Título: Marx, Value and Artificial Intelligence

Título em Espanhol: Marx, Valor y Inteligencia Artificial

Título em inglês: Marx, Value and Artificial Intelligence

Autores: Alexandre Becker[1] e José Paulo Guedes Pinto[2]

Resumo

A discussão sobre Inteligência Artificial é apresentada à sociedade de forma fetichista. Neste trabalho, buscamos mostrar que é possível usar categorias marxistas para compreendê-la como fenômeno socioeconômico, inexistente sem trabalho humano, e voltada à valorização do capital.

Palavras-chave: Inteligência Atificial; Subsunção do Trabalho; Marx.

Resumen

La discusión sobre la Inteligencia Artificial se presenta a la sociedad de manera fetichista. En este trabajo, buscamos mostrar que es posible utilizar categorías marxistas para comprenderla como un fenómeno socioeconómico, inexistente sin trabajo humano y orientado a la valorización del capital.

Palabras-clave: Inteligencia Artificial; Subordinación del Trabajo; Marx..

Abstract

The discussion about Artificial Intelligence is presented to society in a fetishized manner. In this work, we seek to show that it is possible to use Marxist categories to understand it as a socio-economic phenomenon, nonexistent without human labor, and aimed at capital valorization.

Keywords: Artificial Intelligence; Subsumption of Labor; Marx.

---

1 Bacharel em Ciência e Tecnologia e em Neurociências e mestre em Economia Política Mundial pela Universidade Federal do ABC. Atualmente é doutorando em Economia Política Mundial pela UFABC. Email: alexandre.becker@ufabc.edu.br.

2 Professor Associado da UFABC. Professor da pós graduação em Economia Política Mundial (UFABC). E-mail: jose.guedes@ufabc.edu.br. Orcid: https://orcid.org/0000-0003-2457-2882.

Marx, Labor, Fetish and AI

Artificial Intelligence (AI) is impacting several fields of human endeavor (Furman, 2019). Being a practice born inside capitalist societies, its dawn is accompanied by a significant level of fetishism (Marx, 1990). This is partly due to the lack of a precise definition of what an AI actually is (Monett and Lewis, 2018; Wang, 2019), which concedes to any software capable of a certain degree of “autonomy” to be called “intelligent”.

Such a condition is inherent to the field in which the debate around AI's definition is taking place. AI as a topic is not only multidisciplinary but also unconstrained by the walls of academia. Corporations such as Google, Amazon, Microsoft, and others, are responsible for conducting a significant part of the research in the field (Abdalla 2021; Foster, 2020). Anyone trying to define AI will have to account for events in an environment of constant and decentralized movement (Abdalla, 2021), often protected by industrial secrets and/or false/exaggerated claims. (Benaich & Hogardth, 2020)

However, it is widely accepted that one specific set of techniques, often labeled as Machine Learning (ML), may always be treated as AI (Muhamedyev, 2015). Machine Learning itself is usually divided into four main categories: supervised learning, unsupervised learning, semi-supervised learning and, reinforced learning.

Supervised Learning is made through the use of a labeled set of data (or dataset), in which a given point of data is correlated to a dependent variable. After that, the created model is capable of making predictions about data that has no label. In unsupervised learning, the algorithm searches for similarities between the entries in the dataset, without having a label that identifies them. In semi-supervised both techniques are applied at the same time and in reinforcement learning the machine “rewards” success in performing a given task and “punishes” failure by making a given correlation more or less important for a model (Lee & Shin, 2020; Muhamedyev, 2015).

Even though ML techniques are only a part of what is usually called AI (Lee & Shin, 2020), and a definition of the boundaries of what constitutes an AI is still blurry, most authors agree that any AI technique (including ML-based ones) will demand three things to become a real-world application: a) processing power b) mathematical/statistical methods applied through sophisticated software and c) humongous amounts of data (Allen, 2020).

It's through this “decomposition” of AI in its constitutive parts that the political economy can analyze it without falling into a vulgar form of fetishism. This means that instead of adopting a specific definition of AI, it is possible to accept it as a dynamic human economic/social practice, that has risen from a specific historical period and therefore carries with it its contradictions. Also, it's at the same time constrained by the material limits of its time and is affecting these limits, altering the environment that allowed its creation.

We will regard AI models as any other commodity, and its creation as any other industrial activity. The before-mentioned tripod of AI (Processing Power, Data, Software), will be considered as Marxian **means of production**. **processing power** will be regarded as a **tool of labor**, the **data** as a **subject of labor**, and the work of the **data scientist** who both develops and uses existing software to create the model as **labor power itself** (Marx, 1995, p. 129).

This materialistic perspective takes into consideration what role AI plays in the society that created it, making it clear that it exists only through human labor. It admits that, inside capitalist societies, the flow of capital determines the logic around which new techniques develop themselves. Allowing us to explore how AI's constituent parts interact to allow capital to value itself in the current society and what contradictions loom from this process.

To do so, a brief review of the classical marxian understanding of capital fetish and the capital evaluation in the “large-scale industry” (Marx, 1990) will be necessary. Then, two different perspectives on the reproduction of capital when it is associated with “immaterial goods” like knowledge and information will be presented. First, the capitalistic enterprise will be present as a rent-seeking activity that aims to “enclose” the “general intellect” (Paulani, 2016). Then, the concept of “unmeasured value” (Prado, 2005) will be explained, and its implications for value creation and capture. All of them will be accompanied by their respective perspectives on labor subsumption.

After that, each area of AI will be explored on its concrete practices: processing power, data, and labor/software. This work argues that these areas may/should be interpreted through updated marxian categories to form a comprehensive picture of the AI industry, from which concrete analysis may derive.

Capital accumulation and fetishism

In his critique of political economy, Marx (1990) presented a comprehensive explanation of how capitalism works. To the author the wealth in this kind of society “appears as an 'immense collection of commodities’” (Marx, 1990, p. 125), and people relates to themselves by producing and exchanging commodities.

As something visible, the commodity has an objective property, it needs to be useful, but its utility is not something merely subjective, but social. Rice and beans, but also rock concerts or software to be commodities need to have a social utility, need to be desired by society.

But if they are commodities, they also have a price what is the expression of its value, but the objectivity of value is, on the contrary, imperceptible by the senses, “We may twist and turn a single commodity as we wish; it remains impossible to grasp it as a thing possessing value.” (Marx, 1990, p. 138). Value, then, is an objective but social characteristic of the commodities.

> “commodities possess an objective character as values only in so far as they are all expressions of an identical social substance, human labour, that their objective character as values is therefore purely social.”
> (Marx, 1990, p. 138-139)

So, when analyzing capitalism's elementary form, Marx shows how in capitalism we tend to naturalize what is social. He shows how human sociability through the market becomes the source of the fetish; and how the property of having value gives the fetish character to the commodity (and to money, to capital, and its forms - land, means of production, workforce, etc. – as well).

In this sense, Marx shows that in capitalism private labor is a moment of a global social network, and this network is produced by exchanges. These indirect social relationships are seen by producers not as direct social relations between persons in their work, "*but rather as material [dinglich] relations between persons and social relations between things.*" (Marx, 1990, p. 166).

So capitalist relationships are indirect and generalized and happen only inside exchange, only when private labor becomes social labor as well.

> Equality in the full sense between different kinds of labour can be arrived at only if we abstract from their real inequality, if we reduce them to the characteristic they have in common, that of being the expenditure of human labour-power, of human labour in the abstract.
> (Marx, 1990, p.)

The world of commodities – which is a highly complex system of reified social relations – comes from human activity. It is an unintentional, blind social creation that becomes autonomous and begins to dominate its creator. Thus, this world itself also bears the character of a fetish. The man who lives in this world, as always, wants to survive. To do so, he becomes dependent on the system that he recreates every day. He becomes, in other words, a creature of his own creation.

So, commodity, money and capital are, for Marx, objective social creations that subsume humanity, which is subject to its desires. In the case of capital - a dead thing that subjects living people (be they the owners of capital or the workers) - its objective is to constantly value its value (increase the profits), in an automatic, autonomous, and autocratic way. To achieve this, as we know, it is necessary to exploit living human labor, and capitalism does this in different ways.

Historically, primitive accumulation occurred (and still occurs) through force, robbery, rapine, colonialism, etc. But to maintain this modern form of social coercion it was necessary at the same time to dispossess workers of their means of production everywhere. Whether "free" or enslaved, workers had no other choice but to work for the capitalists.

Beyond primitive accumulation, in the initial moments of capitalism, which Marx calls manufacture[3], the production process was still subordinated to the skills and technical knowledge of the worker, who still had a lot of power because he somewhat dominated the work process. Consequently, the capitalist could only increase his profits by increasing working hours.

As capitalism evolves, manufacture becomes large-scale industry. The workforce, which was once subsumed to only a formal form of control, becomes obedient to the machine. The production process is incorporated into the machinery and the worker's skills become less relevant. This allows the capitalists to exchange expensive and specialized workers for a cheap, unskilled, and easily replaceable workforce.

Due to the competition between capitalists, the reduction of the cost of labor becomes an imperative, which leads to the reduction of salaries to the minimal necessary for the survival and reproduction of the working class. After reaching such a limit, the capitalists need to achieve more surplus value through other strategies. They may do it through an absolute increase in the surplus value, which consists in simply expanding the labor time, or through a relative increase in surplus value, accomplished by an increase in the productive capacity of the worker through technology or, what is the same, by decreasing the costs of reproducing the working class (Marx, 1990).

3 As it is a presentation, this phase is not necessarily just historical, it is also part of Marx's logical presentation of capitalism.

The first capitalists that adopt new technology increase their profits or expand their participation in the market. As other capitalists adopt the same innovation, the unity price of the commodity falls, the advantage disappears, and the general profit rate drops. This tendency to a fall at the profit rate, generates successive and increasingly intense crises in capitalist societies (Marx, 1990).

Marx also attributes to this logic of capital competition another tendency of capitalism to produce a relative surplus population (or industrial reserve army) which becomes a social problem in a period of crisis.

These crises allow for capital to accumulate in the hands of fewer capitalists who, eventually, have to fight for resources and markets (Lenin, 2018). Two world wars (1914-1918 and 1939-1945) happened as a result of this tendency (Hobsbawm, 1995). After the Second World War, the United States of America, which rose as one of its winners, started engaging in vulturous investments in new technologies that, already in the 1970's, would lead to a new economic cycle (Prado, 2005). This was accompanied by the promotion of neoliberal policies that, by promoting "*infrastructural speculation, land grabs, climate change and renewed imperialism have resulted in growing levels of poverty, inequality and informalization*" (Neilson & Stubbs, 2011, p.2) allowed a new cicle of accumulation.

This new period is characterized by the prevalence of non-rival goods, which are privatized in the form of intellectual property or industrial secrets. The creation and capture of value of such commodity, despite having zero reproduction cost, are expensive and carry high risks. These new characteristics challenged the established marxian notions about value creation and capture, leading to new theories. Two of them are Paulani's (2016) theory of "rent-seeking capital" and Prado's (2005) theory of "Unmeasured value".

Rent seeking capital vs Unmeasured Value

Paulani (2016) approaches the question of immaterial goods by reviewing Marx's writings on the land enclosures that happened at the XIX century in England. The author argues that immaterial goods have zero cost of reproduction, therefore no value is created or added at the creation of each copy. To reproduce itself, the capital would have to "enclose" part of the general intellect that allows the creation of such goods in a similar manner to what a landowner would do to common farmlands. After taking control of it through legal instruments like Patents or Copyrights, the capital would rent the access, valuating itself through the same strategy that landowners adopted.

Prado (2005) disagrees with such an approach, even though recognizing the importance of rent-seeking capital, the author disputes the idea that no productive work is being made, and affirms that there is value being created and added to these immaterial goods. For Prado, immaterial goods such as software or medicine formulas are the result of intellectual labor that does not carry a direct relationship between abstract labor time and value created. Even though labor is still subsumed to capital, as labor in traditional industries, these new forms of commodities are now capturing value in a different manner.

For Prado (2005), immaterial goods mobilize the general intellect to create something new, whose value is not dependent on the time necessary for its production. The relationship between labor and time ceases to be sufficient to determine how much value is created (and, for the same reason, appropriated). Also, such products demand huge amounts of investments and carry high risks.

As a result of the scale of the value created, this new form of commodity can only be paid for in parts. Their surplus value is not contained in each exemplar of commodity, but in one big immaterial good that, while being sold, repays the invested capital and, after reaching a threshold, generates a surplus value.

Differently from traditional commodities, the uniqueness of such products makes their prices independent of the average amount of time necessary to produce them, not because of the reproduction cost, but because such a measure would not make sense. (Prado, 2005)

It's worth noticing that such products are not like land, buildings, or financial assets. In these cases, competent management would allow them to generate rent potentially forever. In the case of the immaterial goods in question, the use value is constantly reduced by the development of new solutions, demands and societal changes. Even if not devoured by competition, or made obsolete by changes in the technological paradigm, such assets are still doomed by the fact that patents and copyrights, legally, do not last forever.

In the following sections, we will approach the previously mentioned tripod of AI (Hardware - Data - Software) through these theoretical perspectives in order to verify what form the subsumption between capital and labor takes in each one of them. From this perspective, we hope to better understand how different activities work together to create these new forms of commodity that are the Artificial Intelligence models.

Tool of labor - Processing Power

Even though the creation of AI models is often regarded as an immaterial industry, the infrastructure and resources necessary for the storage and processing of the data that generates AI are physical and consume as many natural resources as any other industry.

The United States Environmental Protection Agency (EPA) affirms that data centers consume roughly 2% of US' and 1,3% of the world's electricity (Ebrahimi et al, 2014). This is expected to rise between 15% and 20% in the following decade (Shuja, 2017). processing is responsible for 42% of this consumption, storage is responsible for only 14%. Cooling answers for 15% of the consumed energy (Uzaman et al, 2017).

This is evidence of the cost of data processing. A study made by Strubell et al. (2019) showed that training one single AI model may be responsible for as much CO2 emissions as 125 airplane trips from New York to Beijing and back. This impact is so big that even motivated a strike in 2019, where tech workers demanded their employers to recognize the role of the tech industry in the consumption of fossil fuels (Dhar, 2020).

Tables 1 and 2 show how concentrated this activity is. Google leads by the number of servers it uses, distantly followed by Microsoft. Amazon, Hewlett Packard (HP), Facebook, Yahoo! and eBay together account for less than 25% of the total.

Tabela 1 - Number of servers.

Company	**Servers**	%
Google	2,376,640	69,2%
Microsoft	1,000,000	29,1%
Amazon	158,000	<0,1%
HP	380,000	<0,1%
Facebook	180,000	<0,1%
Yahoo!	100,000	0,1%
eBay	54,011	1,574%

Fonte: Uzaman (2019).

Considering who is providing the equipment, this concentration is also detectable, being the two main providers accounting for more than 50% of the market and five companies controlling more than 80% of it (Uzaman, 2019).

Tabela 2 - Server Manufacturer

Company	Revenue	Market Share
HP	3.839.527.072	24.8%
IBM	3,623,543,805	34,9%
DELL	2,074,167,350	14,3%
Cisco	646,100,000	3,3%
Oracle	574,712,435	4,1%
Others	2,904,637,567	18,7%
Total	13,662,688,230	100%

Fonte: Uzaman (2019).

These servers usually contain "*hordes of GPUs*" (Li, 2020, p.1). This hardware, initially developed for graphical processing in gaming, has a capacity for parallel computing superior to traditional CPUs. In 2007 NVIDIA released libraries that would allow to use this power to process different types of information, such as the Convolutional Neural Networks necessary to create Large Language Models like ChatGPT (Dean et al, 2018; Miailhe at al, 2017; Pandey et al, 2022).

The other resource necessary for operating these data centers is labor. Even though data centers require a lot of employees during their construction (Mayer, 2023), the number of employees necessary for running the facility is relatively low. For instance, security workers are a significant part of the workforce, that, after being built, is "*a fortress – it had barbed wire, cameras, and guards, among other security measures.*" (Mayer, 2023, p.5).

Mayer (2023) explains that the center is run by technicians who are driven by "*Algorithms [who] managed their work by instructing the shift guys where to locate a memory card and where to insert a new one in the massive hall of blinking servers.*". The workers in these facilities report constant boredom accompanied by the tension created by the expectation of something going wrong, which generates an overload of work (Mayer, 2023).

Also, finding skilled labor is challenging, partly because the educational system produces workers who are either under-skilled or over-educated to do these tasks. But also because people do not know about the existence or nature of the activities performed by such facilities. Mayer (2020) argues that such invisibility is made by design, and "reinforces the aura of high-tech hegemons" (Mayer, 2023, p.5) of companies such as Google or Microsoft, making invisible the infra-structure that allow such companies to exist.

This secrecy is part of the policy in all major companies. Microsoft, Amazon, Apple, Meta, and Alphabet, strictly forbid anyone to see the work that is being done inside their facilities (Wark, 2017). Scroggins & Pasquetto (2019) argue that such a situation of secrecy and invisibility reinforces the position of power held by employers.

In a contradictory situation, the managers of the above-quoted companies hold a policy to simultaneously warn about a labor shortage, and, at the same time, to promise data centers operated by computers and robots. This leads to a feeling of perpetual anxiety for the workers, who are constantly being told that their job is one innovation away from being extinguished, rendering useless skills and experience accumulated through years of dedicated work (Mayer, 2023).

Plantin (2021) highlights that such environments also witness what he calls “micro resistance” episodes, in which workers change the working procedures to have more free time (reordering), take pride on their abilities, or just stop working to socialize.

Data Centers are a capital-intensive part of the AI phenomena, in which the labor is under formal and real subsumption. Workers feel both estrangement and alienation from work. The fetishism around the concept of cloud computing is promoted through the invisibilisation of both the capital and work that allow it to exist.

Object of labor - Data

During the last decades, society witnessed a vertiginous increase in the creation of information. Social networks like Facebook have more than 3 billion users that are constantly producing text, video, and interacting with other users' information. On YouTube, 400 hours of videos are posted every minute (Statista, 2024).

AI models cannot be trained on AI-generated data, they must have real-life data to be consistent and updated (Rao, 2023). The most straightforward method to acquire labels to data is paying people to analyze thousands of files and classify them, a task often outsourced to third-world countries, where low-paid workers spend plenty of time classifying data (Le Ludec et al, 2023).

Data can also come from the Internet of Things, which relies in sensors present in several devices that are connected to the internet (Ahmed, 2017). Demographic and GPS data are also extracted from personal devices such as Smartphones and, after being organized and correlated with social media information, used to create clusters of users with specific consumption behaviors (Ahmed, 2022; Kaabi, 2019; Muhammadian, 2020) or political beliefs (Risso, 2018). These information is sold to advertisers or directly used to train algorithms that aim to deliver specific content, keeping the users at the platforms as much as possible.

Social media appears as a major source of data. Media platforms are fundamentally different from traditional media. A television show may exist without viewers, a social media does not exist without engaged users. If traditional media has to pay professionals to produce content, social media gets its content, mostly for free, from its users (Fuchs, 2014; Scholz, 2012).

All of the content that is being consumed is created by other users. This is the opposite that happens in traditional media, in which the content is created through paid labor and then offered to the users. The platform owners collect data from the users' activity and also from the reaction of other users to the flow of new and old text/photos/videos. These reactions given by users at social networks such as Facebook or Instagram are used as labels to train ML models who can identify which words or image patterns generate feelings such as depression, anxiety, engagement, or any other desired emotion in a person with a given profile (McStay, 2020).

But, aside from free labor, there is also paid labor in social media (Kopf, 2020). Platforms such as YouTube, Facebook, and Instagram recurrently reward content creators capable of reaching a wide public. This relationship is subordinated, as the content creator should follow the platform guidelines. The platform subsumes the work of the creators by increasing or reducing their visibility through changes in the algorithm, forcing the creators to have productivity quotas and/or to create a specific type of content. (Craig, 2019).

Nevertheless, being paid to produce online content is a distant reality for most users (Duffy, 2021). Alexander (2020) argues that most of the time promises of payment are used to attract and motivate people to work for free, creating content for such platforms, in the hope of being paid after becoming famous.

Fuchs (2015) argues that not only the expectation of being paid is capable of making people work for free, but also the activity of common users may be considered as unpaid labor capable of producing value. Even though authors such as Bolaño (2015) reject such an idea, Fuchs reiterates that value is being created while users are on the platform, and that the extraction of value from this unpaid labor seems to have a direct correlation with the time spent on the platforms by its users (Fuchs, 2015).

Authors like Scholz (2012) state that there is coercion in this process. He presents the concept of “violence of participation”, arguing that, as new forms of socialization become the norm, engaging in such platforms becomes non-optional. As Andrejevic (2012) puts:

> When we are separated from the means of socialization, this does not mean we do not have access to them; rather, we come to rely upon the provision by their parties of technologies for socialization that separates us from the information upon which our social lives rely
>
> Andrejevic (2012, p.202)

This new form of socialization is characterized by the necessity of sharing information and creating content on social media. This content will become part of the platform owner's assets, turning itself into a product external and outside the control of its creator, which is Marx’s definition of alienation (Ayres, 2007). Also, the way that unpaid content creators engage in social media is modulated by the platforms in a similar manner to the content made by paid creators. The exposition, psychological rewards, or punishments are present and have a significant impact on the users' mental and physical well-being (Duffy, 2021).

Andrejevic (2012) argued that such a process becomes a necessity, crucial for the work of the user and therefore for the reproduction of life. In social media like LinkedIn and GitHub such relationships become more apparent. LinkedIn is currently the main platform for professional business, and being outside of it is harmful to anyone searching for a job (Davis, 2020). Users have to engage in conversations, update their profiles actively search for more visibility on it as a form of becoming good assets for a given company (Twenge et al, 2020).

Another example of imposed participation is GitHub. The platform started as a tool for version control in software development, and allowed for Free and Open Source Projects to be hosted without cost. After being bought by Microsoft in 2018, it has become one of the most important “Electronic Portfolio” for new developers (Hokkanen, 2023). There, 420 million repositories (284 million of them public and 65 thousand in generative AI projects) are available. A total of 4,5 billion contributions were made to projects in 2023 (GitHub, 2024). These contributions are code that is written by GitHub users, in order to create new Free/Open Source Software or improve existing projects. These repositories may contain a couple of lines or thousands of lines of code, most of the time accompanied of documentation and discussions about it.

All this GitHub data is perfect for creating Large Language Models (LLM) such as the ones employed by ChatGPT or Copilot. Both products are capable of understanding natural language[4] and writing complex code by themselves (Wu et al, 2023).

AI is dependent on data as any other industry is on its raw materials. It can't create its own inputs from anything; therefore, human labor is necessary for the creation of new datasets. This data can be commodified, and therefore the process of generating such data generates value. This data comes from a series of sources, but always demands some form of human labor, paid or unpaid. A significant part of the data is extracted from social media, some of it is paid, but most is not. The unpaid data comes mostly from human participation in social media. Such engagement is not optional if one considers a socialization process coherent with a given historical time as a human necessity.

Labor - The Data Scientist

An AI is a mathematical/computational model created based on high processing power and big amounts of data. The coordination of the interaction between hardware and data is the role of the Data Scientist. The term appeared during a Facebook board meeting, when the necessity for a new type of professional arose. (Patil, 2011)

Patil (2011) states that a data scientist must have technical expertise, curiosity, storytelling (the ability to communicate his findings effectively), and cleverness (the ability to look at a problem in different ways). Patil does not restrict data scientists to professionals from the field of computer sciences, believing that multidisciplinary teams are the best approach. The Data Scientist creates AI models by using preexisting software (usually Free and Open Source Software - FOSS) to coordinate processing power and data.

Costa & Santos (2017) have conducted an extensive review of several sources, and came to the following definition:

> The results indicate that a Data Scientist is a multidisciplinary profile that searches for knowledge in several fields of study. It heavily relies on the scientific way of doing things, therefore, research experience is significantly relevant. Data Scientists must also be able to deal with big data in all the stages of data flow, and topics like ethics, privacy and security should also be constantly on their minds. Aspects related to the computing field such as artificial intelligence, machine learning, programming, databases and other data driven aspects have also a strong presence in this profile. In order to communicate findings, Data Scientists must have strong social and personal capabilities, like communication, business acumen and curiosity. According to the

4 Natural language is human language, the language that normal people use.

> proposed conceptual model, we can define a Data Scientist as someone who is able to extract patterns and trends from data, through certain data-related tasks, regardless of its characteristics and challenges.
>
> (Costa & Santos, 2017, p.733)

Hazzan (2023) agrees with Costa & Santos in almost everything but adds cognitive skills such as lifelong learning, the capacity to think in different levels of abstraction, and to perform proper dataset exploration. The author also mentions teamwork and cooperation as necessary qualities of this kind of professional.

Regarding demographics, in the US, most data scientists are male (79%), White or Asian (80%). Fifth percent are older than 40, and 51% have a Bachelor's degree. Among them, 34% have a master's degree and 13% have a doctorate. Male average income is $108.000 a year, while the average income of a female data scientist is $103.000. Black data Scientists receive an average compensation of $104.000, having the lowest wages in the category (Zippia, 2024b).

A Data Scientist earns significantly more than the average US citizen ($74.000 a year) (Statista, 2024b) and is more educated, however, structural differences due to race and gender are still present. The fact that almost half of Data Scientists do not hold a Bachelor's degree shows a high number of professionals are self-taught. This is not uncommon in the field of computer science, in which 66% of the US professionals do not hold a Bachelor's degree (Zippia, 2024a), but is significantly higher than in other areas with high salaries.

These definitions show that a Data Scientist is not someone who just follows commands in a preexisting methodology. The professional must be capable of coordinating social, business, and technical demands. They have to do so while working in a team, which explains why in Brazil, as put by da Silveira et al (2020), the main requisite to be hired as a Data Scientist is experience at the field.

This is not a type of work that is under classical real subsumption present in large-scale industries (Marx, 2015), because the worker has control over great parts of the productive process, but they are still formal subsumed as it was in the pre-industrial era, as the worker works for someone who pays their wages, and it comes along with real subordination as they workers does not control all the productive process, relying on a sophisticated net of professional cooperation among peers and also with different sectors inside a corporation to perform their tasks.

The data scientists hold a crucial node at the company, and know critical information about the most relevant parts of the products that they create. They have control over the work process and the results of it are not predictable. The creation of every new AI model (or product) involves a significant amount of risk, trial and error, and uncertainty. Their success is therefore dependent on the ability of the Data Scientist to mobilize the general intellect of the social body. A condition anticipated by Marx that, in his non-published drafts of the Foundations of the Political Economy (*Die Grundgrisse*) wrote:

> In this transformation, it is neither the direct human labour he himself performs, nor the time during which he works, but rather the appropriation of his own general productive power, his understanding of nature and his mastery over it by virtue of his presence as a social body – it is, in a word, the development of the social individual which appears as the great foundation-stone of production and of wealth.
>
> (Marx, 2005)

Marx imagined that this process could lead to more freedom for everyone, and the end of exploitation due to the destruction of the value as a social relationship, that allows the owners of the means of production to steal time from the working class. But this is not the case.

In our view, this is, therefore, a new form of value creation. The products created by the data scientist are still controlled by the company. Although not exactly estranged, the professionals are still under an alienating condition. The means of production (Tools of Labor - Processing Power, Subject of Labor - Data) and the labor are still under the control of the capital.

What changes is the set of techniques used to subsume this labor. It's necessary to engage the workforce's life as a whole in the production process because their intellect is the main source of value. Ones intellect is difficult to control (we can use our brain to things outside work, to the concurrence, or for our own), and value creation no longer depends on the quantity of time explored, but the quality of the work. Therefore, instead of controlling the worker's actions during the period that the worker is at the service of the company, the capitalist now has to control all of the workers' subjectivity, their free time, their network of friends, and affections, as argued by Prado (2005):

> The workforce that mobilizes knowledge and ensures that production does not stop, engaging its own subjectivity in the production process, is no longer perfectly suited to the exploitation of capital. It is for this reason that the domination of capital, far from softening, must become uncompromising and all-encompassing, extending not only over

> working time but also beyond that time, into the life of the worker as a whole.
> (Prado, 2005, p.136)

Prado evolves his argumentation explaining the necessity of keeping the workforce in a competitive state. Lifelong learning, even though a concrete necessity in a moving economy, becomes also a form of hierarchization of the workforce that facilitates control. At the same time, the workers have to compete and cooperate at the same time, which leads them to a “cooperative rat race” (Prado, 2005, p.134).

Data scientists are highly qualified professionals whose function is to coordinate computational resources, theoretical concepts, creativity, and social nets to create a product that is immaterial and has zero cost of reproduction (an AI model). Even though they a have high level of control over their work routine, they are still intellectually subsumed to capital through new and sophisticated means, that aim to control not only their time but also their subjectivity.

Conclusion

In this work, we tried to offer an overview of the AI phenomena through the lens of the Marxian Political Economy. Most of the attention was aimed at deconstructing the vulgar perception of AI as a technology that is ethereal and independent of human labor. We showed that there isn't a single aspect of AI that is not permeated by concrete (and also abstract) human labor.

We decomposed AI into its “raw materials” (data), tools of labor (processing power), and labor itself (the data scientist).The processing power was presented as performing the function of tool of labor, which means, technical competence crystallized in the form of machines that perform a given task. They demand labor to be maintained, whose subsumption can be considered mostly “real”, as in large-scale industry.

The data is treated as the object of labor, which will be transformed in the process of production. In this case, as it is in other industries, it demands labor to be extracted or created, which may be paid or unpaid. The subsumption may be both material and intellectual, depending on the case, especially when considering the content creators. Their work may generate commodity (datasets) or, when being done inside the same company that uses it to create AI models, it should be treated as part of a bigger productive process, being, therefore, productive work.

Finally, the AI models, in which the data scientists work, are the final and more expensive commodity. They have huge costs of production but nearly zero cost of reproduction, so capital cannot evaluate it by selling copies of them. They must be sold and paid in small installments, in which each consumer pays for the right to access the tool (for example, a certain amount per month) until more value is generated (Prado, 2005). In this type of business, the financial return is not necessarily related to the time spent building the AI model, and can be little or huge, depending on the company's ability to impose itself. As it is a risky business, the companies crowdsource the costs of production of the model and remunerate the capital that was invested in it while avoiding the risk of spending money on failed attempts. The final product is always unique and unrelated to the time applied to its production, so there is no center around whom its price can gravitate. The surplus value is limited only by the technological depreciation of the model, which allows for unmeasured gains..

When approaching AI through the lens or Marxian Materialistic view, the idea that it is something that exists by itself disappears. It becomes clear that it is a form of capital, which means that it is an expression of a social relationship, that there is labor that is subordinated and creates surplus value. This surplus value is unmeasured, but can also vanish as the AI model becomes a public good.

Bibliography

ABDALLA, M.; ABDALLA, M. The Grey Hoodie Project: Big tobacco, big tech, and the threat on academic integrity. In: **Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society**, [S.l.], p. 287-297, 2021.

AHMED, A. Marketing 4.0: The Unseen Potential of AI in Consumer Relations. International Journal of New Media Studies: **International Peer Reviewed Scholarly Indexed Journal**, v. 9, n. 1, p. 5-12, 2022.

AHMED, E.; YAQOOB, I.; HASHEM, I. A. T.; KHAN, I.; AHMED, A. I. A.; IMRAN, M.; VASILAKOS, A. V. The role of big data analytics in Internet of Things. **Computer Networks**, v. 129, p. 459-471, 2017.

ALEXANDER, J. Creators finally know how much money YouTube makes, and they want more of it: More than $34 billion in advertising revenue in three years. **The Verge**, 2020.

ALLEN, G. Understanding AI technology. Joint Artificial Intelligence Center. JAIC. **The Pentagon United States**, 2020.

ALLEN, J. F. AI growing up: The changes and opportunities. **AI magazine**, v. 19, n. 4, 1998.

ANDREJEVIC, M. Estranged free labor. In: **Digital labor**, pp. 149-164. Routledge, 2012.

AYRES, I. Super crunchers: How anything can be predicted. John Murray, 2007.

BENAICH, N.; HOGARTH, I. **State of AI report**. London, UK, 2020.

BOLAÑO, C. R.; VIEIRA, E. S. The political economy of the internet: Social networking sites and a reply to Fuchs. **Television & New Media**, v. 16, n. 1, p. 52-61, 2015.

COSTA, C.; SANTOS, M. Y. The data scientist profile and its representativeness in the European e-Competence framework and the skills framework for the information age. **International Journal of Information Management**, v. 37, n. 6, p. 726-734, 2017.

CRAIG, D. Creator management in the social media entertainment industry. In: Making media: Production, practices, and professions, p. 363-374, 2019.

DA SILVEIRA, C. C. et al. What is a Data Scientist? Analysis of core soft and technical competencies in job postings. **Revista Inovação, Projetos e Tecnologias**, v. 8, n. 1, p. 25-39, 2020.

DAVIS, J., WOLFF, H. G., Forret, M. L., & Sullivan, S. E. (2020). Networking via LinkedIn: An examination of usage and career benefits. Journal of Vocational Behavior, 118, 103396.

DEAN, J.; PATTERSON, D.; YOUNG, C. A new golden age in computer architecture: **Empowering the machine-learning revolution**. IEEE Micro, v. 38, n. 2, p. 21-29, 2018.

DHAR, P. The carbon impact of artificial intelligence. Nat. Mach. Intell., v. 2, n. 8, p. 423-425, 2020.

DUFFY, B. E.; SAWEY, M. Value, service, and precarity among Instagram content creators. In: **Creator culture: An introduction to global social media entertainment**, p. 135-142, 2021.

EBRAHIMI, K.; JONES, G. F.; FLEISCHER, A. S. A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. **Renewable and Sustainable Energy Reviews**, v. 31, p. 622-638, 2014.

FOSTER, D.; ARNOLD, Z. Antitrust and Artificial Intelligence: How Breaking Up Big Tech Could Affect the Pentagon’s Access to AI. **Center for Security and Emerging Technology. CSET.** Issue Briefing, 2020.

FUCHS, C. Social media and productive labour. In: **Culture and Economy in the Age of Social Media**, p. 118-204. Routledge, 2015.

FUCHS, C. Digital prosumption labour on social media in the context of the capitalist regime of time. **Time & Society**, v. 23, n. 1, p. 97-123, 2014.

FURMAN, J. et al. AI and the Economy. **Innovation policy and the economy**, v. 19, n. 1, p. 161-191, 2019.

GITHUB. **100 million developers and counting**. Disponível em: https://github.blog/2023-01-25-100-million-developers-and-counting/. Acesso em: 03 jan. 2024.

HAZZAN, O.; MIKE, K. Professional Skills and Soft Skills in Data Science. In: G**uide to Teaching Data Science: An Interdisciplinary Approach**, p. 165-178. Cham: Springer International Publishing, 2023.

HOBSBAWM, E. J.; CUMMING, M. **Age of extremes: the short twentieth century**, 1914-1991. London: Abacus, 1995.

HOKKANEN, S. The Impact of Electronic Portfolios on Employability of New Full Stack Developers: Insights from Students and Employers, 2023.

KAABI, S., & JALLOULI, R. (2019). Overview of E-commerce technologies, data analysis capabilities and marketing knowledge. In Digital Economy. Emerging Technologies and Business Innovation: 4th International Conference, ICDEc 2019, Beirut, Lebanon, April 15–18, 2019, Proceedings 4 (pp. 183-193). Springer International Publishing.

KOPF, S. “Rewarding good creators”: corporate social media discourse on monetization schemes for content creators. **Social Media+ Society**, v. 6, n. 4, p. 2056305120969877, 2020.

LE LUDEC, C.; CORNET, M.; CASILLI, A. A. The problem with annotation. Human labour and outsourcing between France and Madagascar. **Big Data & Society**, v. 10, n. 2, 2023.

LEE, I.; SHIN, Y. J. Machine learning for enterprises: Applications, algorithm selection, and challenges. **Business Horizons**, v. 63, n. 2, p. 157-170, 2020.

LENIN, V. **O imperialismo etapa superior do capitalismo**. 1977. 2018.

LI, R. Artificial intelligence revolution: How AI will change our society, economy, and culture. Simon and Schuster, 2020.

MARX, Karl. **Capital: A critique of political economy**, Volume 1. Penguin UK, 1990.

MARX, K. Grundrisse: **Foundations of the critique of political economy.** Penguin UK, 2005.

MAYER, V. When do we go from here? Data center infrastructure labor, jobs, and work in economic development time and temporalities. **New Media & Society**, v. 25, n. 2, p. 307-323, 2023.

MCSTAY, A. Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. **Big Data & Society**, v. 7, n. 1, 2020.

MIAILHE, N.; HODES, C. Making the AI revolution work for everyone. **The Future Society at the Harvard Kennedy School of Government**, Cambridge, MA, 2017.

MONETT, D.; LEWIS, C. W. P. Getting clarity by defining artificial intelligence—a survey. In: **Philosophy and theory of artificial intelligence**, pp. 212-214. Springer International Publishing, 2018.

MUHAMEDYEV, R. Machine learning methods: An overview. **Computer modelling & new technologies**, v. 19, n. 6, p. 14-29, 2015.

MUHAMMADIAN, R. Artificial intelligence in marketing. How AI is Revolutionizing Digital Marketing, 2020.

PANDEY, M. et al. The transformational role of GPU computing and deep learning in drug discovery. **Nature Machine Intelligence**, v. 4, n. 3, p. 211-221, 2022.

PATIL, D. J. Building data science teams. O'Reilly Media, Inc., 2011.

PAULANI, L. M. Modernity and capitalist progress in the periphery: The Brazilian case. **European Journal of Social Theory**, v. 19, n. 2, p. 210-227, 2016.

PLANTIN, J. C. (2021). The data archive as factory: Alienation and resistance of data processors. Big Data & Society, 8(1), 20539517211007510.

PRADO, E. F. S. **Desmedida do valor: crítica da pós-grande indústria**, 2005.

RAO, R. AI-generated data can poison future AI models. **Scientific American**, v. 28, 2023.

RISSO, L. Harvesting your soul? Cambridge analytica and brexit. **Brexit Means Brexit**, 2018, p. 75-90.

SCHOLZ, T., ed. Digital labor: **The internet as playground and factory**. Routledge, 2012.

SCROGGINS, M.; PASQUETTO, I. Labor out of place: On the varieties and valences of. in. visible labor in data-intensive science, 2019.

SHUJA, J. et al. Sustainable cloud data centers: a survey of enabling techniques and technologies. **Renewable and Sustainable Energy Reviews**, v. 62, p. 195-214, 2016.

STATISTA (a). **Hours of video uploaded to YouTube every minute**. Disponível em: https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/. Acesso em: 03 jan. 2024.

STATISTA (b). **Wages and salary**. Disponível em: https://www.statista.com/topics/789/wages-and-salary/#topicOverview. Acesso em: 03 jan. 2024.

STRUBELL, E.; GANESH, A.; MCCALLUM, A. Energy and policy considerations for deep learning in NLP. arXiv preprint arXiv:1906.02243, 2019.

TWENGE, J. M. et al. Underestimating digital media harm. Nature Human Behaviour, v. 4, n. 4, p. 346-348, 2020.

UZAMAN, S. K. et al. A systems overview of commercial data centers: initial energy and cost analysis. **International Journal of Information Technology and Web Engineering**, v. 14, n. 1, p. 42-65, 2019.

WANG, P. On defining artificial intelligence. **Journal of Artificial General Intelligence**, v. 10, n. 2, p. 1-37, 2019.

WARK, M. What if this is not capitalism any more, but something worse? NPS Plenary Lecture, APSA 2015, Philadelphia, PA. **New political science**, v. 39, n. 1, p. 58-66, 2017.

WU, T., HE, S., LIU, J., SUN, S., LIU, K., HAN, Q.L. and TANG, Y., 2023. A brief overview of ChatGPT: The history, status quo and potential future development. **IEEE/CAA Journal of Automatica Sinica**, 10(5), pp.1122-1136.

ZIPPIA (a). **Computer scientist jobs**. Disponível em: https://www.zippia.com/computer-scientist-jobs/demographics. Acesso em: 03 jan. 2024.

ZIPPIA (b). **Data scientist jobs**. Disponível em: https://www.zippia.com/data-scientist-jobs/demographics/. Acesso em: 03 jan. 2024.

Agradecimentos e fontes de financiamento

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001