Viewpoint: Human-in-the-loop Artificial Intelligence

Fabio Massimo Zanzotto
University of Rome Tor Vergata
Viale del Politecnico, 1
00133, Roma ITALY

Abstract

Little by little, newspapers are revealing the bright future that Artificial Intelligence (AI) is building. Intelligent machines will help everywhere. However, this bright future may have a possible dark side: a dramatic job market contraction before its unpredictable transformation. Hence, in a near future, large numbers of job seekers may need financial support while catching up with these novel unpredictable jobs. This possible job market crisis has an antidote inside. In fact, the rise of AI is sustained by the biggest knowledge theft of the recent years. Many learning AI machines are extracting knowledge from unaware skilled or unskilled workers by analyzing their interactions. By passionately doing their jobs, many of these workers are shooting themselves in the feet.

In this paper, we propose Human-in-the-loop Artificial Intelligence (HitAI) as a fairer paradigm for AI systems. Recognizing that any AI system has humans in the loop, HitAI will reward these aware and unaware knowledge producers with a different scheme: decisions of AI systems generating revenues will repay the legitimate owners of the knowledge used for taking those decisions. As modern Merry Men, HitAI researchers should fight for a fairer Robin Hood Artificial Intelligence that gives back what it steals.

1. Introduction

We are on the edge of a wonderful revolution: Artificial Intelligence (AI) is breathing life into helpful machines, which will relieve us of our need to perform mundane activities. Self-driving cars (Lipson & Kurman, 2016; Lutin, Kornhauser, & Masce, 2013; Litman, 2014) are taking their first steps in our urban environment and their younger brothers, that is, assisted driving cars (Revathi & Dhillipala, 2012; Trajkovic, Colmenarez, Gutta, & Trovato, 2004; Gray, Ali, Gao, Hedrick, & Borrelli, 2012), are already a commercial reality. Robots are vacuum cleaning and mopping the floors of our houses (Ulrich, Mondada, & Nicoud, 1997; Taylor, Parker, Lau, Blair, Heninger, Ng, DiBernardo, Witman, & Stout, 2005; Huffman & Miner, 2008). Chatbots (Weizenbaum, 1966; Wallace, 2009) have conquered our new window-on-the-world – our smartphones – and, from there, they help with everyday tasks such as managing our agenda, answering our factoid questions or being our learning companions (Kerly, Hall, & Bull, 2007; Beccaceci, Fallucchi, Giannone, Spagnoulo, & Zanzotto, 2009). In medicine, computers can already help in formulating diagnoses (Austin, Tu, Ho, Levy, & Lee, 2013; Kourou, Exarchos, Exarchos, Karamouzis, & Fotiadis, 2015; Ferroni, Zanzotto, Scarpato, Riondino, Nanni, Roselli, & Guadagni, 2017) by looking at data doctors generally neglect. AI is preparing a wonderful future where people are released from the burden of repetitive jobs.

The bright AI revolution may have a possible dark side: a dramatic mass unemployment that may precede an unpredictable job market transformation. People and, hence, think
Zanzotto

tanks (Manyika, Chui, Miremadi, Bughin, George, Willmott, & Dewhurst, 2017; Stone, Brooks, Brynjolfsson, Calo, Etzioni, Hager, Hirschberg, Kalyanakrishnan, Kamar, Kraus, Leyton-Brown, Parkes, Press, Saxenian, Shah, Tambe, & Teller, 2016) and governments (Executive Office of the President of the United States of America, 2016) are frightened. A pessimistic report (Manyika et al., 2017) of the McKinsey Global Institute (MGI) foresees that AI may globally replace the equivalent of the activities of 1.1 billion employees by erasing $15.8 trillion in wages. By releasing people from repetitive jobs, intelligent machines may leave a majority of citizens with the value of their labor insufficient to pay for a socially acceptable standard of living (Stone et al., 2016). The revolution is ongoing. Chatbots are slowly replacing call center agents in some of their tasks. Self-driving trains are already reducing the number of drivers in our trains. Self-driving cars are fighting to replace cab drivers in our cities. Drones are expanding automation in managing delivery of goods by drastically reducing the number of delivery people. And, even more cognitive and artistic jobs are challenged. Intelligent machines may produce music jingles for commercials (Briot, Hadjeres, & Pachet, 2017), write novels, produce news articles and so on. Intelligent risk predictors may replace doctors (Austin et al., 2013; Kourou et al., 2015; Ferroni et al., 2017). Chatbots along with massive open online courses may replace teachers and professors (Gohd, 2017). Coders risk being replaced by machines too (Murphy, 2017). According to the White House report on AI (Executive Office of the President of the United States of America, 2016), this overwhelming progress of AI can initiate long-standing disruptions of local markets and, according to the MGI report (Manyika et al., 2017), nobody’s job will be left unchanged.

Surprisingly, the rise of AI is largely supported by the knowledge of an unaware mass of people who risk seeing a large part of their wages canceled by machines. In fact, along with someone selling knowledge for peanuts with Amazon Mechanical Turk, Crowdflower or SurveyMonkey and along with those aware programmers who set up these intelligent machines, an unaware mass of people is providing precious training data by passionately doing their job – translating, interacting with customers, teaching – or simply performing their activities on the net – answering an email, interacting on messaging services, leaving an opinion on a hotel. These data are a goldmine for AI machines as learning systems transform these interactions in knowledge. By doing their normal everyday activity, many workers are shooting themselves in the foot and unaware people are “donating” their knowledge to machines. This is an enormous and legal knowledge theft taking place in our modern era.

As researchers in Artificial Intelligence, we have a tremendous responsibility: building intelligent machines that “support the parents of their intelligence” (Stone et al., 2016) rather than intelligent machines that steal their knowledge to do their jobs. Moreover, we need to find ways to financially support job seekers as they train to catch up with these novel unpredictable jobs. We need to prepare an antidote as we spread this poison in the job market.

This paper proposes Human-in-the-loop Artificial Intelligence (HitAI) as a novel paradigm for a responsible Artificial Intelligence. The idea is simple: giving the right value to the knowledge producers. Recognizing that any AI system has humans in the loop, HitAI promotes interpretable learning machines and, therefore, artificial intelligence systems with a clear knowledge life cycle. For HitAI systems, it will be clear whose knowledge has been used in a specific deployment or in specific situations. This is a way to give the rightful
credit and revenue to the original knowledge producers. Hence, HitAI is a possible antidote to the poisoning of the job market.

The rest of the paper is organized as follows. Section 2 describes current trends and the enabling paradigms for Human-in-the-loop AI. Section 3 sketches a proposal for a better future. Then, Section 4 draws some conclusions.

2. Human-In-The-Loop AI: Trends and Enabling Paradigms

Nowadays, data seem to be the principal source of knowledge for machines. This section describes: (1) how data have become more important than programmers to “teach” machines (Sec. 2.1); (2) how explainable Artificial Intelligence (Sec. 2.2) and distributed representations for symbols (Sec. 2.3) can be used to understand how these machines learn from data. This description is extremely useful for proposing the agenda of Human-in-the-loop Artificial Intelligence.

2.1 Transferring Knowledge to Machines: from Programming to Autonomous Learning

Since the beginning of the digital era, programming is the preferred way to “teach” machines. Artificial non-ambiguous programming languages have been developed to have a clear tool to tell machines how to solve new tasks or how to be useful. According to this paradigm, whoever wants to “teach” machines has to master one of these programming languages. These people, called programmers, have been teaching machines for decades and have made these machines extremely useful. Nowadays, it is difficult to imagine passing a single day without using the big network of machines that programmers have contributed to building.

Not all the tasks can be solved by programming, so autonomous learning has been reinforced as an alternative way of controlling the “behavior” of machines. In autonomous learning, machines are asked to learn from experience. With the paradigm of programming, we have asked machines to go to school before these machines have learned to walk through trial and error. This is why machines have always been good in solving very complex cognitive tasks but very poor in working with everyday simple problems. The paradigm of autonomous learning has been introduced to solve this problem.

In these two paradigms, who should be paid for transferring knowledge to machines and how should they be paid? In the programming paradigm, roles are clear: programmers are the “teachers” and machines are the “learners”. Hence, programmers could be paid for their work while they are teaching machines that are learning. In the autonomous learning paradigm, the activity of programmers is confined to the selection of the most appropriate learning model and of the examples to show to these learning machines. Nobody is paid while machines are learning.

From the point of view of HitAI, the trend of shifting from programming to autonomous learning is dangerous. In fact, programming is a fair paradigm as it keeps humans in the loop. On the contrary, autonomous learning is an unfair model of transferring knowledge as the real knowledge is extracted from data produced by unaware people. Hence, little seems to be done by humans and machines seem to do the whole job. Yet, knowledge is stolen without paying what it is worth.
2.2 Explainable Autonomous Learning Machines

Explaining the decisions of autonomous learning machines is a very hot topic nowadays: dedicated workshops or specific sessions in major conferences are flourishing (Aha, Darrell, Pazzani, Reid, Sammut, & Stone, 2017; Kim, Malioutov, Varshney, & Weller, 2017). In specific areas of application, for example, medicine, trust in intelligent machines cannot be blind as final decisions can have a deep impact on humans. Hence, understanding why a decision is taken becomes extremely important. However, what is exactly an explainable machine learning model is still an open debate (Lipton, 2016).

In HitAI, explainable machine learning can play a crucial role. In fact, seen from another perspective, explaining machine learning decisions can keep humans in the loop in two ways: 1) giving the last word to humans; and, 2) explaining what data sources are responsible for the final decision. In the first case, the decision power is left in the hands of very specialized professionals who use machines as advisers. This is a clear case of human-in-the-loop AI. Yet, this is confined to highly specialized knowledge workers in some specific areas. The second case instead is fairly more important. In fact, machines that take decisions or work on a task are constantly using knowledge extracted from data. Spotting which data have been used for a specific decision or for a specific action of the machine is very important in order to give credit to whoever has produced these data. In general, data are produced by anyone and everyone, not only by knowledge workers. Hence, understanding why a machine takes a decision may become a way to keep everybody in the loop of AI.

2.3 Symbolic Knowledge and Distributed Representations in Learning Machines

In current AI systems, knowledge is stored in tensors of real numbers called distributed representations. These representations are pushing learning models (LeCun, Bengio, & Hinton, 2015; Schmidhuber, 2015) towards amazing results in many high-level tasks such as image recognition (He, Zhang, Ren, & Sun, 2016; Simonyan & Zisserman, 2015), image generation (Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, & Bengio, 2014), image captioning (Vinyals, Toshev, Bengio, & Erhan, 2015b; Xu, Ba, Kiros, Cho, Courville, Salakhudinov, Zemel, & Bengio, 2015), machine translation (Bahdanau, Cho, & Bengio, 2014; Zou, Socher, Cer, & Manning, 2013), syntactic parsing (Vinyals, Kaiser, Koo, Petrov, Sutskever, & Hinton, 2015a; Weiss, Alberti, Collins, & Petrov, 2015) and even game playing at a human level (Silver, Huang, Maddison, Guez, Sifre, Van Den Driessche, Schrittwieser, Antonoglou, Panneershelvam, Lanctot, et al., 2016; Mnih, Kavukcuoglu, Silver, Graves, Antonoglou, Wierstra, & Riedmiller, 2013).

Explaining AI decisions seems simple when learning machines treat images as this knowledge is stored similarly to distributed representations. For example, in neural networks, input images and layers of the networks are tensors of real numbers. Interpreting these networks is generally done by visualizing how layers represent salient subparts of target images. Hence, these networks can be examined and understood.

However, a large part of the knowledge is not expressed in images but with symbols, which apparently are not similar to distributed representations. Both in natural and artificial languages, combinations of symbols are used to convey knowledge. In fact, for natural languages, sounds are transformed into letters or ideograms and these symbols are com-
bined to produce words. Words then form sentences and sentences form texts, discourses, dialogs, which ultimately convey knowledge, emotions, and so on. Hence, to explain decisions of learning machines, we need to understand how symbolic knowledge is represented in distributed representations.

For HitAI, there is a tremendous opportunity to track how symbolic knowledge flows in the knowledge life cycle. Although symbols seem to fade away in current AI systems, there is a strict link between distributed representations and symbols, the first being an approximation of the second (Fodor & Pylyshyn, 1988; Plate, 1995; Zanzotto, Ferrone, & Baroni, 2015; Ferrone & Zanzotto, 2017). In this way, symbolic knowledge producers can also be rewarded for their unaware work.

3. Human-In-The-Loop AI: A Simple Proposal for a Better Future

The AI revolution is largely based on an enormous knowledge theft, which will possibly be a never-ending source of revenues for companies. In fact, skilled and unskilled workers do their own everyday jobs and leave important traces. These traces are the training examples that machines can use to learn. Hence, AI is stealing these workers’ knowledge by learning from their interactions. The stolen knowledge is going to produce never-ending revenues for companies for years. This is a major problem since only that very small fraction of the population who own shares of these companies can benefit from this never-ending revenue source and the real owners of the knowledge are not participating in this redistribution of wealth.

The model we propose with Human-in-the-loop Artificial Intelligence (HitAI) seeks to give back part of the revenues to the real owners – the knowledge producers. The key idea is that any profit-making interaction a machine does has to constantly repay whoever has produced the original knowledge used to do that interaction. Assigning rewards per decision is very important as it may be an incentive to produce better services today and to have better services in the future. In fact, people have an incentive to work better knowing that their future revenues depend on how they treat difficult and odd cases today.

Realizing HitAI poses two big challenges: the “political” challenge of convincing companies to share benefits with producers of the data; the “infrastructural” challenge of managing ownership of knowledge in the knowledge life cycle. These are two different, yet interrelated challenges.

The “political” challenge is very tough: convincing companies to share the benefits of AI technologies may be impossible. But, if a free society cannot help the many who are poor, it cannot save the few who are rich (Kennedy, 1961). In fact, in the long run, if there is not a way to redistribute benefits to knowledge producers – the poor people –, the overall market will collapse and, hence, companies – the rich people – will also lose these benefits. Although reasonable, this may not be an argument for today’s CEOs focused on the short term.

Hence, if convincing companies is impossible, companies should be forced to share benefits. A possibility to achieve this is to start by protecting personal data with two mechanisms:

- a legal mechanism: protecting unaware knowledge production by extending the copyright law
• a technological mechanism: promoting new “ownership-aware file systems” which release and accept data only if owners are specified.

Both mechanisms should be promoted by governments. The legal mechanism is very slow as it has to go from the national level to the international level. The technological mechanism can be faster as it may be funded by local government grants or by spontaneous social movements such as the nordic model of MyData\(^1\) and, then, spread all over as a novel concept of ownership-aware file system strongly required by final users.

The “infrastructural” challenge is different but again difficult: HitAI needs technologies that produce a trusted knowledge life cycle of AI systems in order to reward knowledge producers. Managing ownership in the knowledge life cycle poses major technological and moral issues and it is certainly more complex than simply using knowledge while forgetting its source. Each interaction has to be tracked and assigned to a specific individual.

To build a trusted knowledge life cycle of AI systems, we need to investigate two core problems: first, building *Knowledge Life Cycle Transparent Artificial Intelligence* systems and, second, building *Trusted Technologies*.

Indeed, building *Knowledge Life Cycle Transparent Artificial Intelligence* is mandatory because, in order to reward knowledge producers, systems need to exactly know who is responsible for a specific decision. *Knowledge Life Cycle Transparent AI* may share techniques with *explainable AI* (Sec. 2.2) but its focus is different: identifying the causal source of each inference of AI systems. For Neural Networks (Haykin, 1998) (NN), it may seem impossible to identify the causal source of an inference. This is a research topic. Yet, the causal source of an inference can be determined in many machine learning models. In support vector machines (Cortes & Vapnik, 1995; Cristianini & Shawe-Taylor, 2000), it is possible to determine which support vector is participating to each single decision and which is its “role”, that is, its weight, with respect to that decision. In decision tree learning (Quinlan, 1993), each decision node is justified by a set of examples. Hence, even if today it seems difficult to trace back causal source of an output in NN, this can be a promising field of research.

Designing *Trusted Technologies* is the second core problem as the knowledge life cycle should be clear and knowledge items correctly tracked. This core problem is largely linked to the new “ownership-aware file system”, which has been mentioned above. However, HitAI needs to guarantee that data in circulation have specific owners without revealing who the owner is. This is also required by the General Data Protection Regulation (GDPR) (European Parliament and European Council, 2016). Hence, HitAI will use *Digital Identity Protocols and Mechanisms* (Camp, 2004) to identify people but, at the same time, HitAI will ensure the use of *Privacy Preserving Protocols and Mechanisms*, which can be obtained by using data encryption and *Block Chains* (Nakamoto, 2008; Tschorsch & Scheuermann, 2016).

HitAI is an alternative to other solutions of wealth redistribution like Universal Basic Income (UBI) (Straubhaar, 2017), which may not be viable (Zheng, Guerriero, Lopez, & Haverman, 2017). In fact, UBI should be governed at the national level through taxes. But, generally, companies act in a transnational level paying taxes on revenues where it is more convenient. Hence, UBI may not be easy to apply.

\(^1\) https://mydata.org/
4. Conclusions

The bright Artificial Intelligence revolution has a dark side: a possible, dramatic job market contraction before its unpredictable transformation. A peasant or, even, a wise politician of the late 19th century would have never imagined that after 100 years yoga trainer, pet caretaker and ayurveda massage therapist were common jobs. Today, the situation is similar with a complication: the speed of the AI revolution. It’s hard to imagine what’s next on the job market and, hence, what are the skills needed for being part of the labor force of the future. We urge a strategy for the immediate future to mitigate the dark side of the AI revolution.

In this paper, we proposed Human-in-the-loop Artificial Intelligence (HitAI) as a fairer AI approach, which leverages the most gigantic knowledge theft of the modern era. In fact, unaware skilled and unskilled knowledge workers are shooting themselves in their feet by passionately doing their normal, everyday work as these workers are producing the knowledge which Artificial Intelligence is making a profit on. HitAI aims to give back a large part of this profit to its legitimate owners. As modern Merry Men, HitAI researchers should fight for a fairer Robin Hood Artificial Intelligence that gives back what it steals.

References


European Parliament and European Council (2016). Regulation (EU) 2016/679 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on

Executive Office of the President of the United States of America (2016). *Artificial Intelligence, Automation, and the Economy*.


