

# THE USE OF WORKERS-PRODUCED DATA IN ARTIFICIAL INTELLIGENCE-BASED SYSTEMS

*Rômulo Soares Valentini* \*

**Abstract:** This paper delves into the problem caused by the development of artificial intelligence-based systems trained on data sets built by harnessing data collected from employees' work activities. The convergence of technology and workers-produced data prompts a critical examination of its implications within the legal standards framework regarding safeguarding workers against automation and respecting workers' data privacy. The analysis underscores the need to establish legal safeguards for workers, proposing elements for building a legal framework to ensure that companies cannot exploit the data generated through workers' labor activities to train AI systems without obtaining proper consent and providing fair compensation for those workers.

**Keywords:** machine learning; language learning models; artificial intelligence; data privacy; labor law; automation; intellectual property.

## INTRODUCTION

Integrating artificial intelligence (AI)-based systems into workplaces raises significant concerns related to labor laws and employee rights. One of the foremost legal issues affected by those new technologies is the protection of workers against automation. The convergence of AI with data-driven technologies has enabled the development of systems that, by learning from extensive datasets, can automate tasks previously performed by human workers.

A critical aspect of this issue is that the success of automating such tasks hinges on AI systems utilizing datasets comprised of prior inputs and knowledge on how to generate optimal outputs. However, these datasets are often created from previous data related to human labor, frequently performed under employment contracts. The legal problem emerges when employers or third-party companies can exploit the data extracted from employees' work —without proper consent or offering fair compensation for the use of that data —to train and enhance AI systems.

This paper seeks to establish the legal significance of regulating the development and use of work-related data produced by employees. It focuses on addressing the following questions: (1) How can modern AI systems pose a threat to workers' rights? (2) What forms of regulation and legal reasoning

---

\* Ph.D in Labor Law from Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil. Email: [rvalentini@gmail.com](mailto:rvalentini@gmail.com) / ORCID iD: <https://orcid.org/0000-0003-4340-9917>

are necessary to protect these rights, given the current state-of-the-art advancements and capabilities of AI technologies?

The methodology employed is primarily cognitive-expository, aiming to elucidate how AI systems are trained using machine learning algorithms and datasets that encapsulate workers' practical knowledge. This approach will demonstrate the legal implications of companies collecting and processing such data without proper consent or adequate compensation for the labor that created it.

This paper is structured into four chapters to explore the multifaceted dimensions of this issue. After the introduction, Chapter Two delves into the mechanics of technological development and the codification of labor processes. Chapter Three examines alternatives to establish a legal framework to safeguard workers from automation and the unrestricted collection of their work data for training AI systems. The final chapter presents proposed conclusions and recommendations.

## I. ARTIFICIAL INTELLIGENCE AND CODIFICATION OF HUMAN LABOR

### *A. Information Technologies and Artificial Intelligence*

The word *machine* originates from the Doric Greek *μαχανά* (*makhana*) and the Ionian *μηχανή* (*mekhane*), which, in turn, derive from *μηχος* (*mekhos*), meaning "means, expedient, solution"<sup>1</sup>.

Unlike a regular tool, which serves as an extension of human faculties and energy to enhance efficiency and performance<sup>2</sup>, a machine can be defined as a system designed to perform specific actions when supplied with the appropriate energy source<sup>3</sup>. In this sense, machines are built to address particular problems or tasks. According to Willis:

Every machine is constructed for the purpose of performing certain mechanical operations, each of which supposes the existence of two other things besides the machine in question, namely, a moving power, and an object subject to the operation, which may be termed the work to be done. Machines, in fact, are interposed between the power and the work, for the purpose of adapting the one to the other<sup>4</sup>

---

<sup>1</sup> Liddell, Henry George, and Robert Scott. *Greek-English Lexicon*. Accessed September 1, 2024. <http://www.perseus.tufts.edu/>

<sup>2</sup> Cf. Latil, Pierre de. *Pensamento Artificial*. Translated to Portuguese by Jerônimo Monteiro. São Paulo: IBRASA, 1959, 31.

<sup>3</sup> Latil, Pierre de. *Pensamento Artificial*. Translated to Portuguese by Jerônimo Monteiro. São Paulo: IBRASA, 1959, 33.

<sup>4</sup> Willis, Robert. *Principles of Mechanism: Designed for the Use of Students in the*

The term *computer* derives from the Latin *computare*, which means "to count, add up, or group together."<sup>5</sup> Today, it refers to electronic or digital computers, a term rooted in *digit*—any number less than ten—which itself comes from the Latin *digitus*, meaning "finger."<sup>6</sup> However, the act of computing—processing data—was originally performed by human workers:

[...] long before any modern desktop computers or tabulating machines, computation was carried out by humans. In fact, before the adoption of tabulating machines, the word computer referred to a person who performed calculation as a profession, and many tabulating machines that came about late on were named using acronyms that ended in "AC," meaning "Automatic Computer," in order to be distinguished from human computers."<sup>7</sup>

In contrast to analog computers, which process information directly from their environment (e.g., thermostats), digital computers<sup>8</sup> function as automatic calculators, operating based on pre-programmed instructions. As such, a computer does not need to be aware of or understand the tasks it performs; instead, it achieves the desired results through a sequence of ordered steps. This step-by-step procedure is known as an *algorithm*, defined as "a systematic process for solving a problem" or "a detailed sequence of actions to accomplish a task"<sup>9</sup>.

In the field of computer science, Thomas H. Cormen et. al defines an algorithm as:

[...] any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output. An algorithm is thus a sequence of computational steps that transform the input into the output. We can also view an algorithm as a tool for solving a well-specified computational problem. The statement of the problem specifies in general terms the desired input/output relationship. The algorithm describes a specific computational procedure for achieving that input/output relationship.

<sup>10</sup>

---

*Universities and for Engineering Students Generally*. London: John W. Parker, 1861, 1.

<sup>5</sup> Law, Edith, and Luis von Ahn. *Human Computation. Synthesis Lectures on Artificial Intelligence and Machine Learning*, no. 13. Morgan & Claypool Publishers, 2011, 1.

<sup>6</sup> Cf. Jacker, Corine. *Man, Memory and the Machine: An Introduction to Cybernetics*. Rio de Janeiro: Forense, 1970.

<sup>7</sup> Law, Edith, and Luis von Ahn. *Human Computation. Synthesis Lectures on Artificial Intelligence and Machine Learning*, no. 13. Morgan & Claypool Publishers, 2011, 1.

<sup>8</sup> Apter, Michael J. *Cybernetics and Psychology*. Petrópolis: Vozes, 1973, 54-56.

<sup>9</sup> Fertig, Cristina, and Marco Medina. *Algorithms and Programming: Theory and Practice*. São Paulo: Novatec Editora, 2006, 13.

<sup>10</sup> Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein.

The effective use of algorithms in programming computers requires clear and precise instructions, as machines operate deterministically, without the capacity to interpret the meaning of those instructions. As Donald Knuth emphasizes:

Instructions given the computer must be complete and explicit, and they must enable it to proceed step by step without requiring that it comprehend the result of any part of the operations it performs. Such a program of instructions is an algorithm. It can demand any finite number of mechanical manipulations of numbers, but it cannot ask for judgments about their meaning. An algorithm is a set of rules or directions for getting a specific output from a specific input. The distinguishing feature of an algorithm is that all vagueness must be eliminated; the rules must describe operations that are so simple and well defined they can be executed by a machine.<sup>11</sup>

Thus, in computer science, algorithms represent a predefined plan of action for computers to follow, enabling them to perform tasks autonomously without the need for further human intervention. The development of sophisticated algorithms and increasingly powerful computers has expanded the scope of tasks that digital systems can undertake, enhancing human work and, in some cases, performing these tasks independently<sup>12</sup>.

This progress led to the emergence of artificial intelligence (AI), a field within computer science focused on emulating human-like thought processes and problem-solving capabilities through advanced algorithms and programming techniques. AI can be defined as “the study of intelligent behavior (in humans, animals, and machines) and the attempt to find ways in

---

*Introduction to Algorithms*. 3rd ed. Cambridge: The MIT Press, 2009, 5.

<sup>11</sup> Knuth, Donald. *The Art of Computer Programming*. 2nd ed. Reading: Addison-Wesley, 1973, 63.

<sup>12</sup> “[...] **as optimizers**. There are many opportunities to leverage machine intelligence to help improve the accuracy and efficiency of human computation algorithms. Machine learning techniques, such as active learning, can help reduce the cost of human computation by choosing only informative queries to ask. [...] **as enablers**. As human computation systems are built to handle increasingly complex tasks done by increasingly larger crowds (e.g., to generate disaster relief plan), we need to use machine intelligence to coordinate individuals, and to make sense of, organize and display information to workers. In other words, AI algorithms can be used to make humans compute better. [...] **as workers**. For many tasks, machines actually outperform humans, both in terms of accuracy and speed. One can imagine future human computation systems to leverage both AI and humans as workers to perform different tasks they are better at. An effective human computation system should be able to interweave machine and human capabilities seamlessly. This idea is not new; many research concepts familiar to the AI community. [...]” (Law, Edith, and Luis von Ahn. *Human Computation. Synthesis Lectures on Artificial Intelligence and Machine Learning*, no. 13. Morgan & Claypool Publishers, 2011, 3).

which such behavior could be engineered in any type of artifact”<sup>13</sup>

The first recognized work in AI was Warren McCulloch and Walter Pitts' model of artificial neurons in 1943, a precursor to the connectionist approach. However, the term "Artificial Intelligence" was officially coined during a 1956 conference at Dartmouth College, organized by John McCarthy, where researchers convened to share and discuss their work.<sup>14</sup>

A pivotal breakthrough in the field was the development of machine learning techniques, which can be defined as “a subfield of AI that studies the ability to improve performance based on experience”<sup>15</sup> or simply as “a field of study that gives computers the ability to learn without being explicitly programmed”<sup>16</sup>.

The advantages of using those learning algorithms are remarkable as they allow the possibility of doing certain kinds of tasks that are too difficult to handle with programs written and designed by human beings<sup>17</sup>.

Several algorithms - designed to perform different tasks - can be submitted to different machine learning techniques. The use of those systems demands not only faster and more efficient computers but also reliable data from which they can be trained and learn the patterns that will be used for teaching the computer to solve the assigned tasks without needing constant human supervision<sup>18</sup>.

---

<sup>13</sup> Whitby, Blay. *Artificial Intelligence: A Beginner's Guide*. Oxford: Oneworld, 1988, 1.

<sup>14</sup> Franco, Cristiano Roberto. *Artificial Intelligence*. Londrina: Editora e Distribuidora Educacional S.A., 2014, 5-7.

<sup>15</sup> Norvig, Peter, and Stuart Russell. *Artificial Intelligence: A Modern Approach*. 4th ed. Hoboken: Pearson, 2020, 1.

<sup>16</sup> Samuel, Arthur L. “Some Studies in Machine Learning Using the Game of Checkers.” *IBM Journal of Research and Development* 3, no. 3 (1959), 210.

<sup>17</sup> Cf. Mackay, David J. C. *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press, 2003.

<sup>18</sup> “Supervised learning --- where the algorithm generates a function that maps inputs to desired outputs. One standard formulation of the supervised learning task is the classification problem: the learner is required to learn (to approximate the behavior of) a function which maps a vector into one of several classes by looking at several input-output examples of the function. Unsupervised learning --- which models a set of inputs: labeled examples are not available. Semi-supervised learning --- which combines both labeled and unlabeled examples to generate an appropriate function or classifier. Reinforcement learning --- where the algorithm learns a policy of how to act given an observation of the world. Every action has some impact in the environment, and the environment provides feedback that guides the learning algorithm. Transduction --- similar to supervised learning, but does not explicitly construct a function: instead, tries to predict new outputs based on training inputs, training outputs, and new inputs. Learning to learn --- where the algorithm learns its own inductive bias based on previous experience.” (Ayodele, Taiwo Oladipupo. “Types of Machine Learning Algorithms.” In *New Advances in Machine Learning*, edited by Yagang Zhang. Rijeka: InTech, 2010, 19).

Nevertheless, one of the huge barriers to the development of viable self-learning systems that can execute tasks with accuracy and efficiency close to human-level work is the need to label huge amounts of data<sup>19</sup>, which is costly and demands a lot of resources to be done.

However, the development of systems called *large language models* (LLMs), which use deep learning techniques to process and comprehend natural language<sup>20</sup> and are trained on “vast amounts of unlabeled text data, which allows them to learn patterns in unstructured sequences and build a knowledge base of language”<sup>21</sup> marks another major breakthrough in the area as these models can perform tasks they were not explicitly trained to handle if provided with clear human language instructions or multiple examples of a specific task.<sup>22</sup>

This capability arises from their ability to generalize knowledge learned during training and apply it to new, related scenarios, often referred to as “prompting”, highlights the flexibility of LLMs in adapting to diverse tasks through minimal guidance, making them valuable tools for contexts requiring adaptability and contextual reasoning and “a highly sought-after research area, because of their ability to generate human-like language”<sup>23</sup>

The class of algorithms behind those models are called *transformers*, a type of machine learning model designed to analyze sequences of data, such as text, by focusing on how different parts of the input relate to one another. They achieve this through a process called self-attention, which enables the model to identify and compute relationships between individual elements (e.g., words or phrases) in a given sequence.<sup>24</sup>

This mechanism allows the model to understand context, of the information it processes, even when important connections or relationships exist between elements and data that are far apart within the human input.

The capacity of “self-attention” makes those systems are particularly powerful for understanding and processing complex and context-dependent texts, and with a huge interesting advantage over the other kind of systems:

---

<sup>19</sup> Newlands, Gemma. “Lifting the Curtain: Strategic Visibility of Human Labour in AI-as-a-Service.” *Big Data & Society* 8, no. 1 (2021), 5.

<sup>20</sup> Cf. Shen, Yun, et al. “ChatGPT and Other Large Language Models Are Double-Edged Swords.” *Radiology* (2023). Advance online publication.

<sup>21</sup> Cf. Brown, Tom B., et al. “Language Models Are Few-Shot Learners.” *arXiv* (2020). Accessed August 31, 2024. <https://arxiv.org/abs/2005.14165>.

<sup>22</sup> Manning, Christopher D. “Human Language Understanding & Reasoning.” *Daedalus* 151, no. 2 (2022): 127–138.

<sup>23</sup> Fan, Lizhou, et al. “A Bibliometric Review of Large Language Models Research from 2017 to 2023.” *arXiv preprint*, arXiv:2304.02020, 2023. Accessed August 31, 2024. <https://arxiv.org/abs/2304.02020>.

<sup>24</sup> Cf. Vaswani, Ashish, et al. “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30 (2017): 5998–6008. Accessed August 31, 2024. <https://arxiv.org/abs/1706.03762>.

LLMs can be trained to execute new and specific tasks without the need of labelling all the data, as explained by Li:

To conduct a machine learning-based task (...) one had to label a large amount of data to train a model. In contrast, one currently needs only to label a small amount of data to fine-tune a pre-trained language model because it has already acquired a significant amount of knowledge necessary for language processing.<sup>25</sup>

These pre-trained LLMs "can be fine-tuned with additional training to specialize them for a variety of language tasks,"<sup>26</sup> with a training cycle that is "much faster than it takes to develop a typical machine learning application, which requires custom databases and hyperparameter searches."<sup>27</sup>

Those models are already available on the market and allow the commercial use of artificial intelligence for building AI-based systems that can be used in all kinds of business at an affordable price and with an immediate impact in labor dynamics.

### *B. Automation and codification of labor*

Automatism, or automation, can be understood as the distribution of control energy by a machine<sup>28</sup>, distinguishing it from mere mechanical action, which is simply a reflex action initiated by human input.<sup>29</sup>

The concept of an entirely autonomous machine remained a utopia until the modern era, when the term *robot* was coined<sup>30</sup> from the Czech word *robotá*, meaning "servitude" or "forced labor."<sup>31</sup> This term reflects the belief that machines would eventually take over all forms of labor, thereby freeing humanity to pursue other endeavors with increased leisure time.

However, this ideal was soon replaced by the realization that machines

---

<sup>25</sup> Li, Hang. "Language Models: Past, Present, and Future." *Communications of the ACM* 65, no. 7 (2022): 56–63, 56.

<sup>26</sup> Sejnowski, Terrence. "Large Language Models and the Reverse Turing Test." *Neural Computation* 35, no. 3 (2023): 309–342, 331.

<sup>27</sup> Sejnowski, Terrence. "Large Language Models and the Reverse Turing Test." *Neural Computation* 35, no. 3 (2023): 309–342, 331.

<sup>28</sup> Latil, Pierre de. *Pensamento Artificial*. Translated to Portuguese by Jerônimo Monteiro. São Paulo: IBRASA, 1959, 43.

<sup>29</sup> Latil, Pierre de. *Pensamento Artificial*. Translated to Portuguese by Jerônimo Monteiro. São Paulo: IBRASA, 1959, 51.

<sup>30</sup> The term robot gained worldwide repercussions when it was used in the play R.U.R (Rossum's Universal Robots), by Czech writer Karel Capek, set in a factory that produced artificial human beings. Capek, however, attributes the creation of the term robot to his brother Josef (Čapek, Karel. "The *přívodu slova robot*." Accessed August 20, 2024. <http://capek.misto.cz/robot.html>.)

<sup>31</sup> Cf. Kurfess, Thomas. *Robotics and Automation Handbook*. Boca Raton: CRC Press, 2006.

did not liberate humans from work but instead reduced workers to mere extensions of the machinery, as Koyré observed:

The machine age, instead of being humanity's golden age, turned out to be its iron age. The shuttles and plectrums moved well enough on their own, but the weaver remained chained to his heir more than ever. Instead of freeing man and making him "the master and dominator of nature", the machine turned man into a slave of his own creation. Moreover, by a surprising paradox, the machine, by increasing man's productive power, undoubtedly created wealth but, at the same time, spread misery. [...] But it was necessary to surrender to the evidence: the machine (or at least the machine working in the given economic and social conditions) has considerably increased the output of labor, but it has also created unemployment. Moreover, by taking the division of labor and its decomposition into elementary operations ever further, the machine made work simpler (which, as Proudhon saw very well, made it possible to replace the craftsman or skilled worker with a manual laborer), but dehumanized it and made it much more monotonous and boring.<sup>32</sup>

The advent of computerized technologies marks a turning point in human history, characterized by "the objectification by the machine of abstract, reflexive brain functions, no longer brain functions linked to the activity of the hand."<sup>33</sup>

This shift represents a fundamental change from previous technological revolutions, which focused on the automation of manual labor through tools and machinery designed to manipulate matter..<sup>34</sup>

---

<sup>32</sup> Author's translation of the original: "A idade da máquina, ao invés de ser a idade de ouro da humanidade, revelou-se a sua idade de ferro. As lançadeiras e os plectos moviam-se bem sozinhos, mas o tecelão permanecia mais do que nunca encadeado ao seu oheio. Ao invés de libertar o homem e fazer dele "o senhor e dominador da natureza", a máquina transformou o homem num escravo de sua própria criação. Além disso, por um paradoxo surpreendente, a máquina, ao aumentar a potência produtiva do homem, sem dúvida criou a riqueza mas, ao mesmo tempo, difundiu a miséria. [...] Mas era necessário render-se à evidência: a máquina (ou pelo menos a máquina funcionando em condições econômicas e sociais dadas) aumentou consideravelmente o rendimento do trabalho; mas, por isso mesmo, criou o desemprego. Além disso, levando sempre mais longe a divisão do trabalho e sua decomposição em operações elementares, a máquina tornou o trabalho mais simples (o que, como Proudhon viu muito bem, permitiu substituir o artesão ou o operário qualificado por um trabalhador braçal), mas desumanizando-o e tornando-o muito mais monótono e aborrecido." (Koyré, Alexandre. "Os filósofos e a máquina." In *Estudos de história do pensamento filosófico*, edited by Alexandre Koyré, 244–245. Rio de Janeiro: Forense Universitária, 1991. Originally published in 1948, 244-245).

<sup>33</sup> Lojkine, Jean. *A Revolução Informacional*. Translated by José Paulo Netto. 1st ed. São Paulo: Cortez, 1995, 80.

<sup>34</sup> Lojkine, Jean. *A Revolução Informacional*. Translated by José Paulo Netto. 1st ed.

Lojkine describes this phenomenon as a new phase in the Industrial Revolution, termed the Information Revolution characterized by the objectification of cognitive functions through standardized information processing:

The information revolution is opposed to the industrial revolution that marked the global development of capitalism from the 18th century onwards. The Industrial Revolution, as its name suggests, is a technological revolution in a specific sector of the economy: industrial production, characterized by the transformation of matter and the production of goods. The heart of this technological revolution is the objectification of the work done by the "hand tool" in machine tools. The information revolution, on the other hand, is characterized by the objectification of certain brain functions in terms of the standardized processing of information. It applies not only to the professional field, but also to all sectors of the economy and, more broadly, to all professional sectors of society [...] [...] a first process is that of the transition from industrial production work to the work of processing complex information, to the social service relationship: the work is more qualified, more intellectualized, it employs new relational and communicational capacities that are difficult to standardize and reproduce: the reaction to events, the capacity for initiative, creation and innovation gradually replace the work of carrying out prescribed tasks.<sup>35</sup>

In the field of labor relations, automation, and technological advancements have the potential to drive significant social transformations. This issue ties directly to the principles of supply and demand, and also the

---

São Paulo: Cortez, 1995, 80.

<sup>35</sup> Author's translation of the Portuguese version: "A revolução informacional opõe-se à revolução industrial que marcou o desenvolvimento mundial do capitalismo a partir do século XVIII. A revolução industrial, como seu nome indica, é uma revolução tecnológica em um setor preciso da economia: a produção industrial, caracterizada pela transformação da matéria e pela produção de mercadorias. O coração dessa revolução tecnológica é a objetivação do trabalho realizado pela "mão ferramenta" nas máquinas-ferramentas. Já a revolução informacional é caracterizada pela objetivação de certas funções do cérebro no que concerne ao tratamento padronizado da informação. Ela se aplica não somente ao domínio profissional, mas também a todos os setores da economia e, mais amplamente, a todos os setores profissionais da sociedade [...]. [...] um primeiro processo é aquele da passagem do trabalho de produção industrial ao trabalho de tratamento das informações complexas, à relação social de serviço: o trabalho é mais qualificado, mais intelectualizado, ele emprega novas capacidades relacionais e comunicacionais dificilmente padronizadas, reprodutíveis: a reação dos acontecimentos a capacidade de iniciativa, de criação, de inovação substituem pouco a pouco o trabalho de execução das tarefas prescritas." (Lojkine, Jean. "O novo salariado informacional: nas fronteiras do salariado." *Revista Crítica Marxista* 25 (2007): 31–37.)

cost of producing intellectual work. According to Lojkine, "the technological innovations of the new complex of productive restructuring are mixed with the organizational innovations of Toyotism"<sup>36</sup>, where increased automation "tends to stimulate wage problems because it increases productivity and also the profit margins of the firms directly concerned."<sup>37</sup>

Lojkine further highlights that new information are showing an increasing potential to deepen labor market distortions. This results in heightened demand for hyper-skilled workers capable of operating machines more efficiently or performing tasks less susceptible to automation.<sup>38</sup>

However, while this leads to substantial income increases for highly skilled workers, it simultaneously precipitates growing precariousness for less qualified workers and as well as for those once deemed skilled before the rapid advancements in AI technology.<sup>39</sup>

This scenario leads to diminished employment prospects and either lower wages or increased workloads for the remaining employees. Ferrari explains:

If the new mode of production presupposes a minimized role for the workforce, with its replacement by knowledge, at the same time as the presence of the capital factor expands and becomes more sophisticated, there is a drastic reduction in the bargaining power of the working class in the conflict between capital and labor over the redistribution of surplus wealth. Thus, the distribution of the surplus will not take place according to the correlation of forces representing the factors of production (capital and labor), but will be the result of institutional adjustments consolidated in some new form of state. Or there will be no redistribution at all. Most of today's specializations will quickly become obsolete and the very concept of the collective workplace will be superseded by the rise of communications and information technology. This implies a challenge for trade unions. If they resist the changes in defense of little or no productive employment, they run the risk of becoming history. Society will not be able to afford outdated, economically uncompetitive forms of production. It's clear that unproductive work, as well as being unhealthy from many angles, becomes costly. The automatic elevator attendant, for example, will cost less for himself, for the company and

---

<sup>36</sup> Schaff, Adam. *A sociedade informática: As consequências sociais da segunda revolução industrial*. Translated by Carlos Eduardo Jordão Machado and Luiz Arturo Obojes. São Paulo: Editora da Universidade Paulista, 1995, 120.

<sup>37</sup> Einzig, Paul. *As consequências econômicas da automação*. Translated by Jorge Eneas Fortes. Rio de Janeiro: Editora Fundo de Cultura, 1959, 179.

<sup>38</sup> Lojkine, Jean. "O novo salariado informacional: nas fronteiras do salariado." *Revista Crítica Marxista* 25 (2007): 35–36.

<sup>39</sup> Lojkine, Jean. "O novo salariado informacional: nas fronteiras do salariado." *Revista Crítica Marxista* 25 (2007): 35–36.

for society if he stays at home living on a socially stipulated minimum income, exercising his right to idleness or preparing himself through studies for productive work.<sup>40</sup>

The application of those technologies in the labor market enabled the conversion of living labor into dead labor<sup>41</sup>, by storing data related to labor activities within machines.

Therefore, when employees carry out their tasks under an employment contract, they not only fulfill their duties but also generate knowledge on how to perform those tasks.

This process, known as the "codification of worker know-how,"<sup>42</sup> involves integrating the worker's expertise into machinery—whether through data or software<sup>43</sup> - such that the machine owner no longer needs to employ or compensate the worker for their knowledge that was once necessary to execute the job.

As a result, human labor is either eliminated or restructured, with employers opting for less skilled workers who can operate these informational machines at lower wages or for a small number of highly skilled workers who can now perform tasks previously handled by multiple professionals, all at reduced costs.

---

<sup>40</sup> Author's translation. In the original: "Se o novo modo de produção pressupõe um papel minimizado da força de trabalho, com a substituição desta pelo conhecimento, ao mesmo tempo em que a presença do fator capital se amplia e se sofisticada, tem-se uma diminuição drástica do poder de barganha da classe trabalhadora no conflito entre capital e trabalho pela redistribuição da riqueza excedente. Assim, a distribuição do excedente não se dará de acordo com a correlação de forças representativas dos fatores de produção (capital e trabalho), mas sim será fruto de ajustes institucionais consolidados em alguma nova forma de Estado. Ou não haverá redistribuição alguma. A maioria das especializações atuais tornar-se-á rapidamente obsoleta e o próprio conceito de local coletivo de trabalho estará superado com o incremento das comunicações e da informática. Isso implica um desafio para os sindicatos. Se na defesa do emprego pouco ou nada produtivo resistirem às mudanças, correm o risco de virar história. A sociedade não terá condições de arcar com formas de produção superadas, economicamente não-competitivas. Está claro que o trabalho improdutivo, além de insalubre sob vários ângulos, torna-se oneroso. O ascensorista de elevador automático, por exemplo, custará menos para si, para a empresa e para a sociedade se ficar em casa vivendo com uma renda mínima socialmente estipulada, exercendo seu direito ao ócio ou preparando-se, através de estudos, para o exercício de funções produtivas" (Ferrari, Levi B. "Revolução Tecnológica e Estado." In *A revolução tecnológica e os novos paradigmas da sociedade*. Belo Horizonte/São Paulo: Oficina de Livros/IPSO, 1993, 109.)

<sup>41</sup> Antunes, Ricardo. "O trabalho e seus sentidos." *Revista Debate & Sociedade* 1, no. 1 (2011), 7.

<sup>42</sup> Antunes, Ricardo. "O trabalho e seus sentidos." *Revista Debate & Sociedade* 1, no. 1 (2011), 103.

<sup>43</sup> Wolf, Simone. "O 'trabalho informacional' e a reificação da informação sob os novos paradigmas organizacionais." In *Infoproletários: Degradação real do trabalho virtual*, edited by Ricardo Antunes and Ruy Braga, 89–112. São Paulo: Boitempo Editorial, 2009, 103.

This shift leads to the replacement of genuinely creative intellectual work with a simplified, mechanized version limited to following machine commands. This phenomenon of "computerized neo-Taylorism,"<sup>44</sup> is evident in the current development of commercial Large Language Models (LLMs) and chatbots that companies can access at affordable prices.

These systems are trained to increase their precision using worker-generated data, effectively transforming labor and expertise into algorithms, models, and automated processes.

By leveraging workers' know-how to fine-tune these systems, companies can enhance their efficiency in performing specific tasks, thereby accelerating the course of codification of labor at a faster pace and lower cost.

Thus, in the rapidly evolving landscape of AI and data-driven technologies, the continual development of new systems capable of performing complex tasks in specialized fields has become central to the modern employment dynamic and the need to find alternatives to provide workers protection against automation regarding AI systems.

This transformation underscores the urgent need for regulatory frameworks to protect workers' rights and adapt labor standards to this new reality.

## II. REGULATING THE USE OF WORKERS-PRODUCED DATA IN ARTIFICIAL INTELLIGENCE-BASED SYSTEMS

The efficiency and accuracy of AI systems and machine learning models are heavily dependent on vast datasets, and some of them can be generated through the daily activities of workers. From text data found in emails and reports to image and voice data from various multimedia sources, the modern workplace has evolved into a data-rich environment. Consequently, work-generated data serves as a crucial foundation for training AI algorithms and models.

As AI technology continues to advance, the boundary between human labor and machine automation becomes increasingly blurred, leading to the growing codification of labor. In this context, striking a balance between technological progress and the protection of worker rights is of paramount importance.

The demand for legal safeguards to protect workers from automation is a pressing issue. However, existing regulations have struggled to establish a comprehensive framework capable of effectively addressing this challenge<sup>45</sup>,

---

<sup>44</sup> Lima, Vinícius Moreira de. *Relação de trabalho versus relação de emprego: A luta pela nova Justiça do Trabalho*. Porto Alegre: Sérgio Antônio Fabris Ed., 2012, 117

<sup>45</sup> As an example, in the Brazilian Constitution, the protection of urban and rural workers from automation is a fundamental right expressed in item XXVII, of Article 7, but there is

particularly given the difficulty of slowing down, prohibiting, or ignoring the adoption of new technologies.

Nonetheless, protecting workers from the adverse effects of automation can be achieved by ensuring their participation in the benefits of technological advancements and their application to production processes<sup>46</sup>.

A potential regulatory framework can aim for two kinds of approaches to establish a fairground regulation able to balance the conflicting interests: (1) granting workers a stake in the economic gains resulting from automation, and (2) developing legal mechanisms that protect their rights to their data and know-how while allowing for continued technological progress.

One of the simplest approaches to granting workers protection against this kind of automation is granting financial benefits or compensation for the codification of work and the use of workers' creativity and inputs to develop innovation and improvements in their workplace for their employers' benefit.

There is room for that reasoning in traditional regulations regarding intellectual property laws in labor relationships. For example, under Brazilian Industrial Property Law (Law No. 9.279/96), Article 91 establishes a framework for the joint ownership of inventions and utility models when they are the outcome of both the employee's contribution and the employer's resources such as data, means, materials, facilities, or equipment.

In these cases, ownership is typically shared equally between the employee and employer, unless specific provisions in the employment contract dictate otherwise.

Furthermore, this regulation underscores the employer's exclusive right to license the invention or utility model, while simultaneously ensuring that the employee receives equitable remuneration for their contributions. This approach reflects a commitment to fairness in the distribution of economic gains derived from exploiting such patents. In scenarios involving multiple employees who have collectively contributed to the invention or utility model, the law mandates an equal division of the ownership share among them, although this can be subject to contractual agreements among the parties.

Additionally, if multiple employees collaborate on an invention, the law mandates an equal division of ownership among them, unless other contractual agreements are in place. The regulation also sets a specific

---

no further legislation or regulation to define the extension and how that kind of protection would be applied.

<sup>46</sup> Cezar, Frederico Gonçalves. "Controle de impactos da inteligência artificial em relações laborais na perspectiva do debate constituinte sobre proteção do trabalhador em face da automação." In *O uso de dados pessoais e inteligência artificial na relação de trabalho: proteção, discriminação, violência e assédio digital*, edited by Thiago Milanez Andraus, Adriane Reis de Araújo, Rodrigo de Lacerda Carelli, and Thaís Fidelis Alves Bruch, Vol. 1. Brasília: Ministério Público do Trabalho, 2022, 197-198.

timeframe for employers to begin exploiting the patented invention; if the employer fails to do so within one year of patent issuance, exclusive rights may transfer to the employee, provided no legitimate reason for delay exists.

Moreover, in cases involving the transfer or assignment of rights, both employee and employer co-owners are granted equal preferential rights to acquire the exclusive ownership share.

At first glance, those regulations may appear sufficient to protect workers from the effects of automation. It has already been argued on lawsuits that the outputs generated by AI systems do not qualify for copyright protection on the basis that human authorship is essential for a valid copyright claim<sup>47</sup>.

Therefore, people who had their work-related data used to create datasets – notably authors and copyright holders - are demanding compensation for the unauthorized use of their intellectual property in training those AI systems<sup>48</sup>.

Given that AI systems can effectively perform tasks by leveraging data generated from human labor and creativity, there is a compelling argument for exploring the possibility of protecting the intellectual property associated with the creative process embedded in the data used for training these systems.

Ensuring the need for compensation and recognition for employees' contributions could help address the potential exploitation of their work.

However, the transformation of worker-generated data into training material for AI systems raises significant questions about intellectual property rights and ownership. Specifically, there is still uncertainty regarding whether employees are entitled to compensation, such as royalties, for the contributions that enhance the performance of these AI systems.

The inability to properly apply the legal protections under Intellectual Property Law to work-related tasks produced by AI systems creates an imbalance in safeguarding workers against the misuse of their creativity and labor.

This deviation from the intended purpose of labor protections becomes problematic when - even if AI-generated outputs may not qualify for intellectual property rights - employers still derive substantial utility from these outputs to enhance production efficiency, thus gaining tangible benefits from collecting and using work-related data without compensating

---

<sup>47</sup> Thaler v. Perlmutter. *U.S. District Court for the District of Columbia*, No. 1:22-cv-01564 (2024). Accessed August 31, 2024. <https://www.courtlistener.com/docket/63356475/thaler-v-perlmutter/>.

<sup>48</sup> A main leading case in US regarding use of copyrighted data for training LLM models is Tremblay et al. v. OpenAI. *U.S. District Court for the Northern District of California*, No. 3:23-cv-03223 (2023). Accessed August 31, 2024. <https://storage.courtlistener.com/recap/gov.uscourts.cand.414822/gov.uscourts.cand.414822.1.0.pdf>.

employees<sup>49</sup>.

This issue extends beyond just workers' rights, as many AI training efforts disregard copyright and intellectual property laws, which has led to widespread legal disputes across various jurisdictions<sup>50</sup>. These disputes also highlight potential violations of data protection regulations, as Hacker emphasizes:

In Article 6(4) of the GDPR, EU data protection law sets out, as an expression of the purpose limitation principle (Article 5(1)(b) GDPR), specific requirements for changing the purpose of personal data. They apply, for example, if data originally collected with a different aim is now supposed to be used as training data. (...) it is conceivable that prospective training data is protected by copyright or related rights (e.g. the sui generis database right). (...) The training typically includes activities relevant for copyright protection. For example, the individual data must be saved on a server and stored in the working memory, which implies a reproduction of the work in terms of copyright. If the original data is pre-processed, an adaptation relevant under copyright law will often take place. Finally, access to databases may involve an extraction requiring permission. As a consequence, the training can only be carried out in conformity with intellectual property rights if either a license is obtained or a specific exception is provided for the respective intellectual property right.<sup>51</sup>

Although the issue of whether workers-produced data used to train AI systems can be granted intellectual property protection remains contentious, the ongoing codification of labor into AI products demands the collection and use of personal and often sensitive data.

This situation underscores the importance of granting protection of workers' rights and compliance with both intellectual property and data protection regulations. That can be done by ensuring that workers need to be informed about the collection and use of their work-related data, providing them with the option to opt out or explicitly consent, and safeguarding their privacy is essential in light of data privacy regulations.

Rather than granting workers co-ownership or intellectual property rights

---

<sup>49</sup> Adding up to this issue is that companies also hire workers – mostly in third-world countries – with the sole goal of training and fine-tuning AI systems. (Cf. Newlands, Gemma. “Lifting the Curtain: Strategic Visibility of Human Labour in AI-as-a-Service.” *Big Data & Society* 8, no. 1 (2021), 5.)

<sup>50</sup> Min, Sewon, et al. “SILO Language Models: Isolating Legal Risk in a Nonparametric Datastore.” *arXiv preprint*, arXiv:2308.04430, 2023. Accessed August 31, 2024. <https://arxiv.org/abs/2308.04430>

<sup>51</sup> Hacker, Philipp. “A Legal Framework for AI Training Data—From First Principles to the Artificial Intelligence Act.” *Law, Innovation and Technology* 13 (2021): 257–301, 216 and 277.

over the data used to train AI systems, employers should focus on preventing the unauthorized use of employees' data. They should also provide compensation or reparations if such data is used internally or shared with third parties for training purposes.

This approach is not entirely novel but is instead a direct consequence of existing rules regarding the protection of workers' personal data. While employees may face certain restrictions on their freedom of movement and attire during work hours, they do not forfeit their fundamental rights to privacy<sup>52</sup>, private life, image, name, and voice while performing their duties.

These personal rights are already granted by laws such as the Brazilian Civil Code, specifically Articles 11 to 21, which are also applicable to employment contracts and under the Brazilian General Data Protection Law (LGPD), data processing must only occur with the employee's explicit and written consent (Articles 5, XII; 7, I; 8).

However, given the inherent power dynamics between employers and employees, obtaining genuine, uncoerced consent can be challenging. As Hendrickx has pointed out, "a major question exists as to whether consent can be a legitimate ground for personal data processing"<sup>53</sup> in such asymmetrical relationships.

To address these concerns, compensation for using workers-produced data should be explicitly negotiated and documented within employment contracts or through separate consent agreements. This ensures that workers' data privacy rights are respected and enables them to seek fair compensation for the use of their data in the context of AI training and automation.

## CONCLUSIONS

- The current technological landscape of AI and its rapidly evolving business environment, enables accelerated development and decentralization of AI systems. This has resulted in a proliferation of diverse services and applications in the market, allowing private companies and entrepreneurs to create personalized, fine-tuned AI systems trained on workers-produced data.

- As the quality of outputs produced by these AI systems improves at an accelerated rate through machine learning—using worker-generated data for training—market pressure intensifies. This dynamic can lead to job displacement and a reduced need for hiring skilled workers or paying high wages.

---

<sup>52</sup> As stated in Article 5.13 of ILO Protection of workers' personal data (International Labour Organization (ILO). *Protection of Workers' Personal Data: An ILO Code of Practice*. Geneva: International Labour Organization, 1997.)

<sup>53</sup> Hendrickx, Frank. *Protection of Workers' Personal Data: General Principles*. ILO Working Paper, no. 62. Geneva: International Labour Organization, 2022, 25.

- Addressing the need to protect workers in this new paradigm of automation requires innovative solutions. Efforts to prohibit or restrict technological advancements have historically proven ineffective, even with earlier forms of automation. An alternative approach involves ensuring fair compensation for workers when their work-related data is used to train AI systems.

- One possible path is to extend the right of fair compensation to workers under intellectual property regulations. However, this approach carries uncertainties due to ongoing debates about granting copyright protection to AI-generated works and for the workers who produce the data for training AI models.

- Therefore, a pragmatic regulatory framework to address this issue is to enhance the rights and principles of data privacy and labor law, especially the understanding of what can be considered valid consent under an employment contract. This would entail that any use of data generated by employees during their contractual work—whether for the benefit of their employer or a third party—must be contingent on obtaining the workers' explicit, informed consent and ensuring appropriate financial compensation for using their data.

The findings of this study highlight the urgent need for establishing new legal frameworks that balance the protection of workers' rights with the ongoing advancements in AI and data-driven technologies. A well-defined approach that integrates data protection laws with labor rights is essential to ensuring fair treatment while enabling technological progress.

#### REFERENCES

- Antunes, Ricardo. "O trabalho e seus sentidos." *Revista Debate & Sociedade* 1, no. 1 (2011).
- Apter, Michael J. *Cybernetics and Psychology*. Petrópolis: Vozes, 1973.
- Ayodele, Taiwo Oladipupo. "Types of Machine Learning Algorithms." In *New Advances in Machine Learning*, edited by Yagang Zhang. Rijeka: InTech, 2010.
- Brown, Tom B., et al. "Language Models Are Few-Shot Learners." *arXiv* (2020). Accessed August 31, 2024. <https://arxiv.org/abs/2005.14165>.
- Čapek, Karel. "The původu slova robot." Accessed August 20, 2024. <http://capek.misto.cz/robot.html>.
- Cezar, Frederico Gonçalves. "Controle de impactos da inteligência artificial em relações laborais na perspectiva do debate constituinte sobre proteção do trabalhador em face da automação." In *O uso de dados pessoais e inteligência artificial na relação de trabalho: proteção, discriminação, violência e assédio digital*, edited by Thiago Milanez Andraus, Adriane Reis de Araújo, Rodrigo de Lacerda Carelli, and Thaís Fidelis Alves

- Bruch, Vol. 1. Brasília: Ministério Público do Trabalho, 2022.
- Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. 3rd ed. Cambridge: The MIT Press, 2009.
- Einzig, Paul. *As consequências econômicas da automação*. Translated by Jorge Eneas Fortes. Rio de Janeiro: Editora Fundo de Cultura, 1959.
- Fan, Lizhou, et al. “A Bibliometric Review of Large Language Models Research from 2017 to 2023.” *arXiv preprint*, arXiv:2304.02020, 2023. Accessed August 31, 2024. <https://arxiv.org/abs/2304.02020>.
- Ferrari, Levi B. “Revolução Tecnológica e Estado.” In *A revolução tecnológica e os novos paradigmas da sociedade*. Belo Horizonte/São Paulo: Oficina de Livros/IPSO, 1993.
- Fertig, Cristina, and Marco Medina. *Algorithms and Programming: Theory and Practice*. São Paulo: Novatec Editora, 2006.
- Franco, Cristiano Roberto. *Artificial Intelligence*. Londrina: Editora e Distribuidora Educacional S.A., 2014.
- Hacker, Philipp. “A Legal Framework for AI Training Data—From First Principles to the Artificial Intelligence Act.” *Law, Innovation and Technology* 13 (2021): 257–301.
- Hendrickx, Frank. Protection of Workers' Personal Data: General Principles. *ILO Working Paper*, no. 62. Geneva: International Labour Organization, 2022.
- International Labour Organization (ILO). *Protection of Workers' Personal Data: An ILO Code of Practice*. Geneva: International Labour Organization, 1997.
- Jacker, Corine. *Man, Memory and the Machine: An Introduction to Cybernetics*. Rio de Janeiro: Forense, 1970.
- Knuth, Donald. *The Art of Computer Programming*. 2nd ed. Reading: Addison-Wesley, 1973.
- Koyré, Alexandre. “Os filósofos e a máquina.” In *Estudos de história do pensamento filosófico*, edited by Alexandre Koyré, 244–245. Rio de Janeiro: Forense Universitária, 1991. Originally published in 1948.
- Kurfess, Thomas. *Robotics and Automation Handbook*. Boca Raton: CRC Press, 2006.
- Latil, Pierre de. *Pensamento Artificial*. Translated to Portuguese by Jerônimo Monteiro. São Paulo: IBRASA, 1959.
- Law, Edith, and Luis von Ahn. Human Computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, no. 13. Morgan & Claypool Publishers, 2011.
- Li, Hang. “Language Models: Past, Present, and Future.” *Communications of the ACM* 65, no. 7 (2022): 56–63.
- Liddell, Henry George, and Robert Scott. *Greek-English Lexicon*. Accessed

- September 1, 2024. <http://www.perseus.tufts.edu/>.
- Lima, Vinícius Moreira de. *Relação de trabalho versus relação de emprego: A luta pela nova Justiça do Trabalho*. Porto Alegre: Sérgio Antônio Fabris Ed., 2012.
- Lojkin, Jean. *A Revolução Informacional*. Translated by José Paulo Netto. 1st ed. São Paulo: Cortez, 1995.
- Lojkin, Jean. “O novo salariado informacional: nas fronteiras do salariado.” *Revista Crítica Marxista* 25 (2007): 31–37.
- Mackay, David J. C. *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press, 2003.
- Manning, Christopher D. “Human Language Understanding & Reasoning.” *Daedalus* 151, no. 2 (2022): 127–138.
- Min, Sewon, et al. “SILO Language Models: Isolating Legal Risk in a Nonparametric Datastore.” *arXiv preprint*, arXiv:2308.04430, 2023. Accessed August 31, 2024. <https://arxiv.org/abs/2308.04430>.
- Newlands, Gemma. “Lifting the Curtain: Strategic Visibility of Human Labour in AI-as-a-Service.” *Big Data & Society* 8, no. 1 (2021).
- Norvig, Peter, and Stuart Russell. *Artificial Intelligence: A Modern Approach*. 4th ed. Hoboken: Pearson, 2020.
- Samuel, Arthur L. “Some Studies in Machine Learning Using the Game of Checkers.” *IBM Journal of Research and Development* 3, no. 3 (1959).
- Schaff, Adam. *A sociedade informática: As consequências sociais da segunda revolução industrial*. Translated by Carlos Eduardo Jordão Machado and Luiz Arturo Obojes. São Paulo: Editora da Universidade Paulista, 1995.
- Sejnowski, Terrence. “Large Language Models and the Reverse Turing Test.” *Neural Computation* 35, no. 3 (2023): 309–342.
- Shen, Yun, et al. “ChatGPT and Other Large Language Models Are Double-Edged Swords.” *Radiology* (2023). Advance online publication.
- Thaler v. Perlmutter. *U.S. District Court for the District of Columbia*, No. 1:22-cv-01564 (2024). Accessed August 31, 2024. <https://www.courtlistener.com/docket/63356475/thaler-v-perlmutter/>.
- Tremblay et al. v. OpenAI. *U.S. District Court for the Northern District of California*, No. 3:23-cv-03223 (2023). Accessed August 31, 2024. <https://storage.courtlistener.com/recap/gov.uscourts.cand.414822/gov.uscourts.cand.414822.1.0.pdf>.
- Vaswani, Ashish, et al. “Attention Is All You Need.” *Advances in Neural Information Processing Systems* 30 (2017): 5998–6008. Accessed August 31, 2024. <https://arxiv.org/abs/1706.03762>.
- Whitby, Blay. *Artificial Intelligence: A Beginner's Guide*. Oxford: Oneworld, 1988.
- Willis, Robert. *Principles of Mechanism: Designed for the Use of Students in*


*the Universities and for Engineering Students Generally*. London: John W. Parker, 1861.

Wolf, Simone. "O 'trabalho informacional' e a reificação da informação sob os novos paradigmas organizacionais." In *Infoproletários: Degradação real do trabalho virtual*, edited by Ricardo Antunes and Ruy Braga, 89–112. São Paulo: Boitempo Editorial, 2009.

\* \* \*

*Rômulo Soares Valentini*

Ph.D. in Law from the Federal University of Minas Gerais. Master's degree in Law from the Federal University of Minas Gerais. Specialist in Substantive and Procedural Labor Law. Bachelor of Law. Visiting professor in the graduate program at UFMG in the research line "State, Reason, and History," teaching the course *Topics in Comparative Labor Law: The Future of Work*. Labor Law professor at FPL/MG. Professor in the postgraduate law programs at Faculdade Arnaldo and IEC-PUC/MG. Judicial Analyst at the 45th Labor Court of Belo Horizonte, Regional Labor Court of the 3rd Region. Email: [rsvaleantini@gmail.com](mailto:rsvaleantini@gmail.com)  
ORCID iD: <https://orcid.org/0000-0003-4340-9917>

 10.59224/bjlti.v2i2.34-53  
ISSN: 2965-1549