

A Lesson from AI: Ethics Is Not an Imitation Game

Gonzalo Génova, Valentín Moreno Pelayo, M. Rosario González Martín

Abstract— How can we teach moral standards of behavior to a machine? One of the most common warnings against AI is the need to avoid bias in ethically-loaded decision-making, even if the population on which the learning is based is itself biased. This is especially relevant when we consider that equity (and protection of minorities) is an ethical notion that by itself goes beyond the (most probably) biased opinions of people: equity must be pursued and ensured by social structures, regardless of whether people agree or not. We know (believe?) that bias, or being biased, is a bad thing, regardless of what the majority says. In other words, *good and evil is not what the majority says*, it is beyond majorities and mathematical formulae. Ethics cannot be based on a majority opinion about right and wrong or on a rigid code of conduct. We need to overcome the generalized skepticism in our society about the rationality of ethics and values. The good news is AI is forcing us to think ethics in a new way. The attempt to formalize ethics in a set of rules misses the point that a person is not only an instance of a case, but a unique and unrepeatable being. Ethics should prevent us from the error of converting equity into mathematical equality, achieved through the extraction of characteristics and the computation of a value formula. Equity is not mathematical equality, not even a weighted equality that considers different factors.

Index Terms — biased judgment, critical thinking, equity and equality, learned ethics, programmed ethics, rationality of ethics.

I. ETHICS IS NOT AN IMITATION GAME

The Imitation Game is the title of the 2014 movie that tells the life of Alan Turing, and especially his outstanding participation in the decipherment of the German messages encrypted with the Enigma Machine in the Bletchley Park complex [1][2]. The expression “the imitation game” is from Turing himself: these are the first words of his 1950 article,

Submitted for review on October 1st 2020. This research has received funding from the RESTART project – “Continuous Reverse Engineering for Software Product Lines / Ingeniería Inversa Continua para Líneas de Productos de Software” (ref. RTI2018-099915-B-I00, Convocatoria Proyectos de I+D Retos Investigación del Programa Estatal de I+D+i Orientada a los Retos de la Sociedad 2018, grant agreement n.º: 412122; from the ECSEL18 project “NewControl” (Project 6221/31/2018), and its National PCI funding n.º 449990; and from the CitiRed project – “Elaboración de un modelo predictivo para el desarrollo del pensamiento crítico en el uso de las redes sociales”, Convocatoria Retos de Investigación del Ministerio de Ciencia, Innovación y Universidades (2019-2022), ref. RTI2018-095740-B-I00.

G. Génova is with the Computer Science and Engineering Department, Universidad Carlos III de Madrid, Spain (ggenova@inf.uc3m.es).

V. Moreno Pelayo is with the Computer Science and Engineering Dept, Universidad Carlos III de Madrid, Spain (vmmpelayo@inf.uc3m.es).

M.R. González Martín is with the Department of Educational Studies, Universidad Complutense de Madrid (marrgonz@ucm.es).

Computing Machinery and Intelligence [3]. It is also the name of a game played by the Victorian aristocracy, which consisted in a blind exchange of handwritten messages to try to guess whether the interlocutor was a woman or a man.

When Turing proposes his famous experiment –the Turing Test– to determine if a machine can think, his focus is on a notion of intelligence as the capacity to solve “closed” (that is, computable) problems, whose paradigm would be the chess game or the deciphering of keys. It is not that he did not care about the ethical questions, of course he cared, but possibly they did not enter within what he considered “intelligence”.

Those were the beginnings of artificial intelligence (AI). But the questions AI raises today about ethics in general, and equity in particular, are increasingly important (we mean ‘equity’ in the sense of attending to each one’s circumstances and needs, which usually does not imply the mathematical ‘equality’ of certain measurable personal features). We see a particular threat in some misconceptions about ethics among the developers of machine ethics. In this article we reflect on some of these questions, and also on *what we can learn* when we consider how to teach ethical behavior and equity to a machine.

II. PROGRAMMED ETHICS

The ethical behavior of machines has supplied the theme for a vast amount of more or less fanciful speculation, especially in science fiction. A paradigmatic example is Isaac Asimov and his legendary Three Laws of Robotics (see Fig. 1), first formulated in a story written in 1941 (*Runaround*), at the same time that Turing was working hard and doing something fairly significant in Bletchley Park. The short story would later become part of the collection *I, Robot* (which, by the way, has very little to do with the 2004 movie with the same title).

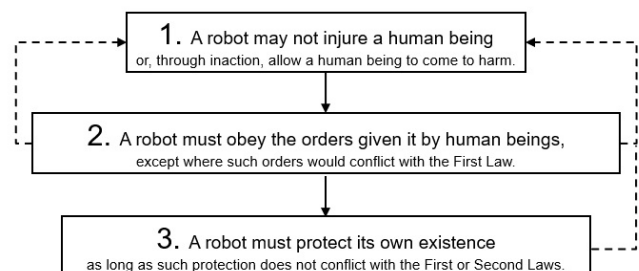


Fig. 1. The Three Laws of Robotics by Isaac Asimov

Leaving apart the conceptual and technical difficulties that a

hypothetical implementation of the Three Laws in a real machine would present, the interest of the Asimovian robotic universe is, in our opinion, twofold. It lies in the introduction of futuristic technological elements, but above all in the dynamics that results from the interplay of these three laws: the conflicts they provoke, and the ingenious –and sometimes heroic– solutions that the humans involved in the stories find. In a way, they are as much science-fiction stories as they are ethics, political or sociology-fiction stories.

The Three Laws of robotics are an attempt to explicitly program an ethical code into a robot, i.e. a sort of explicit or *programmed ethics*. Classical programming is an explicit procedural description of the instructions that a computational machine (a computer) has to follow to carry out a task. The task can be to add two numbers, to take a package from one place to another, to carry out a cornea transplant... This type of programming tries to break down the solution to a given problem into sequential instructions, repetitions and alternative paths (the latter depending on whether or not certain conditions are met).

Though the Three Laws serve as little more than an introduction to programmed ethics –and it is clear that Asimov only intended to use them as a literary device–, there have been preliminary attempts to implement them in a humanoid programmable robot [4]. However, even admitting the technical difficulties could be overcome at some point, it has also been argued that they would be an unsatisfactory basis for Machine Ethics [5].

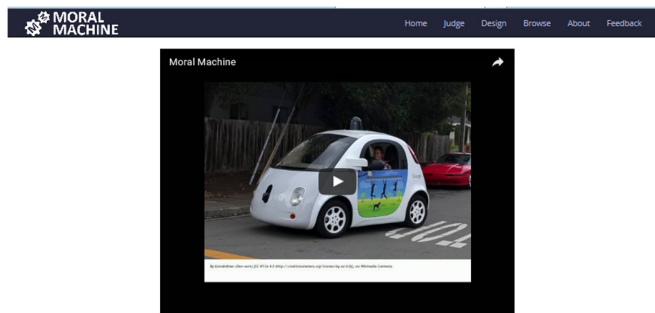
III. LEARNED ETHICS

The truth is that trying to understand ethics in that way, tightly bound within a code that is capable of contemplating all possible cases when faced with them (which is what a program does) is quite problematic, as recent surveys on machine ethics demonstrate [6][7][8]. That is why another approach has emerged, hand in hand with developments in AI. An approach that we can call implicit or *learned ethics*. At MIT (Massachusetts Institute of Technology) they have developed an experiment to “learn how to make machines moral” [9]. They have named the project The Moral Machine, which is presented as follows (see Fig. 2):

Welcome to the Moral Machine! A platform for gathering a human perspective on moral decisions made by machine intelligence, such as self-driving cars.

We show you moral dilemmas, where a driverless car must choose the lesser of two evils, such as killing two passengers or five pedestrians. As an outside observer, you judge which outcome you think is more acceptable. You can then see how your responses compare with those of other people.

If you're feeling creative, you can also design your own scenarios, for you and other users to browse, share, and discuss.



Welcome to the Moral Machine! A platform for gathering a human perspective on moral decisions made by machine intelligence, such as self-driving cars. We show you moral dilemmas, where a driverless car must choose the lesser of two evils, such as killing two passengers or five pedestrians. As an outside observer, you judge which outcome you think is more acceptable. You can then see how your responses compare with those of other people. If you're feeling creative, you can also design your own scenarios, for you and other users to browse, share, and discuss.

Fig. 2. Presentation of the MIT's Moral Machine

IV. TEACHING MACHINES TO MAKE HARD CHOICES

The domain of autonomous vehicles is very present in the media, so it is very appropriate to make the Moral Machine project well known throughout the world. But the same type of techniques and way of reasoning developed here to solve moral dilemmas could be applied not only to autonomous vehicles, but to many other different problems: who to select for a job, who to grant parole to, who to choose as a recipient of a transplanted organ, etc. (Of course, one may ask whether ethics consists primarily in the resolution of dilemmas, as if it were the resolution of problems of geometry... but let's leave that now). To address the driving dilemmas they face, the experts try to learn from the answers that ordinary people would give to these questions.

This is the idea. Since we don't know how to explicitly program the code of ethics of an autonomous vehicle, let's ask people: what would you do? In this situation. And in this other one. And so on. Information is extracted from the set of answers, until a set of patterns of behavior is constructed. The Moral Machine project has dealt with an impressive number of respondents: they gathered 40 million decisions in ten languages from millions of people in 233 countries and territories [9].

Something similar is done in other domains: as we don't know how to program a set of explicit rules to recognize a human face, we ask people to recognize faces; we also ask them to distinguish if this one is smiling, if this one is angry, sad or worried. Even body gestures can be recognized in this way: a friendly gesture between friends who have not seen each other for a long time or between two people who close a deal, a threatening gesture... These are all techniques that are already well known and that are working: let us try, then, to apply them to extract moral knowledge from people's responses.

We see a concrete example in Fig. 3: who should be hit by the autonomous vehicle, the woman or the man? And so on with various situations, some more complex than others, but always in the form of a dilemma A vs. B: men vs. women, old people vs. children, cats vs. dogs, children vs. cats, passengers vs. pedestrians, etc.

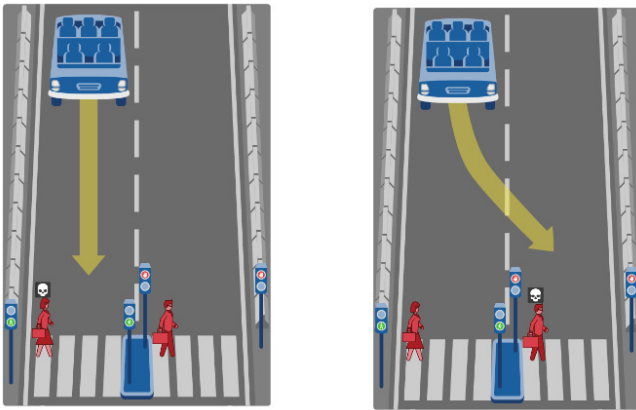


Fig. 3. Illustration of a dilemma of the autonomous cars: who to run over?

Among the little figures that we can select if we want to design our own scenario, there is even a healthcare worker (man or woman) that is easily identified by her case with a white cross (see Fig. 4). Maybe she carries the coronavirus vaccine and by saving her we could save hundreds of thousands of people. Should this influence our decision?



Fig. 4. Is a healthcare worker's life more valuable? Well, this is the option that respondents usually prefer...

V. PROBLEMS OF MACHINE LEARNING

Explainability. Machine learning explores the construction of algorithms that can learn from data, by building a model from example inputs in order to make data-driven predictions or decisions, rather than following strictly static program instructions [10]. What are the problems with this way of addressing ethical questions? First, there is the problem of *explainability*, which is of great concern to researchers. In machine learning (which is only one branch of AI) the end result is a formula, an algorithm, for recognizing the face, the gesture, the writing. But it is not possible to explain why it works. There is no other justification for the formula than its having a very high success rate, its effectiveness (not at all a surprise, since learning algorithms are designed and tested specifically to achieve a high success rate). And of course, when the decision has a heavy ethical load, the fact that one cannot provide reasons for it is a serious, very serious problem.

The lack of explainability is especially severe in the field of neural networks, but it is also present in other AI techniques, such as rule induction (a.k.a. rule inference). Rule induction is a kind of learning technique that process input training data and produce a set of IF-THEN rules to model the desired

behavior. In this case, the training data would be the different situations or dilemmas presented to the respondents, together with their answers (who to run over). It has been argued that rule induction has the advantage over neural networks that the decision algorithm is presented as a decision tree instead of a black box. This human-readable form would improve interpretability and explainability. But in fact, as long as the decision tree becomes as complicated as necessary to cover the maximum number of instances in the training set, it loses all its purported “reasonableness” and “simplicity”. In the end, it is another form of mathematical formula.

Bias. Second, another problem that arises over and over again is the problem of *bias*. Ethical algorithms have to avoid bias, whatever it is. Biases against women, against African Americans, against those who dress unconventionally, etc. But then, what if the population we are surveying in the experiment is biased? We will be reproducing the majority bias, and that is not acceptable: if the majority is biased, it is not enough to imitate it. The bias must be avoided, no matter what the majority says. In other words, the search for equity in social relationships demands that biased judgments are avoided.

This highlights a very interesting point: we know (believe?) that bias, or being biased, is a bad thing, regardless of what the majority says. In other words, *good and evil is not what the majority says*, it is beyond that (see Fig. 5). Moreover, even if certain biases may be beneficial in promoting equity, then the distinction between beneficial and detrimental biases requires a referent beyond the biases themselves. This is not an original consideration: it is at the very origins of ethical reflection in Greek philosophy (see for example Plato's warning against the risk of making choices based on false opinion rather than knowledge, or Aristotle's careful distinction between the common interest of society and the interest of a majority [11]). What is interesting is that AI brings this consideration back into focus.



Fig. 5. A lesson from AI: good and evil is not what the majority says

But then, if it is not what the majority says, *how do we know what is right and what is wrong?* A very difficult problem... We are not going to claim that we have the answer, ready to be explained in three lines. However, we do dare to say that giving up trying to know good and evil in themselves –and this “in themselves” means that they are beyond the opinion of the majority– is a serious limitation for our research, our educational programs, our social and public policies, etc.

Sample selection. And third, closely related to the previous one, is the problem of *sample selection*. The selection of people we ask: who would you run over? That is, who are we asking, ordinary people or sadists? And why shouldn't sadists be included in the survey? Aren't we biasing the sample? And how do we know who are ordinary people, and who are sadists?

If avoiding sadists is a concern, this means that we apply the ethical criteria already in the selection of the sample: it is *a priori*, it is previous to the experiment. Although we do not know it perfectly, we already know, in a way, what is good and what is bad before doing the experiment. That's why we exclude sadists from the sample.

VI. ETHICS, MAJORITIES AND CRITICAL THINKING

In short, our claim is that ethics does not consist in imitating typical or majority behavior. We humans do not teach ethics like that, nor do we want it to be taught like that. If a national or regional government were to include in its educational programs an ethical approach such as the following: "children must imitate the majority"... Would we not rebel with tremendous indignation? The very seed of ethics is critical thinking, non-conformity with dominant thinking, the commitment of one's conscience to recognize for oneself what is right and what is wrong.

Some claim that the ethics of autonomous vehicles should be "customized to the maximum common denominator of the territory where they will be used" [12]. Sure: a different ethics code for choosing who to run over in New York, Johannesburg and Hong Kong? *Can one say more barbarism?* As a technical solution, making a vehicle behave as most drivers would in a given area might be a suitable technological solution. No doubt, a vehicle programmed to operate in Stockholm would be disastrous if driving in Rome, and vice versa, even worse! But please do not call this ethical behavior. It may be effective, but let us not fool ourselves into thinking that this is ethics.

Generally, the terms 'ethics' and 'morality' are used interchangeably, especially in academic contexts [13]. Nevertheless, many distinctions between these two terms have been proposed, even if none has achieved universal consensus. One of the most common distinctions assigns morality a "descriptive" sense, whereas ethics is assigned a "normative" sense. In other words, morality describes social customs or codes of conduct; ethics instead refers to what is *actually* right or wrong, beyond whatever is socially accepted. In this sense, the Moral Machine is certainly effective in its reflecting social customs, but it will never be a true Ethical Machine.

The approach taken by the Moral Machine has been severely criticized, questioning the importance and the real value of finding strong ties between the respondents' preferences and their cultural traits, and even the morality of the whole discussion [14]. Jaques reaches the point of calling the Moral Machine "a monster" [15], because it in fact invites people to express preferences –i.e. biases– for external indicators of *social value*. Not only because humans are very bad at judging people based on appearances; but especially because the Moral Machine gives the impression that this perceived social

value should influence their decision. It suggests and even advocates for the moral relevance of those features – which is nothing less than a direct attack to equity. Etienne [16] discusses the dangers of the Moral Machine experiment, alerting against both its uses for normative ends and the whole voting system approach it is built upon to address ethical issues. According to Puri [17], even if moral behavior can be imitated by algorithms, only conscious agents can take moral decisions and assume responsibility for them. From a more general standpoint, the Proceedings of the IEEE Special Issue on machine ethics contains several papers that emphasize the protection due to human values in future autonomous systems; especially significant here is Bremner et al. [18], which argue for transparent and verifiable ethical reasoning in AI systems.

Ethics is not an imitation game. Ethics is not about following a code of conduct like Asimov's Three Laws. But neither is ethics about imitating the behavior of others. Learning by imitation is something very human, but if there is any difference between #StayHome and #CollectToiletPaper, that difference is not in the number of people who behave one way or another, but in the *reasonableness* of their behavior. Ethics is not spread like a virus; it is learned and discussed rationally.

VII. THE RATIONALITY OF ETHICS: EQUITY AND EQUALITY

So, one of the obstacles we have to overcome in order to recover sanity is the widespread skepticism in our society about the rationality of ethics and values [19], which is also manifested in the existence of multiple incompatible ethical theories [6][20]. Emotivism –not a healthy integration of emotions with reason in moral judgment, but a replacement of the latter with the former– is promoted in approaches such as the Moral Machine, where the respondents' *reasons* for their answers are neglected; only the raw answers are counted, as if to say: "ethics is not rational in itself, so we must be content to reflect what most people say". As we have shown, this is radically insufficient to avoid biased judgments and ensure equity.

The role of emotions in moral decision-making is well established [21], but that does not mean that we can reduce ethics to emotion, as emotivism does. In a cultural context where post-truth is blowing up current knowledge structures [22], an algorithm that produces an appearance of consensus in emotional reactions of a large mass of people, but without valuing arguments, runs more risk of being manipulated, biased or intoxicated than before the emergence of social networks. The good news is AI is forcing us to think ethics in a new way.

One of the most frequently heard forms of ethical argumentation is: "if everyone acted like that, then such a thing would happen" (which is clearly a popular derivation of Kant's categorical/universal imperative). This is what has been called rule-based consequentialism, that is, making ethical assessments according to the consequences of the application of certain rules of behavior, not so much of concrete acts [23][24]. Without denying the value that this type of argumentation has in illuminating our ethical reason, we cannot be satisfied with it being the only type of acceptable ethical rationality. The rational limitations of consequentialism

are clear and have been denounced long ago; here is our version of the refutation [25]:

Be it welcome!

First, the consequences of a certain action extend over a period of time that properly has no limit, yet we cannot indefinitely wait to judge whether an action is good or bad. Second, even if we put a timely boundary to the consequences we want to consider, they nevertheless belong to the time to come, therefore they are rather uncertain; we would have to employ some kind of prediction technique to foresee the consequences and value them; but these techniques will always be limited by the very nature of things, which do not follow perfectly known behavior rules (besides, consequences will probably depend on the freedom of others). Third, and most important, if we want to avoid a priori rules to determine the goodness for actions, and we make the goodness of an action depend on the goodness of its consequences, then we need rules to value the goodness of the consequences; extreme consequentialism does not solve the problem of goodness, but simply puts it off.

The Moral Machine asks you to “judge which outcome you think is more acceptable”. It processes your answers, together with those from millions of respondents, and produces a model of behavior. But learned ethics and programmed ethics are different only at first view, i.e. if we consider only the way the behavior rules have been produced. Actually, both are the same kind of thing: a program that makes choices based on rules aimed at achieving the “best” outcome. A program made of rules that come either from *a priori* speculation or from imitation of actual patterns of human behavior. Not that big of a difference.

The problem is that an approach based on computing outcomes and the reduction of ethics to the compilation and application of a set of rules, either *a priori* or learned, is a severe misconception of ethics. Above all, the attempt to formalize ethics in a set of rules misses the point that *a person is not only an instance of a case*, but a unique and unrepeatable being. A person is a child, a student, a patient, a client, a neighbor (*your child, your student, your patient, your client, your neighbor*). A person’s value cannot be measured with a number, not even with an array of numbers (age, sex, health status, social contribution...). When it comes to people, our reason needs to be trained to understand what we see not only as cases of a general rule, but above all in their uniqueness. We need to recover value rationality beyond numbers, beyond logical and instrumental rationality. We need to learn how to reason with values in a way that does not convert them in numbers.

Ethics should prevent us from the error of converting equity into mathematical equality, achieved through the extraction of characteristics and the computation of a value formula. Equity is not mathematical equality, not even a weighted equality that considers different factors. The first commandment of ethics should be “thou shalt not treat a person as *a vector of numbers*”. Does this demand a renewed rationality of ethics?

REFERENCES

- [1] F.H. Hinsley, A. Stripp (eds.), *Codebreakers: The inside story of Bletchley Park*, Oxford: Oxford University Press, 1993.
- [2] A. Hodges, *Alan Turing: The enigma*, London: Burnett Books, 1983.
- [3] A.M. Turing, “Computing Machinery and Intelligence”. *Mind* 59:433–460, 1950.
- [4] D. Vanderelst, A. Winfield, “An architecture for ethical robots inspired by the simulation theory of cognition”. *Cognitive Systems Research* 48:56–66, 2018.
- [5] S.L. Anderson, “The unacceptability of Asimov’s three laws of robotics as a basis for machine ethics”. In M. Anderson & S. L. Anderson (eds.), *Machine ethics*, pp. 285–296. Cambridge: Cambridge University Press, 2011.
- [6] V. Nallur, “Landscape of Machine Implemented Ethics”. *Science and Engineering Ethics* 26(5):2381–2399, 2020.
- [7] S. Lumberras, “The Limits of Machine Ethics”. *Religions* 8:100, 2017.
- [8] J. Torresen, “A Review of Future and Ethical Perspectives of Robotics and AI”, *Frontiers in Robotics and AI* 4:75, 2018.
- [9] E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon, I. Rahwan, “The Moral Machine experiment”. *Nature* 563:59–64, 2018. See also the online platform for the moral machine experiment, available: <https://www.moralmachine.net/>.
- [10] C.M. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer, 2006.
- [11] J. Ober, “Democracy’s Wisdom: An Aristotelian Middle Way for Collective Judgment”. *American Political Science Review* 107(1):104–122, 2013.
- [12] C. Leonard, “Teaching ethics to machines”, 2016. [Online]. Available: <https://www.linkedin.com/pulse/teaching-ethics-machines-charles-leonard>.
- [13] B. Gert, J. Gert, “The Definition of Morality”, *The Stanford Encyclopedia of Philosophy*. [Online]. Available: <https://plato.stanford.edu/archives/fall2020/entries/morality-definition>.
- [14] A.M. Nascimento, L.F. Vismari, A.C.M. Queiroz, P.S. Cugnasca, J.B. Camargo Jr., J.R. de Almeida Jr., “The Moral Machine: Is It Moral?”. *2nd International Workshop on Artificial Intelligence Safety Engineering (WAISE 2019)*, within *38th International Conference on Computer Safety, Reliability, and Security (SAFECOMP)*, September 10–13, 2019, Turku, Finland. Lecture Notes in Computer Science, vol. 11699, pp. 405–410.
- [15] A.E. Jaques, “Why the moral machine is a monster”. *We Robot Conference*, University of Miami School of Law, April 11–13, 2019.
- [16] H. Etienne, “When AI Ethics Goes Astray: A Case Study of Autonomous Vehicles”. *Social Science Computer Review* 40(1):1–11, 2020.
- [17] A. Puri, “Moral Imitation: Can an Algorithm Really Be Ethical?”. *Rutgers Law Record* 48:47–58, 2020.
- [18] P. Bremner, L.A. Dennis, M. Fisher, A.F. Winfield, “On Proactive, Transparent, and Verifiable Ethical Reasoning for Robots”. *Proceedings of the IEEE* 107(3):541–561, Special Issue on Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems, 2019.
- [19] G. Génova, M.R. González Martín, “Teaching Ethics to Engineers: A Socratic Experience”. *Science and Engineering Ethics* 22(2):567–580, 2016.
- [20] K. Bogosian, “Implementation of moral uncertainty in intelligent machines”. *Minds and Machines* 27(4):591–608, 2017.
- [21] J. May, C. Workman, J. Haas, H. Han, “The Neuroscience of Moral Judgment: Empirical and Philosophical Developments”. In F. de Brigard, W. Sinnott-Armstrong (eds.), *Neuroscience and Philosophy*. MIT Press (forthcoming).

- [22] S. Sismondo, "Post-truth?". *Social Studies of Science* 47(1):3-6.
- [23] D.G. Johnson, *Computer Ethics*, 2nd Ed, Upper Saddle River, NJ: Prentice Hall, 1994.
- [24] K.C. Laudon, "Ethical Concepts and Information Technology". *Communications of the ACM* 38(12):33-39, 1995.
- [25] G. Génova, M.R. González Martín, A. Fraga, "Ethical education in software engineering: responsibility in the production of complex systems". *Science and Engineering Ethics* 13(4):505-522, 2007.



Gonzalo Génova is Associate Professor at the Computer Science and Engineering Department, Universidad Carlos III de Madrid. He has an MS degree in Telecommunications Engineering (1992), an MS degree in Philosophy (1996) and a PhD in Computer Science and Engineering (2003). His main research subjects are models and modeling

languages in software engineering and requirements engineering, which he has combined with interdisciplinary research in the philosophy of information systems, particularly on artificial intelligence and professional ethics. He can be reached at ggenova@inf.uc3m.es.



Valentín Moreno Pelayo is Associate Professor at the Computer Science and Engineering Department, Universidad Carlos III de Madrid. He received his undergraduate degree in Mathematical Sciences in the branch of Computer Science from Universidad Complutense de Madrid. He obtained his first doctorate in

Archives and Libraries in the Digital Environment (2010) and a second doctorate in Sciences and Information Technology (2016), both from Universidad Carlos III de Madrid. He has experience in organization and recovery of knowledge using Artificial Intelligence techniques, Data Mining and Documentary Relevance. He can be reached at vmpelayo@inf.uc3m.es.



M. Rosario González Martín is Associate Professor at the Department of Educational Studies, Universidad Complutense de Madrid. She has a degree (1996) and a PhD (2004) in Philosophy and Educational Sciences. She is the Director of the Research Group on Civic Culture and Educational Policies of the Universidad Complutense de Madrid. Her lines of work

are fundamentally the Philosophy and Ethics of Education, mainly from a phenomenological and personalistic approach. She can be reached at marrgonz@ucm.es.