogs in a machine, unable to claim ownership of their own work histories and achievements. As a result, they are virtually bound out of a broader professional existence, limited to the walled gardens of the data vendors they work for.

Further adding to the cloak of secrecy, service providers frequently use code names for clients, making it difficult for data labelers to even know for whom they are providing services. This practice denies them the opportunity to ask for references, network within their industry, or use their experience in a meaningful way to advance their careers.

The contractual constraints workers face don't just stifle individual career growth; they have far-reaching implications for the socioeconomic development of entire regions and generations. When a significant portion of the workforce is engaged in work that does not allow for career mobility, skill development, or fair negotiation of wages and working conditions, the long-term impact is a stagnating labor market.

Moreover, when an entire generation of workers is tied to such restrictive contracts, the collective skills and professional capabilities of that generation can become narrowed and underutilized. Without the ability to grow professionally, transfer skills, or shift careers, these workers are often stuck in a loop of low-wage, low-skill jobs, which in turn can create a "brain waste" scenario. This undermines national competitiveness and productivity, depriving economies of fully engaged workers who can innovate and lead.

A CALL FOR RECOGNITION AND REFORMS

As we stand at the crossroads of AI's potential and ethics, shifting our narrative is paramount. AI is not solely a product of cutting-edge algorithms and vast datasets; it's a testament to human dedication and ingenuity. The fervor surrounding AI's capabilities should be complemented by a genuine acknowledgment of the individuals laboring behind the scenes. It's imperative to champion their rights, ensure fair compensation, and highlight their indispensable role in this digital renaissance.

In light of this comprehensive review, it is clear that data labelers face a plethora of challenges that demand urgent attention. Enterprises that engage AI vendors have a critical role to play in championing labor standards, ethical sourcing, and sustainability. The path to reforms starts with these businesses deliberately choosing service providers committed to fair labor practices, from transparent pricing and wages to robust benefits and support programs.

Another area that warrants immediate action is the recognition and professional development of data labelers. These workers are not just cogs in a wheel, but vital contributors whose skills are becoming increasingly specialized. The existing system, which often marginalizes them and hinders their career mobility, is unsustainable.

A call for recognition must coincide with action: Systems for transparent and portable reputations, fair wages based on experience, and genuine opportunities for skill development and career mobility must be instituted. The AI odyssey is as much about the human spirit as it is about technological prowess. As we propel into an AI-infused future, it's our collective responsibility to ensure that this journey champions innovation and the dignity and value of every contributing individual. It's crucial to view these calls for reform not in isolation but as interconnected elements that contribute to a worker's overall well-being and job satisfaction. Whether it's dealing with sensitive content or maintaining workforce diversity, the time for piecemeal solutions is over. We need a comprehensive approach to reform that addresses the varied but interconnected challenges faced by data labelers. AQ

Matthew McMullen, senior vice president and head of corporate development at Cogito Tech, drives key technology partnerships, seeks technology alliances that elevate our human annotators' service delivery, and crafts policies for responsible AI growth.

MCE KEY FACTS

MCE is your reliable partner for continuous success with agile people development solutions.

10,000,000

participants on AMA & MCE programmes in the last 10 years

92%

of Fortune 1,000 companies are our business partners

96%

of participants report they are using what they have learnt at AMA & MCE

1,000+

expert facilitators globally

100+

Open Training Programmes running throughout EMEA

98

year's experience working with our clients around the globe

For more information please contact:

+32 2 543 21 20 @ info@mce.eu

Visit mce.eu