Policy Brief 2018-XX -
DRAFT

Artificial Intelligence concerns for workers

Capture of workers' experience, lack of accountability and conservatism embedded in software - DRAFT

Artificial Intelligence improves the efficiency and reliability of industrial processes. It can support the market position of European companies and thus sustain high-quality employment in a globally competitive world. It raises however a number of concerns for European workers in industry, on: (1) the volume of employment and the qualification of tasks remaining for humans; (2) the rules to access industrial data, which can lead to digital monopolies; (3) the inherent conservatism that algorithms based exclusively on past experience entail; (4) the unexplainable nature of decisions or recommendations made by machine-learning systems; (5) the potential abuse by employers of Artificial Intelligence's capacity to predict the health condition of a worker, or to supervise him/her automatically and permanently; (6) the loss of control on self-learning systems after delivery by the producer; and (7) the instability of a system that can use its own output as teaching material. For each of these concerns, some suggestions for policy, regulation or social dialogue are given.

Scope: "weak" Artificial Intelligence

The scope of this reflection is "weak" Artificial Intelligence (AI), which is the type of AI that works and is operational in 2018. This form of AI performs strictly limited tasks, based on "machine learning": computers extract the information embedded in large amounts of unstructured data and develop a capacity to take decisions / make recommendations on cases not yet seen, based on the experience from the past gathered in the teaching data. The underlying software technique is called "neural networks", because it mimics the structure of the brain.

It does not cover the more futuristic prospects of "general Artificial Intelligence", where a single software would be able to engage in a great variety of tasks.

General economic and social effects on employment

Artificial Intelligence performs and will perform tasks that humans currently do. It will perform them often better, at lower cost and with greater reliability than humans. As such, this technology is yet another one **increasing** the **productivity of human labour**, just like many other technologies in history. One important difference is that the tasks that Artificial Intelligence can perform are those relying on human **experience**, and are thus often very **qualified**, e.g. of salespersons, workers driving complex machinery, medical doctors or maintenance workers diagnosing failure or illness, lawyers.

Estimates vary regarding the fraction of human tasks that could be replaced with AI. The same uncertainty reigns regarding the duration necessary for AI applications to penetrate and dominate the market. Whatever the exact figure, the impact will most probably be considerable.

The improvements brought by AI in speed, quality and reliability of professional processes mean that they will be implemented, and legitimately so, in order to make production more **efficient** in the usage of all resources. E.g. the waste rate in a steel rolling mill with a skilled workforce (or which would be driven by an AI system) is around 3%, while it reaches levels in the 25% with an unskilled

work-force. In that sense, the investment efforts announced by the Commission (consolidated public-private investment of EUR 20bn over 2018-20, and EUR 20bn/year beyond) are welcome to maintain the competitive position of European companies in global markets.

The traditional trade union answer to this nature of transformation lies in (1) **anticipation of change**, and strategic skills planning, in order to act before the restructuring takes place; (2) **re-** and **up-skilling** of workers. These answers remain fully valid.

Regarding re- and up-skilling, it is possible that the **quantitative** amount of necessary training be widely above the current budgets available for training of workers. Current budgets cope with a few days of training per year, for ca. 5% of workers (source Eurostat Participation rate in education and training by occupation [trng_aes_104]). The training needed to have workers to change professions entirely, due to their current professional skills being replaced by AI, is probably in the range of **several weeks or months**, for a larger fraction of the workforce per year. This may require a deep reform of the **financing** of workers' training.

An additional area of trade union reflection relates to **working time**, which is currently the purpose of a dedicated document by industriAll Europe.

Access to data: risk of digital monopolies

The development of an Artificial Intelligence system based on "machine learning" relies on the availability of **teaching data**. Without such teaching data, the machine cannot learn, and thus cannot be implemented. This leads to the issue of **access rights** to personal or industrial **data**.

In the current state of the art, collecting large amounts of data requires no inventiveness, and almost no investment, because of the very low cost of sensors, and of data transmission and storage. There is thus no legal nor moral basis for defining any form of "ownership" over such machine-collected data (be it on private persons, on workers or on objects / machines). A private capture of machine-generated data in a professional environment would be particularly damaging, because this data embeds the **professional experience** of workers, so that the data monopolist would *de facto* capture this experience. On the other hand, machine-collected data can find many socially and economically beneficial usages with different players, such as improving the process (in the operating firm), the maintenance procedure (in the maintenance service firm) or the machine itself (with the manufacturer / designer of the machine).

Suggestion for policy / regulation: Considering the collective advantage of sharing broadly the access to such data, and the risks of unjustified rents associated with monopolistic access, there is a rationale for a regime of **non-exclusive, mandatory licensing of machine-collected data**. See our Policy Brief "Sharing the value added of industrial Big Data fairly".

Conservatism embedded in software

Artificial Intelligence relies on the general idea that decisions impacting the future should be based on the past experience embedded in the teaching data, i.e. on the implicit assumption the **future** will be **identical to the past**. This is the very definition of **conservatism**. It leaves **no space** for **change** nor for **innovation**. AI thus risks reproducing the *status quo* forever – including any **discrimination bias** present in our societies, and thus in teaching data.

Suggestion for policy / regulation: Research on means to introduce innovation, experimenting, change and creativity in the operations of machine learning systems.

Neural networks are currently un-explainable

In the current state of science, systems on "machine learning" can neither **explain** nor **justify**

their decisions/recommendations. Contrary to explicit algorithms which can be followed by a human (provided the source code and the supportive data are public), and where all steps having led to the decision / the recommendation are explicit, neural networks are a complete "black box". No scientist is able today to track back, from the teaching data and the learning algorithm, what led to a decision / recommendation.

- This is problematic in general, because it **weakens** further the capacity of humans to **influence decisions**, when "supported" by an AI system, along an argument akin to "nobody was fired for choosing IBM" in the 1970s: a human taking a decision contrary to the recommendation of an AI system will make mistakes, and be sanctioned for having done so, whereas s/he will not for having followed the recommendation of the machine. This leads, statistically, to a situation where *de facto* almost all decisions are taken by the machine.

- This is even more problematic in situations involving the **management of workers**, on decisions impacting their professional development (e.g. promotion, dismissal, training). "Employee assessment" software can now predict the professional development potential of a worker, and make recommendations for the management of his/her career. The manager receiving such a recommendation would not be in a position to explain / justify it, other than the “argument from authority”: there is no other justification given than an opaque reliance on experience. This could deprive the worker from any possibility to discuss, present arguments to support his/her case and obtain redress. This deprival of a human interaction, and of a fair judgement, is very problematic for workers.

- This feature of being unexplainable is also problematic when considering **liability** and **improvement paths** in case of accident, failure or accident – for autonomous cars, production machines, and even more for airplanes or nuclear power plants (i.e. safety-critical artefacts). Finding the cause of the accident, failure or accident is important to determine who must pay for the damage. It is also important in order to improve the system and make sure that the accident, failure or accident of a system does not happen again. If the command and control system is based on AI, and is thus unexplainable, then liability cannot be determined, and no improvement is possible.

<u>Suggestions for policy / regulation:</u>

- One way to alleviate this problem would be to **minimise** the "unexplained" fraction of the modelling software. Many phenomena have been studied by science. They can be the purpose of **explicit modelling**, involving known equations and explicit calculations using known parameters, or parameters that can be explicitly estimated using standard statistical tools (from the simple mean value to more sophisticated Kalman filtering). Thereby, instead of having one single "black box" modelling the whole system, the behaviour of which would need to be anticipated by the AI software, the idea would be to model explicitly all that can be modelled, leaving to AI, and to the unexplainable "black box", only a small fraction of the modelling. This would have the additional advantage of requiring much less teaching data.

- **Mandate** that any machine learning software that takes decisions regarding **humans** and specifically workers (e.g. regarding health or Human Resources management), or that is embedded in a **safety-critical** system (e.g. rail equipment, civil aeronautics, nuclear power), be **explainable** – and prohibit its use as long as it is not the case.

Anticipating the health condition of workers can be misused

AI software can anticipate the **health condition** of a worker, and specifically the appearance of **chronic diseases**. This can be used positively, in order to engage in **preventive measures**. It can also be used in a more malevolent way by the employer: by knowing in advance that the person would develop a chronic disease (or have a high probability of doing so), the employer could dismiss the person beforehand, and thus evade its responsibility.

Suggestions for policy / regulation: **Prohibit** the usage by employers of machine learning systems anticipating the health status of workers.

Machine-learning systems can contribute to the abusive, permanent automated supervision of workers

Traditionally, the supervision of workers by management was technically and economically restricted by the difficulty of having a person looking permanently at the work performed by another to detect non-compliance with prescriptions (regarding speed, quality or safety). Even with cameras, it was difficult for a single person to supervise many, so that this supervision remained costly. As a result, workers had a form of *de facto* space of autonomy for organising their tasks.

With Artificial Intelligence systems, it becomes technically and economically possible to supervise all workers, permanently, and to detect all occasions of non-compliance with prescriptions, in real time. This has the potential to significantly reduce the space of autonomy that workers used to have – to the detriment of the quality of their life at work.

Suggestion for policy / regulation / social dialogue: Define the acceptable means to detect, store and process circumstances of non-compliance with prescriptions.

Systems based on Artificial Intelligence continue evolving after having been delivered to the customer – Loss of control

Machine learning does not stop upon delivery of the product to the customer. It continues, based on the experience and teaching data accumulated while being used by the customer. This leads to a situation where the original manufacturer and the customer / user have **lost control** on the behaviour of the machine. This raises a significant question regarding the **liability** in case of accident, because no one had any means to fully anticipate the behaviour of the machine.

Suggestions for policy / regulation: Define a solid liability regime for autonomous systems, even when their behaviour keeps evolving after purchase. One way forward could be to re-use and adapt the legal regime of animals.

Machine-learning systems using their own output as teaching data are unstable

When humans use the output from machine-learning systems, they produce data. If this data is indiscriminately re-used as teaching data for a further cycle of machine learning, this leads to an unstable amplification of any error or bias in the initial machine-learning system. The most obvious example of such instability is provided by the Google Translate service. As more and more persons use this system to translate, and to publish their translated texts on the Web, the Google Translate service considers these texts as

legitimate sources of teaching to modify its translation machine, and thereby deteriorates its quality, as the proportion of genuine, human-based translation in its teaching database diminishes.

Suggestion for policy / regulation: Mark the output of machine-learning system with an identifier signalling that this data should not be re-used as teaching data for the same system.

DRAFT