PROGRAMA
DE PÓS-GRADUAÇÃO
EM ECONOMIA
FE/UFJF

ufjf
Universidade
Federal de Juiz de Fora

# Close Enough to a Jobless Society: Reflections on Historical Materialism and Artificial Intelligence

Rogério Silva de Mattos - UFJF

Faculdade de Economia

[TD.007/2018]
Juiz de Fora
[2018]

# Close Enough to a Jobless Society:
## Reflections on Historical Materialism and Artificial Intelligence*

**Rogério Silva de Mattos**
Departamento de Economia
Universidade Federal de Juiz de Fora
Juiz de Fora, Brazil
rogerio.mattos@ufjf.edu.br

July 2018

* My work on the present paper started a year ago and was motivated by the reading of another paper written by Hillel Ticktin (2017) on the persistent crisis of capitalism. I wish to thank Lucas Cavalcanti, Eduardo Gonçalves, and Nelio Oliveira for comments on previous drafts of the present paper, without ascribing to them any remaining errors and omissions.

## Abstract

Advancing artificial intelligence (AI) and mobile robotics are producing intelligent machines and algorithms that are better ‘workers’ than humans. The resulting perspective of rising technological unemployment is producing enormous anxiety in the world’s society. Capitalism has always created new jobs during the technological revolutions of the past, but many technology experts believe things will be different this time. The paper uses the historical materialism theory to analyze this context and develop some perspectives for the future of capitalism and employment. Two concepts of this theory are used: ‘relations of production’ and ‘development of productive forces’. Concerning the latter, the exponential increase in computer capacity, the advents of the Internet Cloud and Big Data, and the developments in the AI subfield of machine learning are likely to create conditions for the settlement of a new mode of production in a very long–run future. Concerning relations of production, new possibilities of social production and the emergence of new business models and collaborative schemes may catalyze this process, but understanding the trends involved demand intensive research efforts. An extreme scenario of a jobless society is built and analyzed. It provides relevant clues for the risks incurred by society in an alternative, but more realistic, scenario that is close enough to it. The close enough scenario features a situation in which human labor is replaced by intelligent machines and algorithms at a hight rate but not in full, so that even is this case society is led to a collapse of distribution. Strategic paths for society to avoid it are discussed in the end.

“Labor is not the source of all wealth.”[1]

Karl Marx.

The goal of the future is full unemployment,
so we can play.

Arthur Clarke

## 1. Introduction

The exercise of futurism is often viewed as a premature and fruitless effort. Skeptics claim that futurist predictions are often wrong and serve only to promote bewildering contentions. It happens, they argue, because from now to any point in the future reality can take a multitude of paths, the most of which are unimaginable to us. This is a point indeed, and it is also the case that futurist debates often lead to speculative overstatements. Nevertheless, we are cursed by reality to live forever with uncertainty, and this makes some speculation to be better than do nothing. This is why governments, corporations, other profit and nonprofit institutions, and people are always trying to look into the future. Besides allowing the identification of opportunities, it can be a hedge against being taken by surprise. And it gets overly important as we realize that the cost of being taken by surprise is substantial.

Modern Artificial Intelligence (AI) is producing intelligent machines that are better “workers” than humans. When we consider the present trend of technological developments intensively based on AI and the resulting slow growth of employment opportunities even for highly educated workers, the cost of being taken by surprise escalates. Since the Industrial Revolution, humanity has seen giant step advances in technology but none which was able to put in check capitalism as a social production (and distribution) system. Hitherto, capitalism has always found ways to create new jobs, even during periods of technological revolutions when it was intensively replacing old jobs with machines and automation. Yet, the rapidly advancing technology based on AI seems to be placing a more serious challenge. Nowadays, technology is producing material and virtual machines endowed with AI capabilities that do many kinds of human jobs more competently than humans[2]. As a consequence, not only unskilled but also skilled and highly skilled workers have been replaced by intelligent machines, and it is an upward trend.

Such a scenery is unsettling because it points at a situation in which human labor becomes largely dispensable as an input to the economic system. With a bit more of speculation for somewhere in the future, a situation in which entrepreneurial and firm management skills may also become dispensable. All of this may sound a too futurist perspective, notwithstanding many technology experts believe now that advancing AI is on the go to harshly change modern days’ capitalism. It can be measured by a recent explosion

[1] Marx (1857,2017).
[2] Future of Life Institute (2017).

of books, academic papers, news articles, sci–fi movies, documentary films, and many lectures and panel debates delivered on the subject[3].

Many experts and observers fear certain risks for humanity, like singularity and the danger that super intelligent machines and systems come to subjugate humankind. However, other concerns, which matter most for us in this document, regard the impacts on employment and the capitalist system. In a recent book, technology expert Martin Ford (2015) shows evidence that, in the last decade, GDP growth in the US has not been accompanied by an increase of labor working hours. Ford (2015) also makes the point that AI is to promote disruptive technological changes that will displace highly skilled workers in areas such as medicine, health services, law, and finance, among many others. An earlier though often quoted study by Frey and Osborne (2013), from Oxford University, estimates that 47% of present jobs in the US will simply disappear within two decades as a consequence of computerization[4].

For long, the tendency of technological developments to displace human labor from social production has preoccupied economists, social scientists, and politicians. Much has been written on the subject and the literature abounds with controversies polarizing capitalist believers and non-believers at extreme sides. Nonetheless, we discuss it once more in this document by trying to look into particular features displayed by advancing AI technology. It seems like the present challenges are near to burst the very contradiction between capitalism and the ultimate purpose of an economic system that is to meet human society's needs. It is well known that, within capitalism, the primary goal of social production is to generate profits for the owners of the means of production and not to fulfill the necessities of society's members. Such contradiction is leading to muddle because, while advancing AI technology is going to boost economic productivity in the next decades, the distribution of social output, based on exchange[5], is going to collapse. In other words: a paradoxical future approaches in which the glory of production comes along with the apocalypse of distribution.

In order to analyze this contradiction, we chose Karl Marx and Frederic Engels' theory of historical materialism (HM) as a leading thread. Such a theory provides a point from which to start a discussion on advancing AI technology and the future of modern society[6]. Grounded in dialectics, HM theory predicts certain developments that take place in society in response to economic contradictions. Its core message is that technological advances create new

[3] It is noteworthy that in the last three years The World Economic Forum Annual Meetings – held at Davos, Switzerland –, delivered special panel sessions on artificial intelligence and universal basic income (World Economic Forum, 2017a, 2017b, and 2017c). The latter was motivated by growing concerns on how to respond to perspectives of technological unemployment in the foreseeable future.

[4] Frey and Osborne (2013) refer to computerization in a broad sense as job automation by means of computer-controlled equipment, which include non–AI and AI-based computer systems.

[5] A major outcome of the division of labor within society is the development of exchange. In some primitive and pre–capitalist societies, its members trade leftovers, say, the share of their individual production which exceeds subsistence needs. A merchant society is a more developed one in which the division of labor features exchange as a *necessary form* of distributing social output. Almost all goods and services are produced for exchange and so the subsistence of each individual depends on it. Marxists state that the most developed merchant society is the capitalist one, where most individuals are dispossessed of the means of production. Having to trade only their workforce, they need to work against a salary to have access to social output. The collapse of distribution we refer to here means that in a jobless future most individuals will have literary nothing to trade, not even their workforce, and so will be prevented to access social output.

[6] This document was written by a non–Marxist, though a one who acknowledges the intellectual contributions of Karl Marx and Frederic Engels for our understanding of modern society. By a non–Marxist, it is meant a one who uses many elements of the usual Marxian model of reality interpretation but not all of them. For instance, within our reflections we do not adhere to the theory of labor–value and its derivatives. Although we regard it another relevant contribution of Karl Marx, we find it difficult to apply in the context of labor markets featured by distinct types of labor (unskilled and skilled) that started to prevail since the beginning of the 20th century.

production possibilities that give birth to new relations of production. As the latter develop and spread across the economic system, a big pressure is placed over society to change the prevalent property relations so as to give way to a new mode of production. We argue here that, more than any previous technological revolution, advancing AI is producing effective conditions to bring about a new, post–capitalist mode of production. However, a transition to a new mode will depend on many things, and the purpose of this document is to develop some reflections on the subject.

In addition to using HM theory, we divide our reflections into two branches of analysis. The first tries to understand how AI developments can create new conditions to organize social production in a way that is different and more advanced than capitalism. For so, we develop an extreme and limiting case scenario for the future, comprised by a fully jobless society, and discuss the perspectives of alternative and more realistic scenarios that are *close* to it. The second branch tries to understand how non–capitalist relations of production that are under development now, and others which may develop in the future, can produce a convergence towards the limiting case scenario considered in the first branch. We locate in the second branch the biggest challenges of analysis and the one which is most demanding of research efforts.

In addition to this introduction, the paper is organized as follows. Section two describes in brief the historical materialism theory. Section three recalls and put into perspective the major industrial revolutions of the past. Section four makes an appraisal of the effects of technological revolutions over jobs and employment. Section five develops on the recent artificial intelligence advances. Section six develops the concept of typically human skills from the notions of cognitive abilities and tacit knowledge. Section seven makes an introductory description of a neural network, its potentials, and its relation to tacit knowledge. Section eight describes the conditions that AI and technological advances are creating for the existence a new mode of production featured with a jobless society. Section nine introduces the concept of *close enough* scenario and discusses paths for a safe transition towards it. Section ten closes the paper with some final comments.

## 2. Historical Materialism

Theories of capitalist crises have for long supported Marxists' stances on a (presumed necessary) transition from capitalism to socialism[7]. However, we believe that at the present moment it is HM theory which provides a better explanation for a transition from capitalism to a new mode of production. Theories of crisis are usually based on failures of the capitalist system's functioning in the short– and medium–runs, as is the case, for instance, of underconsumption and profit squeeze theories. Instead, HM theory states that the transition from capitalism to another mode of production will result from long–run changes in the framework within which the capitalist system operates. Such changes are produced by capitalism itself, by persistently improving the productive forces until they get into conflict with the capitalist relations of production.

Karl Marx and Frederic Engels developed and presented the HM theory early in *The Poverty of Philosophy*, written solely by Marx and published in 1847, and *The Communist Manifesto*, written by both and published in 1848. This theory is further presented in different

[7] For a crisis interpretation of nowadays' capitalism with focus placed on political– and socio–economic forces operating in a medium– to long–run, see Ticktin (2017). In a sense, our approach in the present paper complements such a view by exploring the very long–run perspective of HM theory and AI developments.

parts of Marx's works[8]. We take here the famous summary in his *Preface of The Contribution to the Critique of Political Economy*, published in 1859, and which reads:

> " In the social production of their existence, men inevitably enter into definite relations, which are independent of their will, namely *relations of production* appropriate to a given stage in the development of their material forces of production. The totality of these relations of production constitutes the economic structure of society, the real foundation, on which arises a legal and political superstructure and to which correspond definite forms of social consciousness. The *mode of production* of material life conditions the general process of social, political and intellectual life. It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness. At a certain stage of development, the material *productive forces* of society come into conflict with the existing relations of production or -- this merely expresses the same thing in legal terms -- with the *property relations* within the framework of which they have operated hitherto. From forms of development of the productive forces, these relations turn into their fetters. Then begins an *era of social revolution*. The changes in the economic foundation lead sooner or later to the transformation of the whole immense superstructure." (Marx, 1859; italics are ours).

The text above is regarded by many as a theory summarization masterpiece. Notwithstanding, as highlighted by Katz (1993), the study of HM theory is complicated by the fact that Marx himself never provided a systematic treatment of its central principles. Therefore, the task of elaborating HM theory fell over Marx's interpreters, who tried to distill its tenets from Marx's historical writings and from general statements like those in the above citation. As a consequence, controversies exist among interpreters and different views of HM theory compete in the literature. In the sequel, we present our interpretation of the text cited above by filling in some gaps of explanation with our own understandings. We stress that it is a particular view of ours.

In order to exist, every society has to produce the goods and services that fulfill its material needs. The economic base of a society comprises not only the physical infrastructure used to produce those goods and services, but also the set of social relations among society's members in the production process, which was called by Marx as "economic structure". Above the economic base, there is a complex system, called by Marx as "superstructure", which comprises social, legal, political, and ideological (self–consciousness) dimensions of society's organization. Under the materialistic view, it is the economic base that determines the superstructure, and not the contrary as presumed, for instance, by the idealistic view of Hegelian philosophers which was in vogue at the time of Marx and Engels.

Across history, the economic base takes different forms, namely modes of production. It is a central concept in HM theory. A mode of production is a historically determined stage of the economic base. It features a particular stage of development of productive forces and, in connection with it, a particular set of relations of production. The productive forces are associated with the physical infrastructure and consist of labor power and skills plus the productive capacity of tools, machinery, facilities, lands, management practices, and knowledge. Relations of production, by their turn, are social relations the members of society establish among themselves in order to undertake social production. These relations are legally established as property relations: For instance, under the serfdom relation in

[8] As a matter of fact, the first writings of Marx and Engels on HM theory were presented in The German Ideology, dated from 1846. However, it was published only posthumously and for the first time in 1932.

feudalism, barons, who were the legal landlords, had legal rights to coerce the peasant serfs to work for them in their lands; also, under the capitalist relation in capitalism, burgeons or capitalists are the legal owners of the means of production and workers, yet legally free, are the non–owners of those means who have to work for the capitalists in order to survive.

HM theory was developed from the study of how previous forms of organizing society's economic base, each form corresponding to a particular mode of production, changed in time. The quintessence of HM theory is that whenever the economic base changes, or, more precisely, whenever a mode of production transits to another mode, the superstructure follows behind and changes also[9]. Marx listed a historical sequence of four modes of production: Asiatic, ancient, feudalism, and modern bourgeois (capitalism). In spite of controversies, HM theory is well accepted by many Marxists to explain the transition from feudalism to capitalism. However, it was stated by Marx as a universal law that can explain the transition from any mode of production to the next in the historical sequence. Thus, in principle, it might apply also to the case of capitalism and its transition to a new mode of production. Marx called this new mode 'communism', but in what follows we think of it, instead, as simply the next mode.

Each mode of production corresponds to a particular stage in the development of productive forces. Such a stage, by its turn, is suited to particular forms displayed by the relations of production. In other words, the stage of development of productive forces intertwines with the prevalent relations of production conditioning the particular features displayed by each mode of production. However, the productive forces are always developing, either by their own[10] or motivated by the prevalent relations of production. For instance, in feudalism, the relations between barons and serfs in the rural areas, and between masters and apprentices in the urban guilds promoted the technical advances of productive forces in a slow pace. In contrast, within capitalism, the relations between capitalists and workers and among capitalists themselves via competition provide strong incentives for the permanent improvement of productive forces. In such a way that it has no match as compared with previous modes of production.

Within HM theory, it is through the development of productive forces that important things happen. This development creates new possibilities of production that at some moment give rise to new relations of production. As long as these new relations show to be more productive than the old ones, they start to undermine the prevalent mode of production[11]. It happens because while old and new relations co-exist for some time they compete with each other. The new relations pressure by spreading across the economic system, while the old ones react by using the superstructure dimensions and other devices to refrain the advancement of the new[12]. At a certain stage, the tension between the two gets so high that a period of social revolution begins and unfolds toward establishing new property relations and a whole new superstructure. This completes the transition from the old to the new mode of production.

In sum, we might list the following stages of transition within HM theory:

---

[9] Notwithstanding, HM theory possess other essential features about which we develop next.

[10] By "their own", we mean accidentally or randomly.

[11] This particular point, say, the creation of new relations of production by the development of productive forces, is a little bypassed by Marx although he recognizes these relations as "forms of development of productive forces". A few lines ahead in the full *Preface* document, he reinforces the point by writing: "No social order is ever destroyed before all the productive forces for which it is sufficient have been developed, and *new superior relations of production never replace older ones* before the *material conditions for their existence* have matured within the framework of the old society." (Marx, 1859, 2017; italics are ours).

[12] For instance, the medieval guilds used to make pressures over the British parliament against the introduction of machines in factories. The famous Luddites protested at the point of physically destroying machines (Frey and Osborne, 2013).

- *Development of productive forces:* during the prevalence of a particular mode of production, the productive forces are always developing;
- *New relations of production:* the development of productive forces can happen more or less fast but at some moment gives birth to new relations of production;
- *Conflict development:* while the new relations of production co-exist with the old ones (typical of the prevalent mode of production), a conflict between them develops;
- *Social revolution:* as the conflict strain between new and old relations gets high enough, a period of social revolution starts in order to establish new property relations, transform the whole superstructure, and complete the installation of the new mode of production.

The HM theory sketched above highlights two central elements: relations of production and development of productive forces. In the remainder of this text, we'll discuss both in more detail, placing focus on the transition from capitalism to a new mode of production and the effects over employment. With regard to relations of production, a major question concerns *what* are the new relations of production under development now and *whether* they will be able to overthrow the capitalist ones. With regard to the development of productive forces, a central topic regards the *conditions* that new AI advances are producing so that at some point in the future a more advanced mode of production will be able to exist in place of capitalism. For reasons which will become clear ahead, we start by examining the latter issue. Therefore, in the next sections, we present some history of technological developments and jobs in capitalism, and then a discussion on the perspectives being produced by AI developments.

## 3. Technological Developments

For Marx, as he exposed in *Capital*, each capitalist firm faces a double front struggle. The first front is against other capitalist firms, in a race to dominate the provision of goods and services in the market. Such a front pressures the capitalist firm to adopt cost reduction strategies in order to gain extra−profits for some period of time. These cost reduction strategies usually take the form of incorporating technical progress in production. The second front is against the workers employed by the capitalist firm. In order to increase profits, the firm manages to reduce wage expenses as much as possible. It turns the capitalist firm additionally motivated to incorporate technical progress as an attempt to reduce wage expenses. So, both fronts push for the incorporation of technical progress and it is essentially in this way that the development of productive forces happens in capitalism. A major outcome of such a process is the replacement of human labor by machines in social production.

Within capitalism[13], the development of productive forces comprises, overwhelmingly, technical advances in machinery/equipment and improved management practices, including layout and logistics. Since its beginning, capitalism has promoted rapid

[13] In the remainder of this section, we make a brief recall of major technological revolutions which took place throughout the history of capitalism. Our purpose is just to put these revolutions in perspective. We focus on periods and corresponding features using a standard division followed by many authors and our description is similar to the one made by Rifkin (2014).

improvements of machinery and equipment. Particular periods featured technological revolutions, say, disruptive improvements that spanned externalities across the economic system. Researchers generally agree that the first period of technological revolution was the Industrial Revolution era, which endured from late eighteenth century to the first half of the 19th century. Nowadays, it is referred to as the First Industrial Revolution (IR1). It was marked by the generalized adoption of factories, the advent of the steam engine, and lately the widespread use of machines in industries. The Second Industrial Revolution (IR2) started near the turn of the 19th to the 20th century and lasted up to the First World War (WW1). It featured the widespread use of electricity, the telegraph, the internal combustion engine, and the mastery of the assembly line in factories. Such innovations allowed production in large scale for mass consumption. The Third Industrial Revolution (IR3) has developed after the Second World War (WW2). It has been marked by the increasing use of computer and electronic devices in production and services industries; in other words, a persistent and growing incorporation of information technology (IT) in social production.

The 20th century also displayed remarkable advances in management practices. Early, in the era of large-scale production for mass consumption, the first advances came up with the Taylorist–Fordist Revolution, which allowed new factory systems that improved assembly line manufacturing. It was a period of large-scale production but in standardized forms. Famously, Henry Ford once said that a customer could buy a car of any color, as long as it was black. After the WW2, the advances turn to a new direction with the advent of the Toyotist Revolution. New management and logistic practices were developed, turning production processes more flexible in order to allow production customized for client needs. On a step by step basis, these practices eventually became a new paradigm to production management. In the early nineties, technological advances brought about in almost all areas and, notably, in transportation equipment and telecommunications, triggered the Globalization era. Further developments in management practices, like outsourcing and offshoring, were introduced and have been used mainly by large firms to structure global production chains.

## 4. Effects on Jobs and Employment

The technological developments brought about since the beginning of capitalism have had important effects on the kinds of job opportunities and the composition of employment. According to Frey and Osborne (2013), during the IR1, the introduction and widespread of factories produced a job "deskilling" process. Previously, industrial production was mainly undertaken by artisans. These were relatively skilled workers whose set of manual skills was acquired through many years of training. As opposed to the artisan shop, the factory workplace featured a larger space and a different layout that favored an increased division, simplification, and specialization of labor tasks. These specialized tasks demanded little skills and a small degree of education from workers but enabled the production of the same output with fewer man–hours than the artisan shop. With this higher productivity of factories, the artisan shop almost disappeared. In a later stage of the IR1, the generalized introduction of machines pushed further the deskilling process by allowing the automation of many repetitive and routine tasks which were already performed by unskilled workers[14]. As a consequence, workers were relegated to execute ever simpler tasks which depended on ever simpler skills.

[14] Such a process of replacing unskilled workers by machines in industrial production was extensively analyzed by Marx (1867, 200?) in *Capital*. It was also considered by classical economists, such as Adam Smith, Ricardo, and Babbage. For details, see the study of Bruger and Gherke (2017).

Notwithstanding, as the demand for labor at that time was intensive because of a fast–growing industrial production, wages kept increasing. Frey and Osborne (2013) highlight that, in the end, the deskilling process favored the unskilled worker in detriment of the relatively skilled artisan.

The deskilling of labor produced by technological developments was typical of the early history of capitalism and more restricted to the 19th century. In a superb and awarded book, Goldin and Katz (2008) argue that things changed in the IR2 era and the 20th century, when technological developments went hand–in–hand with the demand for educated workers as a result of *capital–skill complementarity*. At the turn of the 19th to the 20th century, with the increasing presence of the US leading the world economy and progressively surpassing Great Britain, the advent of electricity and the improvements in factory systems allowed production in large scale, as we noted above. However, it had the effect of bringing about complex management problems to the factory workplace that fostered the demand for more skilled labor. Skilled, blue-collar workers were demanded to operate and maintain the machines, while highly educated, white-collar ones were demanded to manage the factories. Also, the expansion of office workplaces in urban centers and cities called the clerk worker to enter the scene. Many skilled and highly educated workers, like secretaries, cashiers, file clerks, bookkeepers, accountants, managers, lawyers, and engineers saw a boom in the demand for their services. In the US, as a result of earlier developments in education and particularly in the higher education system with the Morrill Acts of 1862 and 1890, the supply of clerk workers was able to respond fast to the increase in demand, at the point of making the wage differentials between clerks and production workers to narrow[15].

The IR3 era, which started after WW2, witnessed the rise of the digital computer and the growing incorporation of information technology (IT) in economic activities. In industrial production, IT was overwhelmingly important in the adoption of computer–controlled machinery/equipment and eventually of robots. The first robot was introduced in factory production by General Motors in 1961. The automation of job tasks usually performed by production workers was massive within this process, at the point of inducing President Lindon Johnson to create a commission in 1964 to study and recommend solutions to the problem (Author, 2015b). However, the incorporation of IT produced its most disruptive effects over job tasks in clerking activities. There have been mainly two kinds of effects: automation of clerking activities and enhanced computer–skill complementarity. In the 1970s and 1980s, computer development was fast, starting from large mainframe computers and then moving to desktop and notebook personal microcomputers. Stemming from computer's falling prices and escalating processing power, such a process made a large number of clerking activities to vanish or almost disappear: telephone operators, bank tellers, file clerks, and secretaries were the most affected. On another hand, the destruction of such activities gave way to system's analysts, computer programmers, and other computer–skilled workers

[15] The Morrill Land–Grant Act was a bill passed by the American Congress in 1862 that conceded generous pieces of land to American states. The purpose was to induce states to create colleges and universities in the fields of agriculture and mechanical sciences. This was the start of the modern complex of American state universities and at that time was important to boost the American higher education system. It was later extended by the Morrill Act of 1890, which conceded cash instead of land and targeted at the former Confederate States from the American Civil War of 1861–1865. The Acts are noteworthy because the American higher education system had been developing prior to the boom in labor demand for highly educated workers of the IR2 era. Also, the High School movement was in fast development at the end of the 19th century and achieved near universalization of access by the 1930s (Goldin & Katz, 2008 p.12). Thus, American developments in education, which were not matched by European countries at that time, were very important in allowing the supply of educated workers, both blue and white collars, to keep pace with the fast development of capitalism in the US at turn of the 19th to the 20th century.

that were able to operate office software like word processors, spreadsheet calculators, and database managers. Therefore, instead of simply finishing other clerking activities (like in the case of automation), this complementarity between computers and human skills gave birth to new kinds of jobs.

In addition, a sandglass effect over the composition of employment has developed, with the increase in the shares of unskilled and highly skilled workers and the compression in the share of skilled, blue-collar workers (Author, 2015 and 2015b). This sandglass effect is usually referred to in the literature as “job polarization”. As labor markets kept expanding in the last three decades of the 20th century, jointly with job polarization an increasing demand for highly educated workers accompanied the increasing use of IT. Author et al (2003) developed an interesting study on the causes of this empirically verified correlation between IT and demand for highly educated labor. They were concerned with the question: What computers do and what people do with computers that translate into demand for more human–skills in jobs? In the search for an answer, they worked out a simple model based on two criteria for classifying human labor tasks. The first criterion sets down that a task can be *manual* or *cognitive*, with the latter meaning a task that involves analytic problem–solving and/or complex human interaction. The second criterion sets down that a task can be *routine*, in the sense that there are explicit rules indicating how to perform the task, or *non–routine*, in the sense that little is known about the rules involved in undertaking the task[16].

They concluded that computers can substitute for workers with advantage in those tasks that are routine, either manual or cognitive. With regard to tasks that are non–routine, computers are *very limited* to substitute for human workers in the case of manual tasks but have *strong complementarities* in the case of cognitive tasks. These conclusions are illustrated in chart 1, which reproduces with some modifications table 1 of Author et al (2003). With computer prices falling precipitously from 1970 to 2002, they concluded that industries whose labor input was intensive in routine tasks content invested much in computer technology to *substitute for* workers, while industries intensive in non–routine task content invested also in computer technology but to *complement* human labor. Thus, the strong complementarity was likely the major factor behind the increased demand, empirically observed, for highly educated labor. Note that, along with the limited substitution of computers for human labor in non–routine manual tasks, their approach can also explain the phenomenon of job polarization.

Chart 1: Computer versus human skill in labor tasks

| *Type of Task* | *Routine* | *Non–routine* |
|---|---|---|
| *Manual* | Substantial substitution | Limited substitution or complementarity |
| *Cognitive** | Substantial substitution | Strong complementarities |

Source: Adapted from Author et al (2003), Table 1. * Regards problem–solving (or analytic) and complex communication (interactive) tasks.

[16] Author et al (2003) set down this definition of a non–routine task based on the Polanyi’s paradox, which states that “We know more than we can tell” (see also Author, 2015; and the next sections of this paper). Many labor tasks fall into this category, for instance: driving a car, cooking meals, and writing an article.

Author et al (2003) identified the presence or absence of explicit rules within tasks (what defines them as routine or non–routine) as the key component to understanding how computer technology had produced a skill–bias content in labor demand. The authors concluded that IT would keep substituting for skilled workers on manual and cognitive tasks that are routine because the availability of explicit rules allows to computer code these kinds of tasks. In the case of non–routine tasks, and they cited in particular as examples the activities of truck driving and handwriting cuff, IT would hardly substitute for human labor. Frey and Osborne (2013), just ten years later, observe that AI developments applied to IT were successful in creating a driverless car and in developing computer programs able to read handwriting cuff. Thus, the recent AI developments are allowing that IT invades the group of non–routine tasks[17]. In their study, Frey and Osborne (2013) also developed a statistical model to compute the chance that, in the next two decades, each occupation in a set with 702 occupations will be computerized in the US[18]. They obtained a figure of 47% for the share of job occupations that are to disappear. In order to understand why the recent IT developments threaten to substitute for highly educated labor *even in the case of non–routine tasks*, we have to take note of some particularities of advancing AI technology.

## 5. AI Developments

The present state of the art in AI research allows a growing but still limited replacement of human labor by intelligent machines. However, as the present pace of persistent improvements continues, AI experts are betting on just a few decades for a prevalent use of intelligent machines over human labor in social production. Although it may sound too unrealistic now, we shall remember that many things which are possible today were strongly unconceivable in the recent past. For instance, people living in mid–19th century thought it was a mad idea that a ship of the size of a merchant caravel could fly. Also, less than a hundred years ago, people could not even imagine that an equipment such as a personal computer could be connected to another equipment such as a cell phone, each one in a different continent, by means of a wireless telecommunication device called satellite. And that, through this connection, two persons could talk and exchange text messages, documents, and photographs in real time.

At the times those people lived, capitalism developed rapidly but technology did not evolve as fast as it evolves today. This fast pattern of technological improvements is largely a result that technology feeds technology. However, the present stage of technological advances has an additional and distinctive feature: AI. As a formal field of scientific inquiry, AI research started in the 1950s and has since worked to expand the frontiers of information technology by trying to replicate human intelligence in computers. Up to the 1990s, it had developed two major branches of research: expert systems and machine learning. Expert systems are based on programming rules in computers so that computers can solve problems based on these rules[19]. Along the way, expert systems showed to be limited to tackle large-scale problems because of the hard–programming efforts needed to input a large number of

[17] See also Ford (2015) for a number of examples of non–routine tasks that advancing AI technology is getting able to computerize.
[18] BBC News developed a website in which any person can use Frey and Osborne's model to compute the chances that a robot takes her job. The site address is http://www.bbc.com/news/technology-34066941.
[19] An example of an expert system is a wizard that assists users to install programs in computers.

rules into computers. By its turn, the branch of machine learning has shown fruitful and promising developments.

Early computers were highly powerful to perform complex, large-scale numerical computations and to organize, store, and query relational databases filled with structured data. However, they were very limited to perform pattern recognition tasks that are typically easy for humans, like identifying persons and objects in photos or recognizing handwriting cuff. Machine learning is based on replicating human-like processing of information in computers, so as to endow these machines with those typically human skills of pattern recognition. Up to the turn of the century, machine learning research had achieved remarkable progress thanks to its major technique, a device known as neural network (NN). A NN mimics the human brain intelligence in computers using mathematical models that represent information processing functions of the human neural system. A major property of NNs is that they enable computers to develop intelligence through a learning process, just as humans do, what gave the AI branch the name ‘machine learning’. This learning capability is put into action by training, a process in which a dataset is repeatedly presented to a NN. This process of training/learning is quite important for making the programming effort needed to develop a NN considerably smaller than in the case of an expert system. Instead of imputing rules into a computer, one just lets a NN recognize patterns in the data and learn by itself implicit rules. However, this data–dependency for training NNs put AI research nearly dormant for many years since the mid–1990s.

Until recently, data shortage along with limitations of memory size and computer speed had prevented more broadly, large-scale applications of NNs. However, things changed in response to a couple of trends that have developed since the beginning of the new century. The first trend has been what technology experts call Moore’s Law, the fact that computer capabilities increase exponentially by doubling every 18 to 24 months. Indeed, as a result of Moore’s Law, computers today are many times more powerful than they were at the turn of the century. Among many outcomes of such a process, a major one has been the development and widespread use of cloud computing, notably via the Internet. The so-called Internet Cloud is a large scale data processing environment provided by large corporations of the information services industry. The Internet Cloud is physically located in data centers scattered around the world which are equipped with collections of powerful computer servers. The advantages to users accruing from sharing computing resources on the Internet Cloud consist of reduced costs, storage availability, and increased processing speed. Such advantages have made people and organizations move in a massive fashion their domestic data processing to the Internet Cloud, turning it the largest data processing environment used in the Digital Age.

The second trend has been the development of a phenomenon called Big–Data. As computers’ capabilities expanded, a colossal amount of data has been recorded at the Internet Cloud and is permanently increasing. Part of these data consists of information on Internet users’ behaviors, such as transactions, searches, and web site accesses. Another part consists of data captured by sensors present in physical devices connected to the Internet of Things’ networks. Such large amount of data naturally carries with it many new business opportunities. However, the sheer big sizes of the new databases, of the order of zettabytes (trillion terabytes), along with their unstructured nature prevented the use of traditional tools for data management and analysis for some time.

Together, these two trends set the background for a big push to machine learning applications and a recovery of AI research. The plenty availability of data has provided developers with improved conditions for training NN based systems. Indeed, recent experimentations made by startup firms and corporations gave rise to disruptive technological applications of NNs. IBM achieved a breakthrough with IBM Watson, a supercomputer system developed in 2010 with two important features: The first is the capacity to search information in different datasets filled with structured or unstructured data[20]; the second is the ability to process natural language. IBM Watson was originally developed to compete at the famous CBS' TV Game *Jeopardy!*, a quiz show involving general knowledge in which only humans had participated in until then. At the contest, IBM Watson was installed in a large collection of computer servers. It was not connected to the Internet but stored a large dataset filled with structured and unstructured data, including the full text of Wikipedia[21]. IBM Watson could search fast this dataset to provide the answers, but its remarkable capacity to recognize natural language was crucial in the end. In *Jeopardy*!, the questions are presented to contestants in the form of answers for which they have to develop proper questions. Even with this inverted answer–question system, IBM Watson could communicate and respond accordingly, at the point of beating the other contestants.

A similar kind of application of sophisticated NN software was developed by DeepMind, a British startup firm created in 2010 (Silver *et al*, 2017). It was later acquired by Google in 2014 and turned into Google Deepmind (GD) division. A GD team created AlphaGo, a super-intelligent program developed to play the ancient Chinese board game Go[22]. In 2016, AlphaGo was able to win Lee–Sedol, a 9–dan (highest degree) Go player from South–Korea who had won 18 international titles. AlphaGo became the first computer system to beat a game–player champion using a trained, instead of a programmed, system[23]. Indeed, AlphaGo caught the attention of the AI community because of its ability to learn by itself a highly complex system. But, it showed yet another important feature: it was able to develop game strategies completely unknown to Lee–Sedol and many experienced Go players who watch the contest. Such experiment highlights a remarkable potential for machine learning techniques, notably NNs, as it points out not only to the possibility that AI devices can find solutions to complex problems, but also that they can do it in ways unattainable by humans. In addition, AlphaGo's developers claim they started an era of "General Purpose AI" because superintelligent programs such as AlpahGo are able to learn many different things (not only games) and thus have great potential to be applied in a wide range of society's complex problems.

Other large corporations of the information technology business, like Apple, Facebook, Microsoft, and Amazon, have come along with IBM and Google making large-

[20] Structured data consist of quantitative or qualitative data organized in tables, like the data records about employees in a capitalist firm. Unstructured data comprise any other kind of data that cannot, or are usually not, organized in such a table format. Examples of unstructured data may include books, journals, documents, metadata, health records, audio, video, analog data, images, files, and unstructured text such as the body of an e-mail message, Web page, or word-processor document.

[21] Wikipedia (2018).

[22] The Chinese Go played by AlphaGo consists of a 19x19 board game playable by two persons using homogeneous pieces colored white and black. It is a game more complex than Chess because the number of possible legal moves has as lower bound $2x10^{120}$ in the case of Chess and $2x10^{170}$ in the case of Go (Wikipedia, 2018). Therefore, Go is around $10^{50}$ times more complex than Chess.

[23] Previously, other supercomputers had already been able to beat game–player champions, like the famous case of IBM's Deep Blue supercomputer which bet the world's best Chess player Garry Kasparov in 1997. However, Deep Blue did not use a machine learning software, but a kind of an expert system which used "brute force" to calculate several steps ahead of the game before a particular move. The experiment with the AlphaGo program was different in two respects: First, the Go game has a level of complexity remarkably superior to Chess ($10^{50}$ times more); second, AlphaGo used a neural network to learn from scratch how to play the game Go and developed, with human help that provided selected training data, sufficient skills to beat a 9–dan level human champion (https://youtu.be/TnUYcTuZJpM). A recent press release from Google, as of December 2017, announced that a new version, AlphaGo–0, was able to develop from scratch and by playing only with itself sufficient skills to win a 100 times the previous version of AlphaGo.

scale investments in AI systems. What has motivated these large corporations to develop such systems is, ultimately, the purpose to gain competitive advantages in the information services industry. Indeed, they have put available those super intelligent systems on a commercial basis[24] to private firms and other institutions that are using them with many purposes, ranging from business planning and marketing strategies to support of scientific studies in many research areas.

What has also been crucial for making all of this possible are new tools of data analysis that benefit from the machine learning developments embodied in those systems. Labeled under the umbrella of Big–Data Analytics or Data Science, these new tools enable to extract meaningful knowledge from Big–Data and thereby provide highly improved support to decision making. As a major outcome, these novel data processing resources have opened up new perspectives for disruptive technological solutions to rise up in practically all problem domains of society in the next decades. Nonetheless, as a highly important side effect, such disruptive developments are also placing a serious menace to human jobs, even to those dependent on highly skilled workers. To be more precise, it menaces to end the complementarity between fixed capital and human labor that has prevailed since the end of the 19th century.

## 6. The End of Capital–Skill Complementarity

The study by Author et al (2003) that we mentioned earlier[25] calls our attention to peculiar skills of human workers. It regards those skills that enable them to perform *non-routine* job tasks, either manual or cognitive. Acquiring such skills involves the use of human cognitive abilities to develop knowledge. Many scholars see knowledge as featured with two basic dimensions: explicit and tacit. *Explicit knowledge* consists of rules that humans know regarding how to perform routine tasks and which can be easily transferred to other humans using verbal language, visual symbols, or other kind of code. *Tacit knowledge* is the source of Polanyi's paradox epitomized in the statement "we can know more than we can tell" (Polanyi, 1966). It consists of those things someone knows but which he/she cannot easily and directly transfer to others. Usually, tacit knowledge manifests itself when someone is acting or doing something. Examples are the knowledges of how to play a music instrument or how to ride a bicycle. Learning how to perform such activities cannot be fully undertaken without practice and experience.

Philosopher Michael Polanyi introduced the concept of tacit knowledge in mid–1950s and devoted a book to discuss its importance and implications, mostly within epistemology (Polanyi, 1966). Since then, tacit knowledge as a concept has spurred a large body of literature that includes controversial clashes among many interpreters (Yu, 2003). Eventually, it became widely accepted and used in a variety of areas, such as philosophy, cognitive psychology, organization theory, knowledge management, and AI, just to mention some. According to Polanyi (1966), tacit knowledge is more important than explicit knowledge. It comprises the most part of the knowledge a person possesses and grounds the use and development of explicit knowledge. As such, tacit knowledge is a powerful resource allowing humans to perform non–routine tasks.

In the field of knowledge management, it is accepted that tacit knowledge (at least part of it) can be articulated or turned into explicit (Nonaka and Takeuchi, 1995). Hence,

[24] For instance, at the site IBM Cloud (https://www.ibm.com/cloud/) IBM provides many IT services based on IBM Watson.
[25] See also the studies by Author (2015 and 2015b).

experts in this field recommend that capitalist firms convert as much as possible the tacit knowledge of their workers into explicit. The more from tacit to explicit that knowledge can be converted, more job tasks can be automated. But, the difficulty in automating non-routine tasks lies precisely on the tacit knowledge needed to perform them. Tacit knowledge cannot be easily codified, neither within a mechanical device (like a machine) nor within a computer program. In humans, it develops by practice and experience, but can remarkably improve by instruction and education. According to psychological studies, cognitive *innate* abilities and the use of language allows humans to develop intelligence in a far superior fashion than animals (Dennet, 1994). It is such *cognitive abilities* that allow humans to develop and improve tacit knowledge, creating thereby *cognitive skills*[26]. Such skills are germane to the human brain processing and storage of information, the mechanics of which is understood (yet only partly) thanks to developments in psychology, neuroscience, and the AI branch of machine learning. Hereafter, we call as *typically human skills* (THS) those cognitive skills that allow humans to perform non–routine tasks.

The capacity to develop the THS is determinant for the existence of human capital (HC). In economics and the social sciences, the notion of HC was described with different but similar connotations since Adam Smith. Nowadays, it is widely understood as all the knowledge, skills, and health that humans develop from practice, training, and education, and that ultimately increase their productivity (Goldwin, 2016). Note that it is quite close to our notion of THS added with health conditions. For the sake of this paper, we might think that HC can be described symbolically by such a sum:

HC = THS + Health.

Therefore, THS can be seen as a component of HC, and in fact it is the central one. There seems to be a wide consensus among economists and social scientists that HC was, and still is, the major force behind technology, innovation, and economic growth. More than fixed capital and labor, HC is seen as responsible for the remarkable increase in total factor productivity (TFP) displayed by capitalist economies during the 20th century (Goldin and Katz, 2008). In the 1950s, economist Abramovitz (1956) discovered that the increase of US output along 1870 to 1950 had been faster than the increases in fixed capital and labor. Latter, as many economists attributed this phenomenon to HC, the concept gained prominence in the scholarly literature and the society at large. Whilst economists have always recognized that HC resulted from cognitive abilities and skills, it seems they have ascribed these a secondary importance[27]. We argue that the core component of HC is comprised by the THS because these are responsible to endow HC with its capacity to be developed and improved. Hereafter, we'll prefer using the expression THS whenever we need to refer to HC.

Hitherto, the THS have provided human workers with great advantages over machines and software to perform non–routine tasks. As we mentioned before, by the early 20th century, despite persistent efforts of capitalist firms to substitute fixed capital for human labor (automation), the deskilling process crashed into the complexities of producing goods and services to mass consumption and the expansion of cities (urbanization). To deal with such

[26] We shall make clear our distinction between the terms cognitive abilities and cognitive skills. Cognitive abilities regard innate capacity of humans to learn and thereby develop knowledge, either explicit or tacit. Cognitive skills are cognitive abilities developed. It regards the capacity to manage explicit knowledge but is particularly important for using the tacit knowledge to perform non–routine tasks.

[27] Labor economists have also considered non–cognitive skills. According to Heckman and Rubinstein (2001), these comprise "… motivation, tenacity, trustworthiness, and perseverance …". We don't deny the importance of such non–cognitive skills, but they seem as secondary to the cognitive ones when the complexities we refer to here are taken into account.

challenges, fixed capital and technology were important but far from sufficient. What made the difference was not exactly HC, the supply of which was rapidly increasing because of the American pioneering efforts in education. Our present day knowledge about the human brain's processing of information allows that we risk a different interpretation: The fact is that the THS were the most sophisticated "technology" available at the time. Yet a natural and biological one, the THS empowered humans to perform non–routine tasks involved in machine operation and maintenance, factory management, and a wide range of clerical work activities. In other words, the THS fitted uniquely to the non–routine job tasks demanded by the IR2 and were pivotal to allowing capitalism's development from that stage on (Goldwin and Katz, 2008).

What is so important with regard to THS? Under a modern perspective, THS mean human capabilities of sensing, gathering, and processing information from a variety of sources to deal with uncertainties present in ill-structured problems of decision making. Central to these capabilities is human intelligence, defined as human capacity to learn and adapt to uncertain and changing environments within which non–routine tasks are performed. As such, the THS were a *marvel* that the scientific knowledge available at that time was very far from automating. This is why fixed capital technology and skilled labor were but complementary, not substitute, to each other. We might even say that it was not technology that permitted skilled human workers to be more productive, but the THS which allowed technology to fully unleash its potentials. At least, we might say they are equally important. Moreover, the education system in the US, yet largely massified at the early 20th century, permitted to unleash such THS from most of the American people. Thereby, it adds as another major factor, if not the most important, behind the coming to prominence of American capitalism. Of course, the plenty availability of natural resources (scattered across a large geographical area) to be exploited, technological revolutions, and money capital to finance private and public investments were crucial. Notwithstanding, it was the 'technology' embodied in the THS that manage all of this to turn into a reality the impressive expansion of capitalism in US and, with some delay, in Europe and the rest of the world as well[28].

Now, let's turn to a point we made in the beginning of this paper, that capitalism has always been able to create new jobs even after undergoing the technological revolutions of the past. Many technology experts (e.g. Ford, 2015) believe that, with the AI revolution, things will be different now. Why? The answer is that advancing AI technology is threatening to substitute for the THS[29]. Precisely the kind of human skills that hitherto has allowed skilled and highly skilled human labor to be complementary to fixed capital and technology. Even more, the kind of human skills that, since the early days of capitalism in $18^{th}$ century Britain, the scientific knowledge has been unable to automate. However, advancing mobile robotics jointly with AI developments are now turning capital and skilled labor from complementary into substitute factors of production. While human labor has always been object of replacement, the THS have also always made the difference in social production. But now, not only AI threatens to make those THS obsolete at the point of making them not typically human anymore: It also threatens to bring back a deskilling process, similar to the one that prevailed in $18^{th}$ and $19^{th}$ centuries.

[28] Europe and the rest of the world lagged far behind the US in education development during the whole IR2. For details, see Goldwin (2016).

[29] We might say HC instead of THS. However, we prefer using the latter because it goes right to the point as it is the major feature of HC that is to be replaced.

## 7. Neural Networks and Tacit Knowledge

We devote this section to provide a more refined understanding of how a NN functions and why it is so powerful. The reader already familiar with such a topic can skip to the next section. Among AI methods, the NN technique is responsible for the most disruptive technological innovations in recent years. It had a central role, as we saw earlier, in the great breakthroughs of IBM Watson and AlphaGo. In mid–1980s, it triggered the AI branch of machine learning and became the latter's major subfield of research. Besides, NN is also the AI subfield within which the *Deep Learning*[30] (DL) technology is currently developing. Roughly speaking, the NN model, also called *artificial* NN, was initially designed to tackle pattern recognition problems but evolved to be also useful as a knowledge discovery method. Because of DL, it has surpassed human intelligence in many particular applications. Indeed, modern, large scale NNs with built–in DL technology have gone to the point of replicating abstract–like reasoning in applications involving knowledge discovery.

Notwithstanding this remarkable performance, a NN is often blamed for being a black–box because the knowledge it discovers remains embedded within its inner structure. In other words, its knowledge is incomprehensible for humans. It learns things and acquires knowledge but it remains tacit within the NN and is hard to be made explicit. Hence, a NN fails to share its knowledge with us. Since it is often the case that logical and explicit understandings are relevant for users in many contexts where AI technology is applied (e.g., military and hospitals), the last two decades have witnessed the development of a large body of literature on the topic of *rule extraction* from NNs. Rule extraction is a set of methods and approaches to uncover a NN's hidden knowledge in order to help explain how the latter comes to final decisions. The purpose of these methods and approaches is to allow users to better understand the knowledge embedded in a NN (Hailesilassie, 2016). Even so, the high degree of tacitness embedded in NNs (see below) make these to still suffer from the black box stigma.

A NN develops knowledge in a way that is quite similar to the human learning process. As we saw in the last section, such a process involves the use of human cognitive abilities to develop knowledge, and we also said the latter comprises tacit and explicit dimensions. Hence, because of tacit knowledge it is natural to some degree that a NN is a black–box. In this section, we argue that the knowledge a NN incorporates after being trained and tested embbeds the same kind of tacit knowledge that humans develop. In order to make the point, we outline in the sequel a rough explanation of how a NN functions. The description is quite simple, with minimal technicalities and formalisms, and useful to help us understand how a NN functions and develops tacit knowledge.

Basically, a NN mimics the way that humans identify patterns in objects of real life. For instance, we see a crowd of people in front of us and suddenly recognize the face of a friend. Early computers were simply unable to make such recognition. However, the advent of the neuron model, introduced by McChulloch and Pitts (1943), and latter developments that brought about the NN technique allowed empowering computers with such a skill. How? The answer is that a NN is a mathematical model that replicates information processing functions of the human brain. And it is easy to program such a model in a computer. In the description that follows, we'll use a problem of personal recognition as a running example.

[30] Deep Learning is a set of approaches and strategies to build NNs that has dramatically improved the performance of the latter in complex problems of pattern recognition. The present wave of fast developments in DL started in 2006 and for a history and survey, see Schmidhuber (2014 and 2015). DL techniques place focus on incorporating hidden layers of neurons within a NN architecture. We discuss these topics ahead in this section.

An interesting feature of a NN computer program (hereafter simply NN) is that, like a newborn human child, it initially knows nothing. When it runs for the first time, the NN cannot recognize faces but has the ability to learn how to make it. Usually, we can teach the NN by training it, a process called *supervised learning*. We present to the NN a number of pictures (examples) of persons we'd like it to recognize their faces. Such set of pictures is called *training data* and the process of training is labeled *learning stage*. After trained, the NN is then tested with a different set of pictures in order to check if it can *generalize* well. This new set of pictures is called *testing data* and the process of testing is labeled *testing stage*. The NN is regarded *fit* if it passes the testing stage, say, if it displays a high performance of recognizing the persons' faces in the testing data.

Let's make it simpler and consider we want to train a NN to recognize just a single person, say John, in a set of photo pictures (our training data). We assume some pictures in the set contain John and the remaining ones don't. We then proceed as follows:

1. We present the NN one picture and ask it to guess: 'Yes' (picture contains John) or 'No' (picture does not contain John);
2. We then inform the NN whether it has guessed wrongly. We do it by telling the NN it has committed an error (the NN then makes some adjustment and changes itself);
3. We repeat steps 1–2 many times until all pictures in the training set had been presented;
4. After we have presented all pictures, we shuffle them and present to the NN again by repeating steps 1–3 (i.e., now with the sequence of pictures in a different order);

Each time we follow steps 1–3 (present all pictures) is called an *epoch* of training. The NN undergoes many epochs of training in the learning stage. In each epoch, the same pictures are presented to the NN although in different orders because of the shuffling we make in step 4. We shuffle to avoid that the neural model recognizes a spurious serial pattern in the sequence of photos. The only pattern we are interested in regards the visual features of John.

Note that it is like playing with a child whose father is John by showing her many pictures of the family, some with her father, some without. After each picture, we ask the child if her father is present and tell her whenever she commits an error. By repeatedly playing with the child, we help her to develop the skill of recognizing her father in almost any image. Thus, the process by which a NN develops knowledge and skills is very similar to the way humans learn. To check whether the NN had learnt the visual features of John, we present it another set of pictures: again, some with John, some without. If it displays a good performance in this testing set, we regard it is ok and ready to perform the task of identifying whether John is present in any new image.

While the NN is learning the features of John, it is improving on its own errors, like humans do. Recall that, when we learn from our own errors, our brain "adjusts" our perceptions of the object whose features and patterns we are examining. For instance, when we fail to recognize a person in a picture and someone tells us "you didn't see him/her", our brain receives a stimulus to *adjust* its perception of the person's features. It is by repeatedly making such brain adjustments that we finally learn how to recognize the person. Essentially, what the NN does is to replicate such adjustment process.

Before we can grasp how a NN model functions (say, how it replicates the human brain adjustments while learning), we need first to understand the nodes of the net, say, the *neurons*. We shall recall here the visual notion that every net is a set of nodes linked to each other, like the knots in a fishing net. Inspired in a human neural system, an artificial NN is a collection of interconnected artificial neurons. Before we talk about the complete NN model, let's focus on the node, say, let's understand what is a *neuron model*. It is a very simple mathematical model and figure 1 presents a visual illustration of its structure. The three circles on the left, labeled $X_1$, $X_2$, and $X_3$, together represent the *input*, say, the information we give (a picture). The big center circle is the *nucleus* of the neuron where the input values ($X_1$, $X_2$, and $X_3$) are processed to produce an output. The circle on the right, labelled $Y$, is the *output* of the neuron (an answer: 'Yes' or 'No'). The *arrows* on the left indicate that the input information is being furnished to the nucleus and the single arrow on the right indicates that the nucleus produced the output $Y$ (the same as the neuron output).

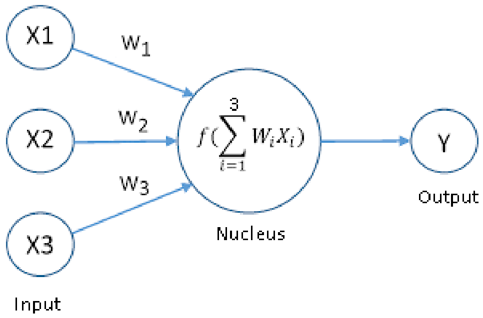



Figure 1. A Model of a Single Neuron.

Now, pay attention to the elements $w_1$, $w_2$, and $w_3$ that are near each input arrow on the left. These comprise the most important component of the neuron model and are called *synaptic weights*. The word 'synaptic' comes from a real, human body neuron and refers to the links to other neurons. The word 'weight' comes from the simple fact that $w_1$, $w_2$, and $w_3$ are the weights of a *weighted sum*, given by the expression $\Sigma_{i=1}^{3} w_i X_i$ inside the nucleus. Indeed, note that the nucleus takes the input values $X_1$, $X_2$, and $X_3$, and simply computes a weighted sum of them, using for this the synaptic weights $w_1$, $w_2$, and $w_3$. The function $f$(,), which we can also see inside the nucleus, then makes another simple operation: it converts the weighted sum into numbers 1 (meaning 'Yes, John is in the picture') or 0 (meaning, "No, John is not in the picture"). These two numbers, 1 and 0, are the only possible values of output $Y$.

A little bit of thinking allows us to realize how simple such a model is. As shown in figure 1, its structure comprises three basic elements:

- a set of *input values*;
- a *weighted sum*, and
- a simple *function* that produces an *output*.

The input values $X_1$, $X_2$, and $X_3$ are numerical values, because the computer only understands numbers and codes. It is through these numbers/codes that a computer 'sees' the image in a photo picture. In a real application with the neuron model, what is difficult is to converting an image (a picture with or without John) into just three numerical values $X_1$, $X_2$, and $X_3$. Indeed, a much higher number of input values is usually needed to represent the image of only a single photo picture. For such an image to be precisely (with high resolution) inputted to a computer, we'd need a thousand of input values $X_1$, $X_2$, … $X_{1000}$, or maybe more. For instance, the image of a picture that we see on our computer screen is formed by an array of thousands of very small squares, named 'pixels'. Each pixel has a different color, and to each pixel color our computer assigns internally a number code. In other words, the array of colored pixels forming with high quality (resolution) the image that we see on our computer screen is nothing but a set of number codes for our computer equipment.

We used only three values $X_1$, $X_2$, and $X_3$ to represent the input (a picture) in figure 1 just to keep the example simple. Hereafter, we might think we used a thousand input values $X_1$, $X_2$, … $X_{1000}$, instead of just three like in figure 1. Note that even using a thousand input values, it does not change the simplicity of our neural model. Its structure still comprise:

- a set of input values;
- a weighted sum, and
- a simple function that produces an output.

We stress this simplicity of a neuron model because an artificial NN is just a collection of interconnected neuron models. Like the knots (or nodes) of a fishing net. It is based on such a simplicity that a sophisticated, large scale NN with millions of neurons is built and used in complex, real applications. But let's go back still for some moments to our neural model of figure 1.

There is another feature of the neuron model that we must understand. We just explained above what the neuron model does when presented to a *single* picture: It processes a 1 or a 0 as output, corresponding to answers 'Yes' or 'No'. But, for the neuron model to learn the features of John so as to recognize whether he is present or not in almost any picture, there is something else. It regards what happens to the neuron model during the learning stage. Remember that in this stage we have to present several pictures (the whole training data) to the neuron model, not just one. What happens is that, after every picture we present and after every answer it gives, the neuron model *updates* and *improves* its knowledge about John. In other words, it learns. But how? What is the mechanism that allows it to learn? We will answer in two steps.

First, let us introduce the concept of *error* made by the neuron model. It is pretty intuitive. Such an error is simply the difference between the *correct answer* (usually called *desired answer*) and the neuron model's *output answer*. Assume now that our training data has 200 pictures. Let's call by the letter $n$ the $n$–th picture presented in a sequence, so that $n$ can be any of the numbers 1,2,…,200. Given this, we can define the error associated to the $n$–th picture as:

$$e(n) = D(n) - Y(n)$$

where $D(n)$ is the correct answer and $Y(n)$ the neuron model's (output) answer. We assume $D(n)$ can be 1 (picture contains John) or 0 (picture does not contain John). This equation is simply saying:

- *While the neural model examines the n–th picture, the error it commits is the difference between the correct answer and its own answer.*

Note that in our example this error can take only three possible values: −1, 0, and 1. If the neuron model's answer is correct, it has committed *no error* and then $e(n) = 0$. If the answer is wrong, the neuron model has committed an error and in this case:

- $e(n) = -1$: correct answer is 'No' and output answer is 'Yes' ($D(n) = 0$; $Y(n) = 1$), or
- $e(n) = 1$: correct answer is 'Yes' and output answer is 'No' ($D(n) = 1$;$Y(n) = 0$).

Truly, these particular values of the error are not important four our understanding. What is important is that the neuron model uses its *own errors* to update itself and improve its skill to recognize John in pictures. In other words, the neural model uses its errors to learn. But how does it use its own errors?

This leads us to the second step. The neuron model 'learns with its own errors' by simply *adjusting the values of the synaptic weights* whenever it commits an error. This adjusting process is carried out by changing the weight values with the simple rule:

$$w_j(n+1) = w_j(n) + \Delta_j$$

Where subscript $j$ represents the $j$–th weight value. Note that $j$ can take the values 1,2,…,1000 if we were using a thousand input values to represent each picture. This formula says that, after the the $n$–th picture is presented to the neural model and it is verified that it has committed an error, the $j$–th weight value $w_j(n)$ is changed to a new value $w_j(n+1)$. The amount of change is given by the variable $\Delta_j$ . We don't need to know much about this variable, only that it is a function of the error $e(n)$:

- If no error is commited, $\Delta_j$=0;
- otherwise, $\Delta_j \neq 0$.

Therefore, according to the formula above, the neuron model can make a change (adjustment) in each weight value at every time it commits an error. Suppose we present the $n$–th picture to the neuron model and it checks whether John is or not in the picture. Then, the following situations can occur:

- *Neuron model hits*: then no error was committed and no change is made in the synaptic weights, say, $w_j(n+1) = w_j(n)$;
- *Neuron model fails*: then an error was committed and a change is made in every weight value because in this case $w_j(n+1) \neq w_j(n)$.

In the learning stage, we train the neuron model by presenting it the sequence of 200 pictures in many epochs of training. After examining every picture in every epoch, the neuron model makes an adjustment inside itself (in its weight values) whenever it commits an error. Through this process, it produces a short improvement in its knowledge about the visual features of John after being exposed to each picture. The process of making such adjustments is the way that the neuron model learns or discovers the visual features of John. All of this mean simply that at a particular $n$–th instant, the weight values $w_1(n)$, $w_2(n)$,…,$w_{1000}(n)$ embody the $n$–th *state of knowledge* of the neural model with regard to John's visual features. After we present the $n$–th picture and tell the neural model it has committed an error, it changes its *state of knowledge* to a new set of weight values $w_1(n+1)$, $w_2(n+1)$,…,$w_{1000}(n+1)$, according to the formula above. The state of knowledge remains the same (weight values do not change) if the neuron model does not commit an error.

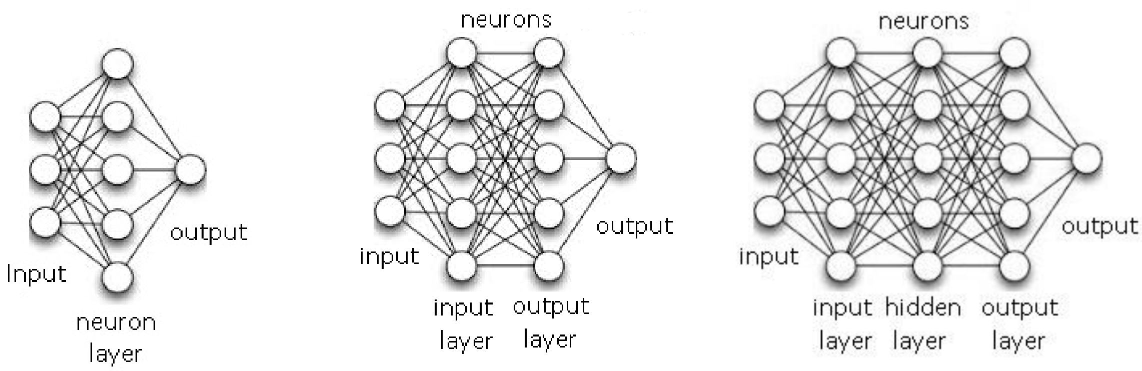
Note that, after the neuron model completes the training process (after all 200 pictures were presented to it in many epochs), it remains that same simple structure that comprises: a set of inputs, a weighted sum, and a simple function that produces an output. But now it is a *skilled* neuron model, with much knowledge about the visual appearance of John. Unfortunately, this knowledge is not explicit. It means it cannot be transferred to us. Mathematically, the neuron model is simply a method to update the weights of a weighted sum. The only thing that changes inside the neuron model, as it learns, is the set of weight values $w_1$, $w_2$, … $w_{1000}$.

In other words, the knowledge incorporated by the neuron model is embedded in a set of 1000 weight values. This set of weight numbers has no meaning by its own. We cannot interpret it, neither with regard to John's visual characteristics, nor with regard to anything else. But it is this set of 1000 weight values that provides the neuron model with the knowledge necessary to recognize the presence or absence of John in pictures. We can say that such knowledge is like human tacit, for three basic reasons:

- the neuron model replicates the human brain processing of information;
- its acquired knowledge cannot be communicated or transferred to us, and
- its acquired knowledge was developed following the same process used by humans when they learn.

The neuron model is the simplest NN we can conceive of because it features only one neuron. If we want to apply the NN technique to more complex problems, for instance to recognize the faces of many persons, not only of John's, we need to enrich the simple structure of the (single) neuron model. We have to add more neuron models (hereafter simply neurons), and there are two basic strategies to do that. First, we can simply add more neurons reading the same inputs $X_1$, $X_2$, ..., $X_{1000}$. An example is portrayed in figure 2.(*a*) In this case, we have a simple NN with a single *layer* of neurons.

Second, we can add more neurons distributed in different layers. Figures 2.(*b*), and 2.(*c*) display two cases of this second strategy. In figure 2.(*b*), we have a first layer of neurons, which reads the set of inputs and produce a set of neurons' outputs, and a second layer of neurons, which reads the neurons' outputs of the first layer and produce another set of neurons' outputs. The first layer of neurons is called the *input layer*, and the second the *output layer*. The overall output of the NN model is the output circle. The structure of this NN model can be further enriched by putting another layer of neurons between the input and output layers. This new layer is called *hidden layer*, because it does not have contact neither with the NN's inputs nor with the NN's outputs. This is portrayed in figure 2.(*c*).




(*a*) (*b*) (*c*)

Figure 2. Types of NN Models and Architectures.

The basic setup of a NN comprises the choices regarding the number of neurons, the number of layers, and the distribution of neurons among layers. It is called the NN's *architecture*. The art of determining the architecture is key for a NN model to succeed in particular applications. Modern NNs used in complex applications such as natural language processing and speech recognition come to feature millions of neurons and many hidden layers. These NNs also make use of a variety of additional settings related to the types of:

activation and error minimization functions, learning process (supervised, semi−supervised, or unsupervised), and weight updating rule, among many other settings. The NN's research subfield of DL is concerned with the inner structures of NNs' architectures, say, the use and set up of hidden layers. Recent progress have shown that properly designed architectures with hidden layers can allow the NN to replicate abstract level reasoning and retrieve deep knowledge from complex systems.

In a NN model, the tacit knowledge comprises a large set of weights. Each link (trace) that we see in figures 2.a), 2.b), and 2.c) corresponds to a weight value. Note we have the weights associated to the links:

- from the input values to the input layer,
- from the input layer to the next layer (a hidden or the output layer),
- from the ouput layer to the output value.

With millions of neurons in different layers, a NN may have millions or billions of weights to update. Therefore, the potential for a NN model to develop a large amount of tacit knowledge is remarkable, and depends on two things: computer capacity (storage and speed) and data availability. As we explained earlier in section 5, two things whose supply has been developing fast thanks to the rapid advances in IT.

## 8. Conditions for a New Mode of Production

In this section, we discuss some conditions that AI developments are bringing for the installation of a new mode of production. In order to discuss this and other topics regarding the effects of AI over capitalism and employment, in this section we use a strategy of considering a limiting case scenario. We explicitly assume that, somewhere in a very long–run future[31], advancing AI technology *fully displaces human skills in jobs*. We are aware that such assumption is extreme, but it provides two advantages: First, it is well known that exploring an idealized situation grounded in basic and explicit assumptions can provide relevant insights about a more complex reality. Second, we find it unimportant whether such extreme scenario will be possible or not. It suffices that society approaches it to some degree for its major consequences to realize.

Advancing AI has been pervasive in many domain specific fields and, as in the case of IT, is evolving to be a general purpose technology. It has been applied directly to innovations in IT proper or indirectly as a tool to technological developments in other research fields. These include leading domain specific fields such as mobile robotics, precision mechanics, molecular biology, and nanotechnology. Furthermore, technological advances in these and other fields have fostered innovations in medicine, health services, pharmaceuticals, other chemical products, and industrial processes. All such developments have persistently morphing the way social production is undertaken. In a very long–run future, as AI becomes widely pervasive in society, it is possible that human labor becomes obsolete for social

[31] It is important to observe that the effects we consider to be produced over society by advancing AI technology depend on how far in the future goes our scope of analysis. In a near future, there is little sense to assume AI fully displaces humans in jobs. On the contrary, it is likely that advancing AI technology comes to create many waves of new kinds of jobs. But, we are working here under the assumption that any new jobs still dependent on human skills to be created in the short− and long–runs by advancing AI technology will eventually vanish in a very long–run future.

production. If it comes true, social production will be undertaken under conditions completely different than today's. Our point in this section is that such conditions will be strong enough to allow the existence of a new mode production which is different from capitalism. Mostly, these conditions will be the result of particular features already displayed by advancing AI technology. In the following paragraphs, we explore some of these AI features and the implications of those conditions.

We start by depicting the most immediate product of AI: an intelligent machine or algorithm (IMA). Basically, an IMA can be classified into one of the following cases:

a) physical robot equipped with and controlled by an intelligent computer;
b) physical robot remotely controlled by an intelligent computer; and
c) intelligent virtual or algorithmic robot.

An IMA can operate individually or in a network connection with other IMAs. We list in the sequel some features of IMAs that are noteworthy:

- *Productivity*: As in the case of traditional (non–intelligent) MAs, IMAs can perform tasks much faster than humans and on a continuous basis, say, without the need to stop for resting, feeding, sleeping, socially interacting, etc. Essentially, this is what makes IMAs more productive than humans;
- *Non–compensation*: IMAs do not demand *wages* or any kind of *compensation*. They work for free for their owners. It means that IMAs operate, or perform tasks, for only two simple reasons: they have goals and are plugged into electrical outlets (the goals are usually set up in their programming by the IMAs' owners[32], which can be a private or a state organization);
- *Technical efficiency*: Human workers only produce efficiently when strongly motivated, for instance by earning a good salary or by getting in risk to be fired. They achieve their maximum productivity rate, or do their best, only when they receive strong incentives, whether positive or negative. They follow the rules of the capitalist game. As opposed to human workers, IMAs do not depend on particular motivations or incentives. They always work under their maximum productivity rates (which are higher than humans'). They follow only the physical laws. As technical devices, IMAs' only *motivation* to function, or to 'work', is their energy inlay. In this sense, they will stop working only if they were unplugged from the electrical outlets. Also, IMAs operate with the same stamina whatever the goals and tasks posed to them: whether the goal is maximum profit or minimum price, or whether the task is recognizing a face or extracting meaning from a text. Goals and tasks can vary in complexity, but IMAs always operate with the same devotion. They don't prefer a task to another. They simply 'work' and always do their best. No more, no less;
- *Impartiality*: IMAs don't suffer from human bias. A human person can make different judgments about a particular person or issue in different circumstances. For instance, empirical evidence exists that court judges display different degrees of tolerance in their sentences before and after taking meals (Frey and Osborne, 2013). In principle, IMAs are not affected by such biases unless they were trained to act differently in different circumstances;
- *Replicability*: a highly sophisticated IMA, for instance a highly trained NN with significant knowledge like IBM Watson, is a software program installed in a

[32] AI experts believe that in the future IMAs will be able to reprogram themselves and thereby will develop their own goals, what has bringing much anxiety to those concerned with the issue of singularity.

computer. Therefore, it can be easily copied (replicated) to other computers[33]. A corollary of this property is that a NN's tacit knowledge can be easily conveyed between IMAs. This is in hard contrast with humans, who have to socially interact for some time under a master–apprentice relationship for tacit knowledge to be transferred.

Although these five features have many implications for depicting IMAs, we'll focus only on the first three. The productivity, non-compensation, and technical efficiency features have a particular importance for the issues we are discussing. These features are shared by traditional MAs, but they loom in importance and perspective under our assumption that, in the future, IMAs will reach the capacity of replacing human labor even in firm management tasks. In such a setting, society would be able to have firms supplying goods and services *efficiently without profits motivation*.

The last statement needs more explanation. A capitalist firm, as we know it today, is a production unit of the economy devoted to providing some good or service. However, it pursues that goal primarily by generating profits. Say, a production unit in capitalism has as primary goal generating profits for the firm's owners (human capitalists) and only as a secondary goal the provision of goods and services to society. If there were no possibility of profits, the good or the service would not be produced or supplied by the firm. Motivated by the perspective of obtaining profits, the human owners of the capitalist firm decide that it purchases inputs and hire human workers. The basic motivation of the latter to work for the capitalist firm is that they need wages to access social output. In connection with that, after being hired workers develop additional motivations resulting from the fear to be fired and get unemployed for an uncertain period of time. Workers' motivations do not end here, as many of them also expect to improve their living conditions. For so, they endeavor to upgrade within the firm to better positions that pay higher wages. These human motivations, either of capitalist owners or of workers, are behind the relations of production in capitalism.

Such (humanly motivated) capitalist mode of providing goods and services to society is known, and often hailed, to be the most "efficient" in history. Indeed, it has shown in practice to be efficient with regard to quantity and quality of the goods and services produced, minimum production costs incurred, and incentives provided for technological developments[34]. But, it is a result of how the capitalist system manages human motivations and all its efficiency is fundamentally dependent on these motivations. Because such capitalist efficiency is produced by the *human motivations* to earn profits and work, we call it here as *behavioral efficiency*.

Now, let us consider IMAs operating the capitalist firm in place of a team of human workers. When IMAs perform tasks, they are optimizing an objective function. Their goal is to attain the extremum (maximum or minimum) of the objective function. According to the non–compensation feature, this goal is set up by the IMA owner and can be different things.

[33] Of course, the presently existing superintelligent IMAs, like IBM Watson and AlphaGo, are computer systems which are private properties protected by patents. Their replicability in practice has to respect the interests of the corporations and enterprises which own them under the umbrella of the law. We list replicability among the major features of IMAs because it implies a range of new possibilities of production. The development of new (maybe free) IMAs by cooperative schemes, like for instance the free software movement, are not forbidden if patent rules were respected.

[34] In standard economic theory, capitalism is also hailed to possess a particular kind of efficiency known as Pareto efficiency. It is achieved by free markets under perfect competition in all economic industries and corresponds to an equilibrium state in which one individual can improve only by worsening the situation of other individuals. It is a particular and quite theoretically concept of efficiency that is difficult to be measured and observed in practice. However, free markets based capitalism can be ascribed to displaying those more practical efficiency concepts discussed here.

For instance, IMAs may alike pursue profit generation and maximization for the owners of the production unity to which they belong. Because of the productive and technical efficiency features, a team of IMAs will be able, in a very long–run future, to do it more efficiently than a team of human workers. However, an important fact is that IMAs can also be programmed or trained to have *different goals*, say, different than pursuing profits or compensations. For instance, IMAs can be programmed/trained to optimize the provision of the good or the service so as to fulfill a set of society´s needs, including the needs for environmental protection and technological development. And again, because of the productivity, non–compensation, and technical efficiency features, IMAs can do it efficiently and in this case much better than a team of human workers who depend on those typically human motivations to do their best. Because of the technical efficiency feature, IMAs would not work harder motivated for getting a better position in the firm (targeting a higher wage) or because of a fear to be fired. They would simply work and do their best better than humans. And trying to replicate those human motivations in IMAs, through a different programming/training, would make no sense.

Let us try yet a simpler example. Suppose we have a firm somewhere in the future operated only by IMAs and dedicated to providing a single good in the market. There is no human worker in this firm. All the tasks associated with production and firm management are performed by IMAs (thus, robots and virtual algorithms). Now, suppose IMAs were programmed/trained to manage the firm with the goal to provide this good at the least possible price to consumers, under a restriction of zero–profit, or of just a small amount of profit only to preserve financial balance. Note we are not saying "maximum profit" but "minimum price", so that it is a nonprofit oriented firm. Because of productivity, non-compensation, and technical efficiency features, IMAs will be able to manage the firm *efficiently* (given its minimum price goal) and the consequence would be lower price and more plenty provision of the good to consumers than if the IMAs had operated the firm with the goal of profits maximization. The efficiency of IMAs' work within this firm is not altered by having a different goal. They work under *technical efficiency*, not under *behavioral efficiency* as do human workers (or economic agents).

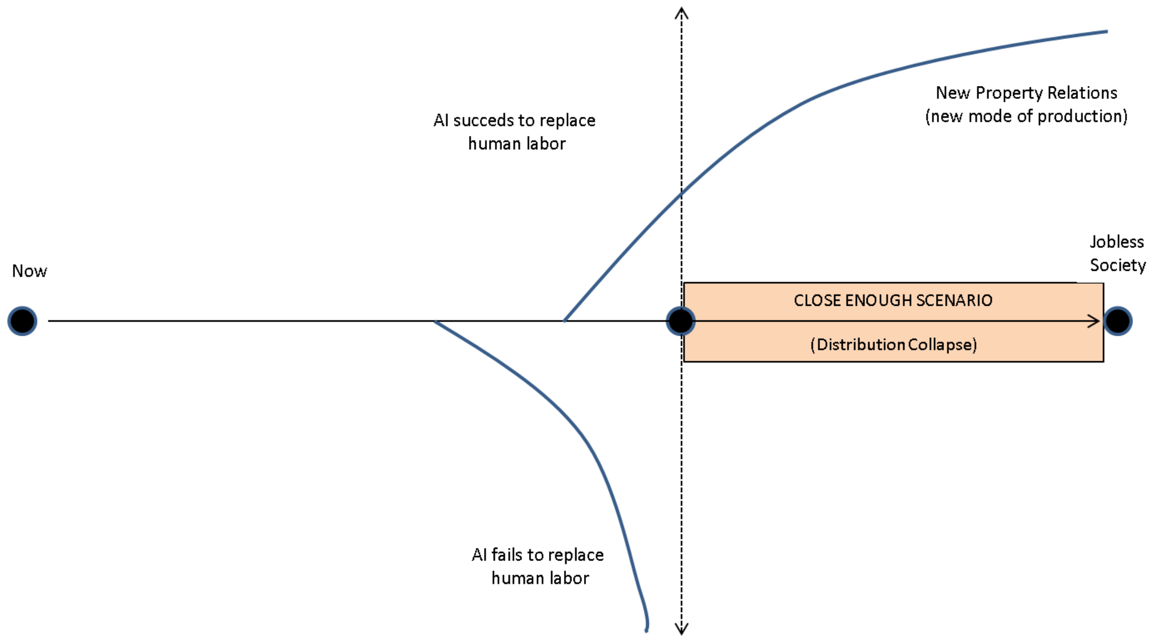
These considerations lead us to the following conclusion. The productivity, non-compensation, and technical efficiency features plus the assumption that IMAs will replace all human workers in the future comprise conditions for a new, post–capitalist mode of production to exist. *A mode that may be efficient without depending on agents searching for profits*. Advancing AI technology is leading us to such a limiting case future. A future in which human society as a whole will not need to work and the 'IMA workers' (or better, IMA based systems) will be able to provide with abundance the needs of all society's members. It would be a heavenly world, except for the fact that property relations in society would have to be different than today's. For non-employable human members of society be able to consume and live, they will need somehow to have rights to access IMA based social output. If not, the heavenly future turns into a distribution collapse.

Whether it is possible or not that advancing AI technology will take us at a limiting case, jobless future is not relevant. What matters is that by getting close enough to it, we'll be already exposed to its hazardous consequences. Can advancing AI technology take us so far? We do not have a well-defined probability distribution for the degree of closeness to such a future. Nevertheless, up to this point our reflections lead us to the conclusion that *close enough* is not an unlikely event. By a *close enough scenario* (CES), we mean a situation in which the consequences would be essentially the same as those of the pure limiting case scenario. If IMAs are not to replace 100% of human workers in jobs, it suffices to be a high figure like, for instance, 47% (as predicted by Frey and Osborne, 2013, for the next two decades) or a bit more. It would also create a distribution collapse.

We see just two possibilities in which AI might not be a threat. The first is a scenario in which advancing AI technology eventually *fails* to replace much of the THS. In this case, it would simply be a frustrated promise. The second is a scenario in which advancing AI and other technologies can physically improve the THS by transforming the human body. Since a few years ago, technology expert Ray Kurzweil has expressed his belief that nanotechnology will be able in the future to allow the human brain connect directly to computers and the Internet[35] (Miles, 2015). In September 2017, biomedical engineers at the Wits University, from Johannesburg, South Africa, connected for the first time a human brain to the Internet using a (non–nano) technology based on small chips (Medical Express, 2017). This second scenario thus involves the body–improvement of THS by technology, turning human workers each time more productive. It can be further enhanced by other technological advances in other areas such as medicine and molecular biology.

The fact is that both scenarios ('frustrated promise' and 'body–improvement') would keep human labor still complementary to fixed capital. As yet, we can only assume that AI threatens, not that AI will, replace in full the THS in the future. While technology remains dependent on the THS, human labor will not be replaced in full. The second scenario, body–improvement of THS, seem more effective to saving human jobs from disappearing and thereby keeping the capitalist system, once more, able to create new human jobs.

There seem to be two trends here: The first, that AI technology *succeeds* to replace human labor and thereby ends the complementarity of fixed capital with human labor (eventually creating a jobless society); the second, that AI *fails* to replace human labor (scenarios 'frustrated promise' or 'body–improvement') and thereby keeps the latter still complementary with fixed capital. We illustrate both trends in figure 3. Exactly which trend will prevail in the long run, it is early to say. But we shall bear in mind that, as long as the capitalist system keep seeing the replacement of human labor and its THS as an advantageous business strategy, the trend towards a jobless future is to prevail.




[35] The sci-fi movie industry has recently explored such possibility in the film Anon, by director Andrew Niccol. The script, however, is more audacious and features a world in which people's brains connect directly to each other and the Internet without the need of computers, mobiles or tablets.

Figure 3. Scenarios on the run towards a jobless society. If AI is leading us towards the jobless society and we cross the middle point, we will be in the CES (pink region). It is a dangerous area because contemporary society undergoes a distribution collapse. In case AI *succeeds* to replace human labor, we enter the CES but to avoid the distribution collapse, new property relations warranting social output is accessible to the whole society need be implemented (blue curve growing). Other way to avoid the distribution collapse is by not entering the CES, which would be possible should AI *fails* to replace human labor (blue curve dropping; in this case, capitalism stands still).

## 9. Relations of Production

In the previous section, we concluded that advancing AI technology is (likely) leading us to a future scenario that is *close enough* to a fully jobless society. It is an important conclusion, but it is not clear how we'll arrive at such a CES. For a proper assessment, we have to look also into how the transition from now to such a future can happen. We learned that property relations in society will have to be different[36] to avoid the distribution collapse or, better, to protect society's members against the consequences of a generalized technological unemployment. Now, the basic questions are: What can we do as a society to arrive safe at such a CES? What are the present opportunities to help us fight this challenge? It is here that futurist efforts can be important and HM theory particularly useful. We consider in this section some possible paths toward the *close enough* scenario. A first one is society making conscious efforts to work out solutions and policy responses. The other paths are associated to the natural unfoldment of HM theory.

A first path is that society develops conscious, strategic movements. Some important efforts are on the go. For instance, the fact that many experts have alerted the potential AI threats to employment has sensitized institutions, political leaderships, and elite groups. A measure of society's awareness of the new challenges is the many discussions undertaken recently in international forums (it means the issue has gone further than the media and academic institutions). For instance, the WEF Annual Meetings held in the last three years delivered many discussion panels and experts' interviews on subjects such as advancing AI, technological unemployment, fourth industrial revolution, and universal basic income (WEF 2017*a*,*b*,*c*,*d*). Some consensus has emerged on two major courses of actions: retraining workers to the new, upcoming environment and redesigning safety nets, including the adoption of universal basic income. We have short room here to dive deep into such issues, but it matters to say that such conscious efforts are undoubtedly the best avenue for a safe transition towards the CES. It is society itself trying to work out solutions in a negotiated and peaceful way.

A second path links to one of HM theory's predicted developments. As we described in section 3, HM theory's last stage of transition between modes of production is a period of social revolution. A very undesirable outcome, it can happen even before new, more advanced relations of production develop. It can result from a generalized distress of the population with a lack of responses by part of society's elites in the event that technological unemployment grows out of control. We must not disregard it as a possible scenario. It can happen, for instance, due to a lack of agreement among elite groups over policy responses. A major risk that society runs was pointed out by technology expert Erik Brynjolfsson (WEF,

[36] If they have to be different, it means we'll have a new mode of production that is different than nowadays' capitalism. HM theory predicts that the drive towards a new mode of production after capitalism is new relations of production.

2017d): AI-based technology develops faster than society's institutions can change to adapt to the new environment.

In case social stability is restored after a revolution, we see two major possibilities here. The first is the attempt to implement socialism again. The second, a moderate development, is a return to social democracy and a heavily enhanced welfare state system, without displacing capitalism. Anyway, inasmuch as both developments resulted from a (possibly violent) social revolution, it reinforces the importance of the first strategic path based on conscious and negotiated efforts to work out a peaceful transition.

Another possible path associated with HM theory regards the other stages of transition before the last one. Applying HM theory to the present context, the development of productive forces accruing from advancing AI technology is to enter into conflict with the capitalist relations of production. As we saw in section 3, it means that new relations of production were to spring up from the new production possibilities. The challenge here is to properly identify these new relations of production. For so, we have to answer questions such as: What are the relations of production under development now that are different from the capitalist relations? Which ones are the most likely, in the sense of more advanced, to displace the capitalist relations? There are no easy answers to these questions. What seems clear is that the issues most demanding of research efforts stay here.

For instance, the rapid changes are every time introducing new possibilities of production, mostly for private businesses but also for collaborative schemes. The new possibilities for private businesses may not imply new relations of production, just different business models. The new possibilities for collaborative schemes may be new relations of production, but not advanced enough to overthrow the capitalist relations. In fact, a clue to identify new relations of production is to look upon the alternative schemes developing at the margins of the capitalist economy. For instance, there are workers' owned cooperatives producing particular goods for the market. They are profit–oriented but the work is developed in a cooperative fashion based on common interests of all participants. Other examples are the Open Software movement and the collaborative, nonprofit based scheme of the Wikipedia group. These alternative schemes have been able to compete and even displace some profit–oriented firms in their industries. However, it is difficult to imagine them spreading throughout the economy displacing most of profit-oriented, capitalist firms.

Nevertheless, it is a fact that the rapid development of new technologies and production possibilities are at every moment creating new business models and opportunities for cooperative schemes. It is a process of small transformations within the capitalist system that can lead to a redesign of it in the future. Bestselling author Paul Mason (2015) calls our attention to this rapidly changing environment, arguing that in the last decade we have seen the emergence of:

> Parallel currencies, time banks, cooperatives and self-managed spaces, … [n]ew forms of ownership, new forms of lending, new legal contracts: a whole business subculture … which the media has dubbed the 'sharing economy'. Buzzterms such as the 'commons' and 'peer-production' are thrown around, but few have bothered to ask what this means for capitalism itself." (Mason, 2015)

For capitalism itself, these small transformations may be the embryo of new and more advanced relations of production. By more advanced, we mean more productive and more suitable to environment preservation and human health. And a nice strategy for society might be to incentive the development of such new relations. At the moment, the collaborative schemes seem the most likely of bearing the embryo of these new relations.

At this point, it is opportune to cite also Jeremy Rifkin (1995), the famous author of *The End of Jobs*, published in the mid–1990s. There, he already predicted the importance of AI as one of the sources of technological unemployment in the future. In a more recent book, *The Zero Marginal Cost Society*, Rifkin (2014) explores further the impressive advances in technology and develops an interesting narrative on what he calls the *collaborative commons* (CC). Roughly, CC is the concept that IT-based developments are leading people to work each time in a more collaborative way, notably via the Internet, as if they were at a large virtual public square. If such a collaborative system is going to replace capitalist relations, it remains to be seen. The CC is still far from being a dominant way of undertaking social production, but as Rifkin (2014) makes clear, it is not a marginal way anymore. Conversely, as long as the advancing AI technology falls somehow in public domain, it can provide a boost to such a collaborative system. At the moment, it is mainly affecting the collaborative schemes based on IT and the Internet, but it can also reach some tangible industries in the foreseeable future.

In sum, it is still early to foresee precisely the main directions of the developments even using HM theory. As we pointed out above, research efforts are needed. But, in loose terms, there seems to be a new contradiction under development. This is the one between behavioral and technical efficiencies. It has been a natural development of capitalism since its beginning, stemming from the latter’s propensity to replace human workers by machines in social production. However, it also represents a bottleneck for solving the problems we are discussing. Technical efficiency winning the race against behavioral efficiency means the distribution collapse, as we discussed earlier. But, it also means the opportunity of an abundant world in which social production can be undertaken efficiently without depending on profit oriented activities. In other words, an economic system not dependent on profit oriented activities may become in the future both a necessity and an opportunity to solve the distribution collapse.

Care must be taken here because ‘not dependent on’ does not mean ‘without’ profit–oriented activities. It just means that the *dominant form* of undertaking social production is not comprised by profit oriented activities anymore. Nothing prevents that the latter might co–exist with non–profit oriented activities. Such a view may be advantageous for the negotiations of a peaceful transition to a new mode of production because it preserves some economic freedom. It may be an acceptable solution for the conflicting parties involved (e.g, elite groups and workers) and can avoid social distress. At the moment, the most visible opportunity for this seems social democracy with a reinforcement of its safety nets (for instance, with universal basic income). It might be progressively implemented depending on how the contradiction between technical and behavioral efficiencies evolves.

## 9. Final Comments

In this paper, we developed some reflections on the challenges that advancing AI technology is placing to capitalism and employment. The excitement with AI goes far beyond such specific topics, reaching also the perspectives of singularity and what humans would do in a jobless society, among many others. For sure, these comprise important dimensions of the new challenges posed to us and naturally deserve much attention. Rather, we decided to focus on capitalism and employment because we see in the new developments good opportunities for solving the capitalism’s problems of high and increasing inequality and of limited effectiveness to end poverty.

Our analysis placed focus on the extreme case of a fully jobless society, which we labeled as the limiting case scenario. We also analyzed a similar, but also more realistic, case of a CES. Both share as a major feature the distribution collapse. Of course, we cannot determine in precise terms which is the best alternative path that society should follow to arrive save at a CES. The most we can say is that society efforts to anticipate the future will be crucial for a safe transition to any future scenario. Futurism seems unavoidable: the challenge is how to make it good. It would ground a good development in the negotiations towards a safe transition. In the case of the CES, HM theory gives us some clues about the ongoing developments. We shall pay attention to the new relations of production surging from the fast technological developments. Also, what happens in the area of collaborative schemes of production is of obvious importance because these already embrace alternative relations of production.

We'd like to finish by stressing Brynjolfsson's remark: AI technology advances faster than society's institutions can change. It is possibly among the major challenges brought about by the present context. In order to properly address all the questions and issues involved, research and other efforts are needed. In addition to the ongoing discussions coming about in the media, social nets, universities, and international forums, society must use all the available resources to develop good policy responses, including advancing AI technology.

**References**


Author, D. H. Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, v. 29, n. 3, 3–30. 2015a..

_________ Polanyi's Paradox and the shape of employment growth. Federal Reserve Bank of St. Louis: *Economic Policy Proceedings*, Reevaluating Labor Market Dynamics. 129–177. 2015b.

Author, D. H., Levy, F. and Murname, R. J. The skill content of recent technological change. *The Quarterly Journal of Economics*. November 2003.

Brugger, F. and Gherke, C. Skilling and deskilling technological change in classical economic theory and its empirical evidence. Working Paper 2017–02, Karl–Franzens–University Graz, Faculty of Social and Economic Sciences. 2017.

Cellan–Jones, R. Google Deepmind: AI becomes more alien. BBC News. Retrieved 1/24/2018 from http://www.bbc.com/news/technology-41668701.

Dennett, D. C. The Role of Language in Intelligence. in *What is Intelligence?*, The Darwin College Lectures, ed. Jean Khalfa, Cambridge, Cambridge Univ. Press. Retrieved 6/19/2018 from https://ase.tufts.edu/cogstud/dennett/papers/rolelang.htm. 1994.

Ford, M. *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books: New York, 2016.

Future of Life Institute. Benefits & risks of artificial intelligence. Retrieved 5/5/2017 on https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/. 2107.

Frey, C. B. and Osborne, M. A. The future of employment: how susceptible are jobs to computerization? Oxford Martin School. Manuscript. Retrieved 11/9/2017. (https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf), 2013.

Goldwin, C. Human Capital. In Diebolt, C, and Haupert, M. (ed). *Handbook of Cliometrics*. Berlin: Springer. 55–86. 2016.

Hailesilassie, T. Rule Extraction algorithm for deep neural networks: a review. *International. Journal of Computer Science and Information Security*, v. 14, no. 7, 2016.

Heckman, J. and Rubinstein, Y. The importance of noncognitive skills: lessons from the GED Testing Program. *The American Economic Review*, v. 91, no. 2, 145–149. 2001.

Kohs, G. Alphago. Documentary Film. Netflix. 2017.

Marx, K. *Critique of the Gotha Programme*. Marxists Internet Archive. Retrieved 6/8/2017 from https://www.marxists.org/archive/marx/works/1875/gotha/. 1875.

________. *The Contribution to the Critique of Political Economy*. Marxists Internet Archive. Retrieved 6/8/2017. from https://www.marxists.org/archive/marx/works/1875/gotha/. 1859.

Mason, P. *Postcapitalism: A Guide to Our Future*. New York: Allen Lane, 2015.

Medical Express. Biomedical engineers connecting a human brain to the internet in real time. Retrieved 5/29/2018 from https://medicalxpress.com/news/2017-09-biomedical-human-brain-internet-real.html. 2017.

Miles, K. Ray Kurzweil: In the 2030s, nanobots in our brains will make us 'Godlike'. Hungtinton Post. Retrieved 3/3/2018 from https://www.huffingtonpost.com/entry/ray-kurzweil-nanobots-brain-godlike_us_560555a0e4b0af3706dbe1e2. 2018.

Nonaka, I. and Takeuchi, H. (1995) *The Knowledge-Creating Company*, New York: Oxford University Press

Polanyi, M. *The Tacit Dimension*. University of Chicago Press: Chicago. 1966.

Rifkin, J. *The End of Jobs*. New York: Putnam's Sons. 1995.

Rifkin. J. *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*. New York: Palgrave McMillan. 2015.

Sæbø, S. Polyani and the neural networks. Metacognition. Retrieved 5/2/2018 from http://blogg.nmbu.no/solvesabo/2015/11/polyani-and-the-neural-networks/. 2015.

Schmidhuber, J. Deep Learning in Neural Networks: An Overview. Technical Report IDSIA-03-14 / arXiv:1404.7828 v4 [cs.NE]. 2014.

Schmidhuber, J. Deep Learning. *Scholarpedia*, 10(11):32832. Retrieved 5/6/2018 from http://www.scholarpedia.org/article/Deep_Learning#References. 2015.

Silver, D., J. Schrittwieser, K. Simonyan*, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, and T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, D. Hassabis. Mastering the game of Go without human knowledge.. Retrieved 5/1/2018 https://deepmind.com/documents/119/agz_unformatted_nature.pdf. 2017.

Ticktin, H. The Permanent crisis, decline and transition of capitalism. Critique: Journal of Socialist Theory, 45(3). 359-386. 2017.

World Economic Forum. A basic income for all: dream or delusion? Panel Session with speakers: Michael Sandel, Amitabh Kant, Neelie Kroes, Tamzin Booth, Guy Standing. Available at https://www.weforum.org/events/world-economic-forum-annual-meeting-2017/programme. Davos: 2017a.

World Economic Forum. An insight, an idea with Guy Standing. Panel Session. Speakers: Guy Standing, Zanny Minton Beddoes. Available at https://www.weforum.org/events/world-economic-forum-annual-meeting-2017/programme. Davos: 2017b.

World Economic Forum. Artificial Intelligence. Panel Session, with speakers: Ron Gutman, Robert F. Smith, Ginni Rometty, Satya Nadella, Joichi Ito. Available at https://www.weforum.org/events/world-economic-forum-annual-meeting-2017/programme. Davos: 2017c.

World Economic Forum. Jobs and the Fourth Industrial Revolution. Issue Briefing. Speakers: Erik Brynjolfsson, Suzanne Fortier, Saadia Zahidi Available at https://www.weforum.org/events/world-economic-forum-annual-meeting-2017/programme. Davos: 2017d.

Yu, Z. Tacit knowledge/knowing and the problem of articulation. *Tradition and Discovery* 30, 1, 11–22. 2003.