

Taming the Techno Leviathan: Why We Should Adopt a Society-in-the-Loop Model Inside IoT Utilities



Alfredo Adamo

Abstract While the Internet of Things (IoT) has made significant progress along the lines of supporting individual applications, it is only recently that the importance of people as an integral component of the overall IoT infrastructure has started to be fully recognized. Several powerful concepts have emerged to facilitate this vision, whether involving the human context whenever required or directly impacting user behavior and decisions. As these become the stepping stones to develop the IoT into a people-centric utility, this paper outlines how to include the “Society-in-the-loop” approach to govern a lot of ethical, moral concerns. Rapid advances in IoT, Artificial Intelligence and Machine Learning have raised many questions about the regulatory and governance mechanisms for autonomous sensors, machines and infrastructures. We discuss about, in the context of the IoT utility, a lot of concerns raised about algorithms governing our lives, and how the adoption of “Society-in-the-loop” paradigm could be a solution.

Keywords IoT (Internet of Things) · Artificial intelligence · Machine learning · HITL (human in the loop) · SITL (society in the loop) · Humanities

1 Introduction

In our modern society, science and humanism are often considered two distinguished, often perceived as opposite, entities. The figure of the scientist is regarded as a rational individual, which pursues the research of the truth through experiments and factual plans; while the humanist is a creative individual, with compelling arguments at best, but whose research of truth falls merely in an epistemological discourse.

However, this hasn't always been true. Before the introduction of the scientific method, science and philosophy walked hand in hand in the academic universe. The Italian renaissance, particularly, witnessed the birth of polyvalent intellectuals, who mastered both the scientific, mathematic side of arts and the creative, theoretical facet of natural sciences. As instance, Galileo Galilei was indeed one of the most

A. Adamo (✉)
Alan Advantage, Rome, Italy
e-mail: alfredo.adamo@alanadvantage.com

© Springer Nature Switzerland AG 2020
G. Di Francia et al. (eds.), *Sensors and Microsystems*, Lecture Notes
in Electrical Engineering 629, https://doi.org/10.1007/978-3-030-37558-4_18

119

relevant men of science, but he also contributed to the European philosophic scene with his *The Assayer*, where he poetically states that mathematics is the language with which the entire Universe is coded and created. Leonardo Da Vinci created masterpieces as the Gioconda and The Last Supper, but he was also a pioneer of revolutionary engineering breakthroughs as the helicopter, the tank, or the parachute. It is obvious how technology (although understood in a different way) and humanism were not perceived as mutually independent. The recent, gargantuan developments in the various fields of technology could not help but exacerbate such rupture.

However, I argue that this is not just an intellectual, academic conundrum. The loss of humanism in the recent, techno-scientific progress has led to actual ethical dilemmas, caused by such increasing, uncontrolled social responsibility computers (one should only think of Facebook's data breach by Cambridge Analytica or the, for instance). Dismissing the relevance of the human component in something so influential as technology—and focusing only on the progress for its own sake—is distracting from the main reason technology exists, that is being of support and improvement to the human life.

Nonetheless, it is safe to say that, in recent years, the need to reintegrate the human component in the hi-tech universe has resurfaced in the academic world, with many prominent scholars sharing their publications on such an important matter. Borrowing the words of Luciano Floridi, professor of Philosophy and Ethics of Information at the University of Oxford, it is exactly now, in this age of “ubiquitous computing” [1: 43] the Information and Communication Technologies (ICTs) provide, that we should pursue the need to reintegrate humans, and society at large, in the technological progress.

In his *The Onlife Manifesto*, Floridi explores how ICTs' power to permeate every fiber of the social context, together with the newest Artificial Intelligence technologies (as well as apps, robots, and devices of different kinds) allowed the creation of a human-computer hybridization, changing humankind's vision. The result is the creation of a new social framework in which men, and society itself, are struggling to adjust its pre-existing norms, values, and behavioral codes [1: 43]. Such is the impact of ICTs on society that Floridi argues their advent can be seen as “a fourth revolution in our political anthropology”. That is after the innovations which changed our understanding of the world, as the ones by Copernicus, Darwin, and Freud [2: 21]. Hence, humanity ended up in a new phase of the information age, in which it almost became an appendix, a stranger who needs to learn how to live in such a new, disruptive reality.

It is obvious, then, that there is a compelling necessity to make the human component central to technology again, to reunite the humanistic side of progress with the hard, scientific facet of the developmental process. Or better: there is the necessity *to reintegrate society in the loop of supervisory control*, to avoid many of the ethical, moral dilemmas we face in our age.

In the next paper I will argue how a wider supervision by society can be achieved by the so-called “Society-in-the-Loop” model (SITL) [3]. Briefly, it is an evolution of the pre-existing Human-in-the-loop system (HITL), which implied supervision from an individual being. SITL, however, does not stop at the individual supervision, but

calls into action the wider social context, providing a more inclusive, democratic supervision, avoiding discriminatory algorithms as well.

Following the study of many scholars, my intention is to present a conceptual framework in which the adoption of a SITL system could provide a solution to our conundrum. A reconnection between humanism and technology is possible, and it should start with the inclusion of society in the technological, developmental process.

2 Modern IoT and Humans

Beyond legacy embedded systems with constrained applicability, the emerging IoT solutions are becoming more open and integrated by adaptively combining sensors and actuators with actionable intelligence for automatic monitoring and control. However, as a multitude of interconnected and intelligent machines communicate with each other and autonomously adapt to changing contexts without user involvement, the fact that present technology is made by humans and for humans is often overlooked. Indeed, modern IoT systems are still widely unaware of the human context and instead consider people to be an external and unpredictable element in their control loop, because of their unpredictable behavior both as user of the IoT scenarios and as a person who is present in the environment. Therefore, future IoT applications will need to intimately involve humans, so that people and machines could operate synergistically. To this end, human intentions, actions, psychological and physiological states, and even emotions could be detected, inferred through sensory data, and utilized as control feedback.

Many steps in the fields of supervisory control have been achieved in the last years, feeling the compelling necessity to regulate such complex, fast-developing world, but it is only recently, with the evolution of human-computer interaction, that scholars and theorizers feel the necessity of a more relevant human presence in the AI learning processes. An important milestone in the field has been the conceptualization of the Human-in-the-Loop model, as we saw in the introduction to this paper above. Taking part in the loop of reciprocal learning, the human supervisor takes active part in AI Machine Learning process, providing not only a better performance, but also controlling possible computer misbehavior and serving as a legally accountable subject, minimizing the probability of misbehavior at the expense of third parties [3: 7] (Fig. 1).¹

However, being the essence of progress an unrestrained run towards the future, computers and bots are having more and more computing power and influencing

¹As fictional as it could seem, however, the European Union, in 2016, has approved a motion aimed at the creation of a body of civil laws to make computers and bots legally accountable for their "actions" [4]. The motion also provides a framework for the establishment of a European agency for the regulation of hi-tech production and several, ethical considerations on the use of machines.

The documents the EU voted upon can be found at:

<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0005+0+DOC+XML+V0//EN#title10>.

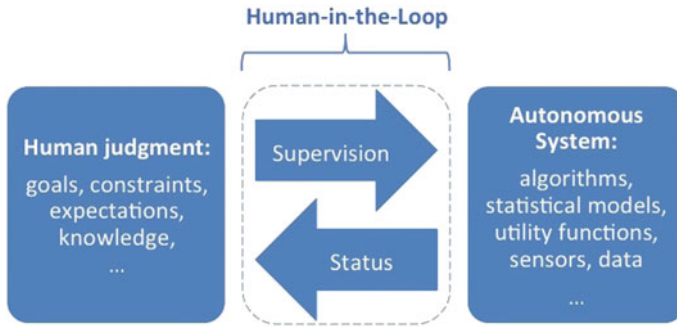


Fig. 1 Human in the loop

forces on our everyday life. Machines are entrusted with tasks increasingly affecting our actions and decisions, let alone the ways we connect to other humans.

Human behavior may be impacted in either space (for example, the users are encouraged to move to a less congested location) or time (for example, the users are convinced to reduce their current data demand in case the network is overloaded); this is known as the “user-in-the-loop” (UIL).

With UIL, often referred to as “layer 8”, the space-time user traffic demand may be shaped opportunistically and better matched with the actual resource supply from the people-centric wireless system. While HITL involves the user whenever human participation is desired or required and UIL extends the user’s role beyond a traffic-generating and traffic consuming black box, these trends must account for the fact that people are, in essence, walking sensor networks. Indeed, a wide diversity of user-owned companion devices, such as mobile phones, wearables, connected vehicles, and even drones may become an integral part of the IoT infrastructure.

Hence, they can augment a broad range of applications, in which human context is useful, including traffic planning, environmental monitoring, mobile social recommendation, and public safety, among others. Therefore, we envision that—in contrast to past concepts where the user only assists the network to receive better individual service—future user equipment will truly merge with the IoT architecture to form a deep-fused human–machine system that efficiently utilizes the complementary nature of human and machine intelligence.

3 Our Dilemma in a Nutshell

To start with, we should first ask ourselves why it is so important to (re)integrate humans in the supervisory loop. Deep down, isn’t progress existing to improve human livelihood, to remove every day’s hassles for users and consumers? Shouldn’t we instead, through a classic, heuristic trial and error, arrive at the point in history where

we could lie down, relax, and let machines wipe out the sweat of our historical fatigues?

Sure, it would be a great life thinking about it now. But, as already demonstrated, the world of hi-tech works in ways much similar to those of a hierarchical, political pyramid, where the top of the system rules and governs the ways algorithms are coded and implemented in our everyday life. That is to say: who decides and programs computer algorithms will end up having a massive influence on other people’s lives, in the same fashion as autocratic politicians. Without a more involving, democratic regulation of hi-tech “politics” we could be going towards a “new feudal order”, in which a restricted circle of people decides for the vast majority [5: 19].

Although the conceptualization of the HITL model has been a most relevant achievement in the solution of our dilemma, what happens when a computer is entrusted with a task which implies a broader social impact (as an algorithm which could influence mass political preferences or mediate resources and labor within a given country)? Can a single supervisor act on behalf of the entire social context?

It is for this new challenge that Iyad Rahwan, professor at the MIT Media Lab, has proposed an extension of the HITL model in order to integrate the wider social context into fundamental decisions and supervision of digital behavior: The Society-in-the-Loop system. With the integration of society in the loop we could meet the needs of a larger majority of the population, serving as a supervisor not only for the computer performance, but also for programmers and experts behind. In short, such a model would protect the rights of the various societal actors and allow them to enjoy a more sustainable and democratic use of AI and algorithms in their life (Fig. 2).

Nonetheless, if we accept the compelling necessity to integrate what Rousseau called the general will of the people, how do we define, then, what is best for *all* the social actors? How can we safely state that *x* is better than *y* in the full respect

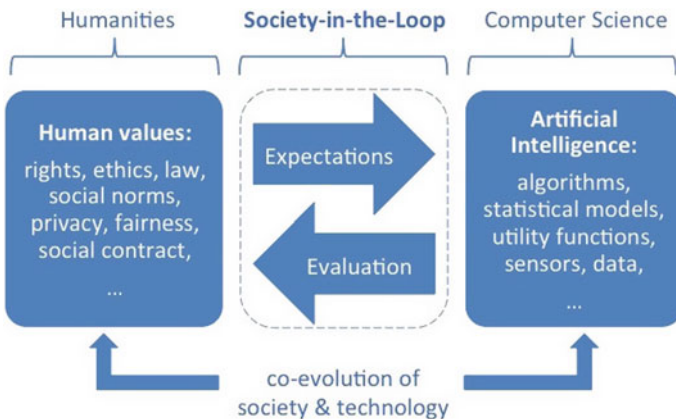


Fig. 2 It is worth noting how while a HITL system involved individual judgment on the computer’s performance, a SITL involves a further consideration on human values, expected to be implemented in the algorithmic applications

of everyone's human rights? Let's have a look to this equation, with which Rahwan summed its SITL model:

$$SITL = HITL + SocialContract$$

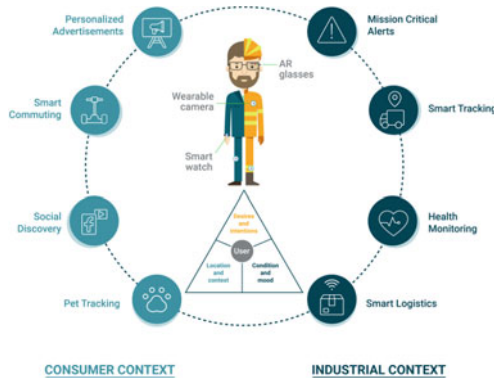
Having in mind what HITL and SITL stand for, the only remaining variable to define is exactly the "social glue" for the interests of every individual within the social context: The Social Contract.

4 Many Categories of People-Centric IoT Services Emerge and Are Expected to Be Deployed Over the Following Years

Intention- and mission-aware services. These services primarily reflect user's current intention or desire and assist by enabling, for example, situation-aware smart commuting for pedestrians, cyclists, and drivers of scooters, trucks, and other vehicles. This group of applications can help people in a variety of use cases, from highlighting the nearest available parking space on a vehicle's head-up display in urban areas to status reporting on a display or augmented reality (AR) glasses in challenging environments, such as mines, construction sites, etc.

Location- and context-aware services. Another group of services is formed by location- and context-aware applications, such as those communicating alerts from environmental sensors (for example, "put on/take off your mask" when entering/leaving a polluted area). Many more of these services are envisioned to be deployed in the coming years, such as identifying slippery floors and low ceilings, notifying about forgotten trash when a user is about to leave the house, and many other examples.

Condition- and mood-aware services. A deeper level of IoT penetration into people's lives can be achieved by integrating city/area infrastructure with personal medical and wellness devices. For instance, dietary restrictions could be applied on a menu when ordering food or a squad leader may be advised to give a break to a worker whose blood pressure has recently gone up.



Summarizing the above examples of services, we note that depending on the environment the set of requirements and challenges to implement a particular application may vary considerably.

To further offer a challenges-based grouping, we propose to differentiate between two major contexts: consumer and industrial.

The former is characterized by the presence of numerous devices that are heterogeneous in terms of their communication means and ownership. Therefore, the major challenge in this context is to provide sufficient scalability of the deployed connectivity solution.

On the contrary, the latter context is more challenging in terms of maintaining communication reliability due to more difficult propagation environments. At the same time, the system operator has more control over device population in such areas.

We continue by addressing how people-centric IoT applications are to be engineered, that is, which radio technologies need to be employed in particular scenarios and how to ensure their suitability for the target operating conditions.

5 Definition Problems: How Do We Determine What’s Best?

Imagine a country devastated by a civil war. The government splits in two factions, those loyal to the former power and those willing to get rid of the old, antiquated political order. Armed conflict is at the door and, eventually, the two factions meet on the battlefield. National confusion and political unrest lead to economic crisis, deaths, and destruction. Finally, the former ruler is beheaded and a new political system is established, changing the governmental landscape of the country. This was Great Britain in the 17th century, when Oliver Cromwell and the Parliament rebelled against the king and established the “republican” Commonwealth.

In hindsight, it is understandable why the Social Contract Theory made its first appearance in England, where the lack of a stable, central power before—and the

sudden change in the political system after—the revolution inspired Thomas Hobbes to write his famous *Leviathan*. Briefly, the English philosopher suggested that the political order (granted by the existence of a superior power, embodied in the state and the government) is the result of a compromise, a *contract*, men agreed upon to enjoy social order and security. Before such an agreement, men lived in a pre-political, pre-moral State of Nature, which he calls the *bellum omnia contra omnes*, the war of all against all. To escape the ferial stage men had to give up part of their freedom and accept a *Pact Unionis*, where they undertook to respect each other and live in harmony, and a *Pact Subjectionis*, where they agreed upon the establishment of a superior, ruling power which could defend and guarantee social order.

The Social Contract Theory, then, was reexamined by other philosophers, as John Locke and Jean-Jacques Rousseau, who re-elaborated its theoretical foundations—the former moving toward a constitutional contract, where the ruling power is not uncontested; the latter beginning a democratic, fundamental discourse on people’s *volonté générale*. However, although their arguments are of the utmost philosophical importance, in all its developments the Social Contract kept the fundamental characteristic of being an agreement all the individuals must agree upon to secure their wellbeing in a given social context.

But how does this pertain to our conundrum? Simple: *if we want to move towards a more involving, democratic use of AI and technology at large, we have to understand what can best serve the general will of all the social actors*, integrating the wider social context into the supervisory loop, as argued above. The SITL system extends decisional power to a larger group, involving two important features a HITL model does not:

- (1) Better treatment of common-sense choices an AI cannot face nor process, since it has not the morale a human mind has (as choosing between efficiency and safety or favoring fair and just options although not mathematically perfect).
- (2) Better understanding of the social costs and benefits in the implementation of a new innovation. For instance, whether a self-driving car (a much controversial technology, especially after a fatal crash in Arizona, where a pedestrian was killed in the collision) has to prioritize the safety of the driver or of the pedestrians it cannot be decided by the machine itself. It has to be programmed beforehand.

In brief, then, the “Social Contract” is what society agrees upon to deal with many of the problems affecting the individuals’ wellbeing and security. As humankind needed a political order to preserve their selves in the pre-political State of Nature, we need, today, a protection for—and more involvement in—our digital selves.

Helbing and Pournaras [6], in an article on *Nature*, express thoroughly how a “top-down control has various flaws”. First off, it can be subject to corruption and hacked by extremists and criminals. Second, it fails to address local needs, being bound to limitations in data-transmission rates and processing power. Third, forcibly intervening in individual choices undermines collective intelligence, a most important factor in the creation of a diverse, inclusive society. Fourth, “filter bubbles” consequent to personalized information cause people to be less exposed to other

opinions, increasing polarization and conflict. The reduction in pluralism and diversity is very dangerous for ecosystems relying on interdependencies as society and economy. Finally, top-down control can alter people's decisions, disrupting everyday decision-making skills and undermining stability and order, rather than providing a more secure environment.

6 Easier Said Than Done

However, although such a model could be a milestone in the achievement of a more democratic hi-tech development, there are still many doubts on how to successfully carry it out. For instance, a fundamental problem is the disciplinary gap existing between machine programming and the legal, ethical values of social sciences. Although professionals and scholars in the various social and legal subjects are able to identify possible computer misbehaviors, it is not that simple to mathematically articulate how a machine should behave lest a computer inflict moral and ethical damages. Furthermore, the human-computer relation is characterized by a constant, reciprocal learning. Human users' IT skills are in constant evolution, thus changing what the wider social context deems acceptable.

Modern demography, moreover, does not help our efforts in implementing such a revolutionary solution, since we should be dealing with a huge amount of people, each with their own opinions and ideas. In this way, we should address every individuals' perception of what is "fair", "correct", and "acceptable", which is almost impossible or, at least, very time consuming. Hence, how could we reunite all people's awareness and have an arithmetic mean of all their feelings? How can we quantitatively define their desires and point of views on ethical matters?²

Rahwan [3: 11–12] tries to put forward some solutions to the conundrum, considering some of the solutions put forward by scholars in the field. One of these is the use of *crowdsourcing* techniques and tools to produce a database which could store the general preferences of a society.³ Together with his colleagues at the Massachusetts Institute of Technology, he developed a public-facing survey tool exhorting the participants to answer ethical dilemmas as: "If this self-driving car is doomed to crash, is it better that it kills x number of pedestrians (including a pregnant woman) or

²One should only think that *computational science* is still trying to give an answer to this, researching algorithms that could calculate the optimal results in trade-offs decisions involving several stakeholders. In our case this can be seen through the lenses of a Steinhaus' "fair cake-cutting" problem [7]: we have a number N of stakeholders (all the individuals society is consisted of) and a good (algorithmic computer behavior) they are interested in (or, in this case, they demand for security and wellbeing). Thus, we have a good which should be *fairly divided* among all the stakeholders, without discrimination and in the full respect of everyone's opinion. Steinhaus expressed this conundrum in the form of a cake which should be divided among various gluttonous children. The problem arises in the moment that the cake has assorted toppings, meaning that the good offered is heterogeneous, complicating our research of an *optimal, fair division* among the children/stakeholders.

³Crowdsourcing as a solution to deal with trade-offs is also well explored by Conitzer et al. [8].

the passengers (including a family with children, for instance)?" [9]. The perk here is that collecting and analyzing their answers can tell us a lot about how the wider social context deals with such ethical, moral dilemmas.

Moreover, the possibility of having a human audit of algorithms and programs is explored in the paper, but Rahwan believes that such kind of supervision could benefit from automation, entrusting other algorithms with the task to audit computer behavior. This is the overall point we find in Etzioni and Etzioni when they state that: "To ensure proper conduct by AI instruments, people will need to employ other AI systems" [10: 155], foreseeing the use of "oversight programs" to achieve a successful automation of the auditing process.

From a technical-scientific point of view, the progress of GANs seems to indicate a viable path to have an automatic treatment at the base of the SITL-based approach. A generative adversarial network (GAN) is a class of machine learning systems. Two neural networks contest with each other in a zero-sum game framework. This technique can generate photographs that look at least superficially authentic to human observers, having many realistic characteristics. It is a form of unsupervised learning.

However, I find that employing algorithms or GANs to check for possible computer misbehavior is rather tautological. Is defying today's habit to put technology at the center not our main dilemma? The technochauvinist morale to zeroing in on technology for its own sake, excluding the importance of the human component, is what I am trying to point out in this paper: outsourcing the auditing responsibility to other machines would not help (re)integrating humans in the loop, but would confine them again at the borders of the loop, placing at the center of progress the mere will of unrestrained technological development.

I am not, by all means, trying to develop a luddite argument, in favor of the complete destruction of technology to achieve an Amish-like society. I do know that our society will be—if it is not already—data-driven and enmeshed more and more in the massive use of innovative technologies. However, as I already argued above, the use of technology for the sole mean to launch new innovations on the market is creating more problems than improving our lives.

What we need is not *only* the control of machines and algorithms (whether automated or by humans), but the control of those behind the machines, namely the entrepreneurs, the programmers, and the experts coding the algorithms affecting our lives and society. The supervisory loop needs to be extended to the use humans themselves make of our data and our digital selves, aside from the supervision of the pragmatic "actions" of human-programmed computers.

Fortunately, we might see some of these improvements. Back in 2013, Alex "Sandy" Pentland, entrepreneur and professor at the MIT, wrote that "to achieve a data-driven society, we need what I have called the New Deal on Data", where data would be seen as an asset and "individuals would have ownership rights in data that are about them". This means that an individual has full control over his/her personal data, including possession, crystal-clear terms of use, and the right to dispose of or distribute your data [11: 83]. Does this sound familiar? The General Data Protection Regulation (commonly known as GDPR) issued by the European Union in 2016

foresees exactly many of the points explored by Pentland. After the scandal of Cambridge Analytica breaching the data of Facebook’s users, this has become even more relevant to the good use of Internet tools as socials and websites.

7 Conclusion

Computers and machines are miracles of technology. They have had the power to reshape our understanding of the world and of the relations we build with other humans. Socials have changed the way people know and perceive each other, aside from having changed the way politics is conveyed to—and conceived by—the general public. They are amazing tools for our needs and whims. But that’s what they should remain: *a tool*.

In this paper I tried to put forward all the perks, as well as the *current* limitations, of a SITL model. I do not aim at concluding the debate with a finale, sole possible solution. Quite the contrary, I believe it is time to consider the possibility of such a system and to go back to a human-centered vision of technology. As I started this paper, with a title inspired by the disruptive paper by Iyad Rahwan, I would like to conclude quoting his words: “We spent centuries taming Hobbes’s Leviathan, the all-powerful sovereign. We must now create and tame the new Techno-Leviathan.”

References

1. Floridi L (2015) Background document: rethinking public spaces in the digital transition: the Onlife Manifesto. In: Floridi L (ed) The Onlife Manifesto. Springer, Cham. https://doi.org/10.1007/978-3-319-04093-6_4
2. Floridi L (2015) Commentary on the Onlife Manifesto. In: Floridi L (ed) The Onlife Manifesto. Springer, Cham. https://doi.org/10.1007/978-3-319-04093-6_4
3. Rahwan I (2017) Society-in-the-loop: programming the algorithmic social contract. DOI, Ethics Inf Technol. <https://doi.org/10.1007/s10676-017-9430-8>
4. Delvaux M (2016) Motion for a European Parliament resolution: with recommendations to the commission on civil law rules on robotics. Technical Report (2015/2103(INL)), European Commission
5. Citron DK, Pasquale FA (2014) The scored society: due process for automated predictions. Wash Law Rev 89:1–33
6. Helbing D, Pournaras E (2015) Society: build digital democracy. Nature 527:33–34
7. Procaccia AD (2016) Cake cutting algorithms. In: Brandt et al (eds) Handbook of computational social choice. Cambridge University Press, New York, pp 311–329
8. Conitzer V, Brill M, Freeman R (2015) Crowdsourcing societal tradeoffs. In: Proceedings of the 2015 international conference on autonomous agents and multiagent systems. International Foundation for Autonomous Agents and Multiagent Systems, pp 1213–1217
9. MIT (2017) The moral machine. Retrieved 01 Jan 2017, from <http://moralmachine.mit.edu>
10. Etzioni A, Etzioni O (2016) AI assisted ethics. Ethics Inf Technol 18(2):149–156
11. Pentland A (2013) The data-driven society. Sci Am 309(4):78–83. <https://doi.org/10.1038/scientificamerican1013-78>