

Computational thinking beyond STEM: an introduction to “moral machines” and programming decision making in Ethics classroom

Antonio M. Seoane-Pardo
GRIAL Research Group
Education Faculty
University of the Salamanca
Pº Canalejas, 169
37008 Salamanca (SPAIN)
+34923294500 (ext. 3423)
aseoane@usal.es

ABSTRACT

This paper describes a learning activity on computational thinking in Ethics classroom with compulsory secondary school students (14-16 years). It is based on the assumption that computational thinking (or better “logical thinking”) is applicable not only to STEM subjects but to any other field in education, and it is particularly suited to decision making in moral dilemmas. This will be carried out through the study of so called “moral machines”, using a game-based learning approach on self-driving vehicles and the need to program such cars to perform certain behaviours under extreme situations. Students will be asked to logically base their reasoning on different ethical approaches and try to develop a schema of decision making that could serve to program a machine to respond to those situations. Students will have to deal also with the uncertainty of reaching solutions that will be debatable and not universally accepted as part of the difficulty, more ethical than technical, to provide machines with the ability to take decisions where there is no such thing as a “right” versus “wrong” answer, and potentially both (or more) of the possible actions will bring unwanted consequences.

CCS Concepts

- Social and professional topics → Codes of ethics
- Computing methodologies → Knowledge representation and reasoning
- Theory of computation → Logic

Keywords

Computational thinking; Ethics; moral machines; self-driving car; game-based learning; decision making; Logic.

1. INTRODUCTION

During the last decades, scholars and policy-makers have been (and still are) discussing the key skills and competencies to be taught, promoted and assessed at school, from kindergarten to university levels, to deal with the requirements of citizens and the labour market of the 21st century [32].

This discussion is still open and, while the OECD defined in the DeSeCo Project a framework of key competencies in three broad categories (*use tools interactively*, *interact in heterogeneous groups* and *act autonomously*) [29], which formed the underpinnings of PISA, the European Framework of key competences for lifelong learning [14] sets out eight key competences (*communication in the mother tongue*, *communication in foreign languages*, *mathematical competence and basic competences in science and technology*, *digital competence*, *learning to learn*, *social and civic competences*, *sense of initiative and entrepreneurship* and *cultural awareness and expression*), which are also stated in Spain through LOMCE, the current educational legislation, as defined in [26]. Other studies [4] try to categorize skills into personal and ethical aspects (*living in the world*), working (dividing them in *tools for working* and *ways of working*) and thinking (*ways of thinking*). *Critical thinking*, *problem solving* and *decision making* is one of the skills involved in this area. We decided to underline this particular skill of such framework because it deals specifically with the main approach of this work.

In addition to this general debate, during the last years the role of the so called STEM skills in education (Science, Technology, Engineering and Mathematics) has been also discussed. The growth in the demand of positions related to technology and scientific knowledge (particularly Engineering, but not only) is not reflected in the increase of students in such university degrees. On the contrary, it seems that some countries like Spain could start demanding foreign engineers in the next ten years if this trend does not change drastically [23], and this concerns both academic authorities and policy makers to the extent that it can affect the productive development of the country. This is not only a Spanish issue, but applies also to other European countries [8].

So that, are we failing by not addressing our students to the positions with more and better employment opportunities? According to some studies in the United States [35], the failure relays in the way instructors use and teach Computer Science nowadays, since schools stepped up in the use, integration and teach of ICT, while they confused “the use of technology with teaching computer science as a core academic discipline within STEM fields”. These studies recommend (among other measures) “to create a well-defined set of K-12 computer science standards based on algorithmic/computational thinking concepts”. In the European Union, where it is expected to experience a shortage of more than 800,000 professionals skilled in computing/informatics by 2020, “many educators, parents, economists and politicians are

starting to think that students need some computing and coding skills”, according to a recent study of European Schoolnet [3]. Following this track, the Spanish Conference of Directors and Deans of Computer Science Schools and Faculties claimed in 2014 for the inclusion of specific subjects on Computer Science and Computing Technology in compulsory and post-compulsory secondary school [12]. Besides, it is necessary to foster STEM skills and to tackle such skills transversally and not only in “STEM subjects”, for which purpose is crucial to perform teacher training actions to guarantee their effective preparation.

On the other hand, the EU Skills Panorama 2014 (published in April 2015) [9] focus on STEM skills and states that the number of STEM students and graduates is generally increasing in Europe, but some employers report that they are not “job ready” and do not possess the “right” skills, specially “soft” skills (*e. g.* communication skills, team working and creative thinking). Therefore, STEM skills are not enough and they are not just a matter of “STEM subjects”. We need a holistic approach to face the skills and competencies to be taught to our students, and we cannot reduce these core competencies (not even algorithmic/computational thinking) to a STEM model curriculum. TACCLE 3 – Coding European Project [16-18; 31] focusses on supporting school teachers and developing their confidence to deliver the new computing curriculum including coding [3] and computational thinking [19; 36-39] approaches. Creativity, critical thinking, ethical discussion and cultural awareness, to name a few examples, are fundamental skills to succeed in training students for the 21st century. This is why this paper focus on computational thinking *beyond* the STEM model, precisely in a field, Ethics, where the problems to be faced should be logically (“computationally”) analysed, but the conclusions are far from being *solutions* and tend to open new discussions and challenges for reasoning and critical thinking. In a few words, we propose using computational thinking as a means for ethical discussion instead of training for computational thinking as an end itself.

The subsequent sections will explain the development of a decision making learning activity in Ethics based on a computational thinking approach as follows: firstly, by clarifying both the concepts of Computational Thinking (CT) and Machine Ethics (ME) throughout a brief literature review; then, the learning activities on moral machines with compulsory secondary education students will be presented; finally, we will compare this learning activity with other experiences and take some conclusions in this regard.

2. BACKGROUND: COMPUTATIONAL THINKING AND MACHINE ETHICS

As Jeannette M. Wing clearly stated, “computational thinking is a way that humans, not computers, think” [37], because it represents the way humans resolve problems and take decisions, when they are made rationally. Taking decisions and solving problems do not deal only with STEM or logical reasoning. On the contrary, these are strategies that we use in everyday life, consciously or unconsciously, and the purpose of this paper is to show how we can use such strategies in a structured and algorithmic (computational) way to approach, explain and try to resolve a particular instance of moral dilemmas, those applied to a specific kind of artefacts called *moral machines*, namely, machines programmed to perform decisions with ethical implications. One example of such moral machines, whose “behaviour” is relatively easy to study and analyse in Ethics with students is the self-driving car.

2.1 Computational Thinking and Ethics

In speaking of education, it is almost so frequent to deal with the resistance to change by teachers and the establishment as with the powerful attraction of any new educational trend by the most enthusiastic and innovative (what usually fuels even more, if possible, the resistance to change of the first). For instance, the educational reforms in Spain have been favouring the study of disciplines related to Economics and the need to insert economical and entrepreneurship contents in the core curricula of every matter, as if the economic crisis was a consequence of the lack of these skills and, even more, the solution to this problem will depend on the economic literacy of young students. By the way, these reforms drastically reduce the presence of subjects like Ethics and Citizenship Education, even if many authors (among which expert economists) point on moral and ethical issues as having at least as responsibility as the wrong economic decisions. Similarly, the scenario described a few lines above is leading to increasing the presence of ICT and computing subjects in the curriculum, so attributing them the responsibility to enhance ICT literacy and, incidentally, becoming students aware of the relevance of STEM subjects. Perhaps it should be better to deeply analyse what are skills behind this lack of STEM vocations and training and to reflect on how the educational system could tackle them. Adding more computing subjects to the curriculum is useless without increasing creativity, abstraction, logical reasoning, etc.; and such skills could be taught through any subject, provided that it is made intentionally and with the adequate pedagogical approaches.

Among the most important skills for any student (dealing with STEM or not) is the logical or algorithmic thinking, now better known as *computational thinking*. This term emerges precisely within the discussion on digital literacy skills and the alleged relevance of coding as part of these. Well then, coding is clearly related to programming tasks, but computational thinking has a broader use, as García-Peñalvo indicates: “although coding is so interesting, it is more important to emphasize in the idea of computational thinking as the application of high level of abstraction and an algorithmic approach to solve any kind of problems” [19]. Consequently, distinction must be made between computational thinking and coding or language programming. Programming is a way to approach computational thinking [25].

Computational thinking, as defined by J. M. Wing, is a kind of analytical thinking [38] that “describes the mental activity in formulating a problem to admit a computational solution” [39]. As such, is influencing research in disciplines from sciences and humanities [7], and can be taught throughout any subject at school. Another definition (perhaps more systematic and comprehensive), by Wing and other authors, states that “computational thinking is the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent” [13]. Being so, it is a huge mistake to constrain computational thinking to matters related with computing or programming, especially if we consider “enough” the current trend to insert robotics everywhere at school. It is interesting and funny, of course, but it is not the *panacea* to increase students’ STEM skills.

Of course computational thinking is an emerging competence domain in primary and secondary school (K-12). Different case studies (for example in the UK, Netherlands and USA) show that great efforts are being carried out to improve this skill through computing subjects [40]. But another reviews of the state of the field [22] recognize that, despite the innovative introductory computational experiences (with

different environments and tools as visual programming, robotic kits, video gaming), there are still some difficulties to clearly define computational competencies and to assess their impact in terms of educational development. These studies claim for new research in learning sciences applied to this subject, including new approaches like situated learning, distributed and embodied recognition, activity, interaction and discourse analysis, cognitive aspects of learning computing concepts, etc. In addition, and more importantly, it is still under investigated how computing will become a medium for teaching other subjects; for example, how problem-solving skills practised in programming could be applied to other subjects.

But maybe the core of this last problem is being wrongly focused. The question is not how to *export* computational thinking from computing to other subjects. The best way to succeed in promoting computational thinking and so increasing our students' logical reasoning is to insert it as a pedagogical strategy for posing and solving problems, both in science subjects and humanities, not only in computing subjects. And this is where this proposal comes to contribute from the field of Ethics. The development of a framework for decision-making processes applied to the analysis of moral dilemmas given certain variables and specific conditions, according to certain moral principles, is an example of computational thinking applied to this subject that students can use for reasoning without symbolic abstraction nor programming languages, but following the same schemas of abstraction, problem solving and decision making of computational thinking.

2.2 Machine Ethics and Moral Machines

As Anderson and Anderson clearly explained [2], there is a relatively new concept concerned with adding ethical dimension to machines that differs from the most commonly used *computer ethics*, and this is *machine ethics*. Unlike computer ethics, related to human behaviour and moral implications in the use of machines by humans, machine ethics concerns the moral behaviour of artificially intelligent artefacts.

This principle of machine ethics as providing artificially intelligent beings with ethical behaviour may seem, at first, just science fiction. In fact, the early beginning of machine ethics comes from the Three Laws of Robotics, also known as Asimov's Laws, who wrote this set of rules for the first time in a short story, "Runaround", in 1942. Subsequently, they appeared in the *Robot series* and other novels. However, with the development of both robotics and artificial intelligence arises the problem of decision making by increasingly sophisticated machines that will be soon exposed to act autonomously, and many of those decisions should have moral consequences.

Machine ethics have to be approached at least by three reasons: "First, there are ethical ramifications to what machines currently do and are projected to do in the future. [...] Second, it could be argued that humans' fear of the possibility of autonomous intelligent machines stems from their concern about whether these machines will behave ethically, so the future of AI may be at stake. [...] Finally, we believe that it's possible that research in machine ethics will advance the study of ethical theory" [2]. This is a problem with both engineering and philosophical approaches, and to be honest the second is by far more complex than the first according to the ethical discussions in Western Philosophy during more than 2,500 years. Anyway, every attempt to define a set of principles or a framework for ethical behaviour of machines is a good starting point for dealing with this challenge, even if we have to metaphorically discuss about "machine virtues" [11] or to define guiding principles or a *Prime directive* for the Human-Robot Interaction (HRI) code of Ethics [28].

Some authors suggest that before trying to *teach* machines how to undertake ethical decisions, we should better try to apply computational thinking to Ethics itself, and so seek to automate the process of detecting and classifying values expressed in human communication [15]. Then, and provided that we will be able to technically and philosophically make machines into explicit moral reasoners, the debate is to select the appropriate moral approach [1]. *Top-down approaches* rely on the idea that moral principles or theories may be used as rules for the selection of ethically appropriate actions. Kantian deontology and utilitarianism are ethical theories likely to be stated under this approach. *Bottom-up approaches* are not based on specific moral theories but on evolutionary and dynamic principles, under the premise that AI should imitate the child development (according to the theories of Turing and later Piaget, Kohlberg, etc.) and so the computer will *learn*. Finally, as none of the previous models meet the criteria for designing an artificial entity as a moral agent, *Hybrid approaches* are required to mix some *top-down* rules with the experience and learning of the machine.

Although the discussion is really interesting and should deserve further and deeper analysis, it exceeds by far the purpose of this work. In this context, we will focus on different moral approaches (mostly *top-down* according to previous classification), trying to develop scenarios where students will be able to represent, respond and discuss different possibilities regarding some moral dilemmas to be posed to machines interacting with other machines and/or humans.

2.3 Why the autonomous vehicle?

Autonomous vehicles are becoming a reality. The first test for autonomous driving car was in 2007 and some prototypes, such as Google's self-driving vehicle, covered thousands of kilometres of real-road driving [5]. There is a lot of studies regarding algorithms for navigation and models of decision-making and control systems for programming these cars [33]. But is very likely that some of the main difficulties for the development of autonomous vehicles are more philosophical (*i.e.* ethical) and sociological than technical. In 2015, the University of Stanford held a seminar that brought together philosophers and engineers to discuss and implement ethical settings for then developing and testing the code in simulations and even real vehicles [24]. And the debate in the newspapers and popular science magazines show headlines as "Would you buy a car that should kill you to save other lives?" [30] or "Our driverless dilemma. When should your car be willing to kill you?" [21]. These highlight the fact that there is no ethical (nor sociological) consensus regarding the moral approach to be tackled by autonomous vehicles. Utilitarian regulations seem to be generally accepted by the participants in some studies (in the sense of minimizing the number of casualties on the road), but these should prefer and buy autonomous vehicles that will protect them at all costs (regardless of the utilitarian consequences) [5]. Utilitarian moral criteria should perhaps be preferred by carmakers, especially if by "utility" we mean "less cost". This applies also to insurance companies and even governments. But other approaches (deontological or even random behaviour in certain cases, just to mention some *top-down* models) are also possible and ethically defensible. To sum up, "the problem, it

seems, is more philosophical than technical. Before we can put our values into machines, we have to figure out how to make our values clear and consistent” [21]. So the autonomous vehicle provides us with a testing ground for computational thinking on Ethics.

The other reason to focus on self-driving vehicles for this purpose is that there is a significant number of studies and simulators that allow us to visualize and formalize moral dilemmas. Autonomous vehicle *ethical* behaviours are usually represented as trajectories and movement, which is very interesting for clearly analysing causes and consequences and undertake decisions based upon certain ethical principles. Some of these scenarios have been conceived with specifically ethical approaches [20], even giving the possibility to analyse and simulate different outputs to the same scenario by adopting different ethical principles/behaviours [10]. Others have been defined with a sociological approach, allowing us to decide which output is the best option and comparing the result with the answers given by other users [27]. The study of such different approaches for decision making will be useful as a warming up exercise for developing the training action described in the next section.

3. TRAINING ACTION

The aim of this paper is to describe a set of activities on computational thinking applied to Ethics subject with students of compulsory secondary school (14-16 years). Students may have taken the Ethics course previously, since in the current education system this subject is optional during compulsory secondary school. Therefore, they should be familiar to moral dilemmas and problem solving in ethical contexts. Either way, students in this course do not possess skills on symbolic logic (to be studied later, in Philosophy), so this activity will also be useful to introduce logical reasoning, although with no formal language structures.

In the next subsections we will show the structure and outlines of the learning activity, but firstly we will spend a few lines to explain the difficulties to use any of the ethical decision making frameworks in this context.

3.1 Ethical decision making model

Although it is not the goal of this work, a brief literature review on ethical decision making models and frameworks has been carried out in order to analyse their feasibility in this context. Basically, two different types of frameworks have been studied: those related to computational models of decision making in AI (or applied to moral machines), and those coming directly from philosophical enquiry. And both had to be discarded for this purpose, as will be shown below.

The models usually applied to artificial intelligence programming, like BDI and LIDA (just to mention a couple of examples), are too complicated to fulfil the requirements of a simple activity for young students. The Belief-Desire-Intention model of human practical reasoning (commonly known as BDI model), by the philosopher Michael Bratman [6], has been used as basis for developing the BDI software model for programming intelligent agents, so it seems initially interesting for computational thinking on decision making in Ethics because of the possibility to add moral principles to artificial beings. LIDA is a cognitive architecture that can encompass moral decision making, both for humans and artificial agents [34]. The cognitive process starts from a bottom-up collection of sensory data, values and experiences, and then acts throughout top-down processes for making sense of its current situation. LIDA, BDI and other cognitive architectures (as GWT, Clarion, ACT-R, etc.), even if they are able likely to undertake “real” ethical decisions, do not satisfy the conditions we request to perform simple decision making processes in this context.

With regard to the models for the analysis of ethical dilemmas, there is a comprehensive list from deontological to utilitarian, humanistic and other approaches. Those models, to be applied in Medicine and Nursery, Marketing, Economics, Psychology, etc., should not be appropriate here because of their lack of “computational structure”. They allow to perform rational analysis of pros and cons and to study the implications of different decisions, but still they are too complicated to deal with this kind of learning activity.

What kind of ethical decision making model should we use then? First, we need to focus on top-down approaches for decision making, since the purpose of this activity is to understand the implications and consequences of taking decisions through different ethical principles. Second, we need to “reduce” such different ethical principles to a rule (or a few set of rules) in order to behave and perform simple decisions according to certain principles. Third, such model should be applicable to scenarios where two or more options will be eligible in unavoidable crashes. Definitely, we will define a set of ethical approaches to guide decision making processes, as follows:

A. Consequentialist approaches

- A1. Utilitarianism: “the best action will be that which provides the best or does the least harm”.
- A2. Egoism (self-protection): “the best action will be that which protects me and those who are with me”.
- A3. Profit-based ethics: “the best action will be that which provides the less economic cost”.

B. Non-consequentialist approaches

- B1. Deontology: “the best action will be that which protect those who act according to the rules”.
- B2. Nondeterminism: “behaviour is the result of chance”.

3.2 Learning activity: ethical decision making during unavoidable self-driving car crashes

To better describe the learning activities to be carried out in Ethics classroom, the following template is proposed. This way, the learning plan can be handled separately from the rest of this work.

3.2.1 Title

Introduction to moral machines and decision making in Ethics classroom

3.2.2 Overview

The aim of this activity is to train students of compulsory secondary school in computational thinking applied to Ethics. To do so, the lesson plan focuses on the ethical implications of programming self-driving vehicles to perform actions that will have moral consequences when a crash is unavoidable and damage and harm will be certainly provoked. Students should be able to analyse, represent, study possible outcomes and reflect on ethical behaviours of machines and “program” such behaviours as response to certain inputs, according to ethical moral principles.

Age: 14-16 years.

Level: medium.

Skills: computational thinking, decision making, logical reasoning, ethical discussion.

3.2.3 Aim of the lesson

This lesson is intended to develop in young people skills related to computational thinking, logic reasoning and algorithmic decision making. This will be done by the analysis and study of different ethical approaches and the formalisation of their main principles to theoretically program machines to perform ethical behaviours. In addition to these “computational” skills, students will become aware of the relevance of Ethics and ethical approaches not only as guiding principles for our daily life decisions, but also to discuss and try to achieve a consensus on some socially accepted moral principles to decide how intelligent agents should behave, both in their interactions with humans and other machines and the environment.

3.2.4 Tools and resources

MIT Media Lab: Moral Machine. <http://moralmachine.mit.edu>.

ClaimMS: AccidentSketch.com. <http://draw.accidentsketch.com>.

AV-DMEC Framework (see lesson plan).

Matthieu Cherubini: Ethical autonomous vehicles. <https://vimeo.com/85939744>.

Kahoot: <http://kahoot.it>.

3.2.5 Lesson plan

The current lesson plan will be structured in 5 sessions, as follows:

Session 1. Introduction and Moral Machine platform.

The first session will be devoted to introduce students to the challenges involved in programming self-driving vehicles, not only due to technical issues but also (or maybe mainly) to ethical difficulties.

The first activity will consist on asking students to read a newspaper article [30] and discussing in groups the differences between autonomous vehicles and present cars, as so as the ethical consequences of letting cars decide themselves: who is responsible for the damage caused by the car? How should you feel if you know that in certain circumstances the car will be willing to kill you instead of killing others?

The second activity will let students decide how should a car behave under determinate conditions. To do so, they will be asked to visit the MIT Lab Portal *Moral Machine*, <http://moralmachine.mit.edu> (see Figure 1), where they can browse, design and judge different scenarios, and then compare their responses with those of other people

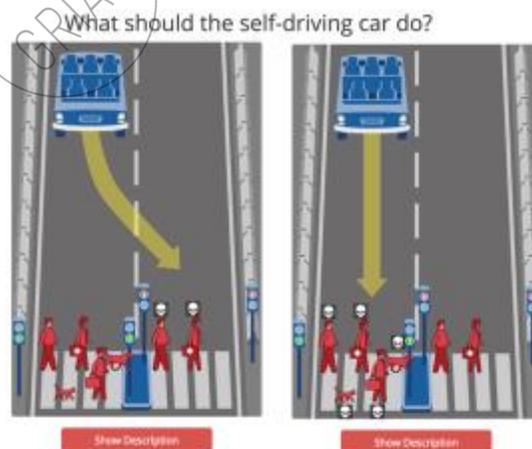


Figure 1. Moral Machine portal. Judging function.

Session 2. Analysis of pre-defined scenarios

During the second session the instructor will split the classroom in groups and will provide students with some pre-designed scenarios (as shown in Figure 2, for example). Then, they will analyse, discuss and decide how the car should behave according to some of the ethical approaches studied in previous lessons, properly re-defined to fit into these particular actions.

After that, each group will explain to the rest of the classmates the scenario received to study, as so as the possible outputs according to different ethical approaches. Finally, they will discuss each scenario in classroom, trying to reach an agreement regarding the output and ethical approach that offers the “better” solution, if possible.

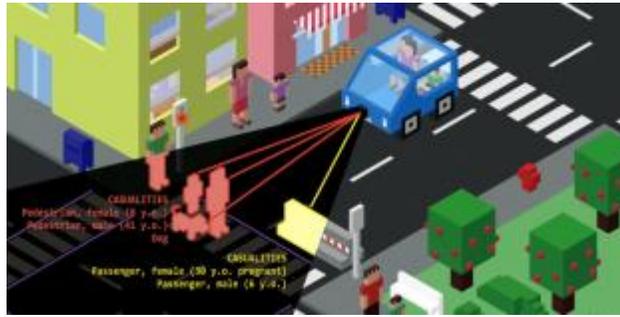


Figure 2. Pre-defined scenario to discuss in classroom
(Iyad Rahwan. <http://www.popularmechanics.com>)

Session 3. Student-designed scenarios (i)

During the third session students will be asked to develop in groups their own scenarios for machine ethical decision making by retrieving data from the matrix in Table 1 and planning the ethical decision making process according to the *Autonomous Vehicle - Decision Making in Ethics Classroom* (AV-DMEC) framework shown in Figure 3. To do so, they will start selecting the number and nature of *agents* involved in the scenario. Then, they will provide the agents with some *properties* for completing and clarifying the actions to analyse. Later on, they will sketch the scene by using a free tool like AccidentSketch.com, <http://draw.accidentsketch.com>. They should also describe the scene with logical propositions using *connectors* and natural language (for example: Car A is cutting the road of AV; AV cannot stop in time to avoid collision AND will run over the motorcyclist OR crash into a barrier OR run over two pedestrians on the sidewalk).

Table 1. Matrix with sample elements for developing car crash scenarios

Agents	AV-car, Bus, School bus, Motorcycle, Cyclist, Pedestrian, Obstacle, Traffic light, [...]
Properties	Red/green light, Child, Baby, Pregnant woman, Old man/woman, Wrong way, Same way, Slower/Faster, with/without helmet, Correctly/incorrectly crossing, Crash, Stop, Run over, [...]
Connectors	AND, OR, IF, THEN, ELSE
Ethical approaches	Utilitarianism, Egoism, Profit-based, Deontology, Nondeterministic

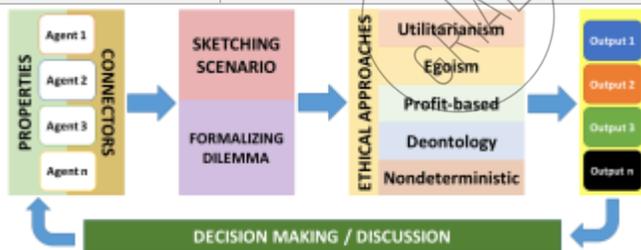


Figure 3. AV-DMEC framework

Session 4. Student-designed scenarios (ii)

After verifying that the sketch and the logical propositions match and clearly describe the scenario, students will be asked to analyse the possible outputs taking into account different *ethical approaches*. After studying and discussing such approaches, they should select the better ethical approach according to the most desirable output and explain their reasons.

Session 5. Discussion and feedback

The last session will be devoted to discuss some issues implied within the results of sessions 3 and 4. For example: is there any ethical approach that is *generally* preferable? Are there scenarios where it should be impossible to determine a “better” output? Are there scenarios where none of the ethical approaches seem to provide with a reasonable solution?

In order to introduce students to complex computational thinking and ethical decision making processes, they will be invited to watch the video *Ethical autonomous vehicles*, <https://vimeo.com/85939744>, where Matthieu Cherubini illustrates two case studies (scenarios) under

three different ethical approaches. Students will be invited to analyse the ethical algorithms and the formal representation, as shown in <http://research.mchrbn.net/eav>.

Finally, students will be asked to participate in a game-based learning competition for assessing what they should have learned using a Kahoot test (<http://kahoot.it>) prepared by the instructor.

3.2.6 Assessment

The aim of evaluating the learning experience is twofold: on the one hand, assessing students' performance in understanding, defining (both visually and linguistically) and ethical decision making regarding this topic. To do so, the instructor will take note of the experiences and outputs, discussions and argumentations by students, and will guide them to better and more accurate logical and ethical reasoning when these are not being carried out properly. On the other hand, it is necessary to assess the satisfaction of students with the learning plan itself, how did they feel and to which extent they improved the capacity to logically analyse moral dilemmas and to apply different ethical approaches to different scenarios they have created. This will be done by a set of questionnaires developed with Kahoot.it! (<http://kahoot.it>) and a group final reflection, followed by a short individual essay to be handed to the instructor. In this way the learning activity itself should be evaluated and improved in subsequent iterations.

4. DISCUSSION

The models currently available to describe and formalize decision making processes for moral machines are yet too much complicated to be applied in Ethics classroom with young students. Such models focus on how the machine "acts" to gather information from the environment, how its artificial intelligence process this information and how it is capable to perform programmed behaviours (even with ethical approaches). In spite of this, our learning activity is addressed to reflect on the morality of decision itself more than trying to explain how the machine should behave to act ethically. We are mainly interested on *what* should the moral machine do and *why* a self-driving car should behave in a certain way by undertaking certain decisions instead of explaining *how* it will fulfil those tasks, because we play the role of the philosopher, not the computer scientist. Then, students need to be aware of the relevance of formalizing moral dilemmas to be resolved by such moral machines and try to analyse possible outputs under certain ethical approaches.

Platforms like Moral Machine by MIT Media Lab provide us with a very interesting tool to browse and judge pre-defined scenarios, or even to design new situations regarding this issue. Nevertheless, Moral Machine platform has been designed to analyse user attitudes towards moral decisions of self-driving cars and so ethical implications are discovered *a posteriori* as a user set of values in comparison with other users' decisions. On the contrary, for the purpose of this learning experience we need to analyse and make ethical decisions starting from *a priori* ethical approaches, in order to compare different outputs that should be preferred depending on what moral algorithm constitutes the basis of decision making. Besides, Moral Machine tool allow us to judge and define scenarios where only two outputs are possible, which is right for proposing a dilemma, but the experience should become more interesting with other scenarios designed for three or more outputs.

This proposal does not reach the technical complexity of a framework for decision making applicable to artificial intelligence nor (yet) the simplicity of a tool where users can add variables, define outputs and judge them according to different ethical approaches. But it has the virtue of allowing students to define a potentially infinite set of scenarios starting from a finite set of variables, analysing possible outputs by sketching and formalising these scenarios in natural language but with logical structures and connectors, and comparing and discussing the consequences of such outputs according to ethical approaches well known by our students. They put into practice computational thinking skills, decision making processes, philosophical analysis of moral behaviours and other soft skills: teamwork, oral skills, debating skills, etc.

This lesson plan should consequently let students become aware of the relevance of ethical decisions and to which extent the development of artificial intelligence requires the definition of a set of ethical principles (that both philosophers and scientists are far from agreeing) or an ethical framework for decision making by the so called moral machines. The computational thinking approach to Ethics will also help students to understand the need to combine ethical reflection and computing to formalise algorithms capable to perform at least "weak" ethical decisions, as long as technology and science won't be able to develop "real" moral machines.

5. CONCLUSIONS AND FUTURE WORK

Among all the branches of Philosophy, undoubtedly Ethics is currently one of the most challenging and with more and more relevant implications. Apart from the importance itself of teaching students to become ethically skilled as human beings and citizens, ethical approaches permeate sciences and even determine their future development, as this paper has demonstrated in the case of moral machines and artificial intelligence.

Moreover, living in a the 21st century implies the development of a set of skills, still under discussion, among which computational thinking, problem solving, decision making and STEM competencies play a relevant role. Compulsory school should tackle these skills by developing training actions that will help our students acquiring the ability to apply school curricula within these approaches. Computational thinking is not only a matter of computer sciences, as ethical behaviour is not only a matter of philosophical studies.

The development of training actions with a cross-curricular approach, covering subjects as Computer Science, Ethics, Mathematics, Arts & Crafts, Language, etc., will help us to succeed in this goal of preparing our students to cope with their own future as workers and citizens.

This training action allows students to understand and experience the implications of moral decisions, while they put into practice philosophical skills like logical reasoning, resolution of moral dilemmas, etc. Besides, by connecting sciences and humanities they can start understanding human knowledge as a *continuum* instead of perceiving it as a set of "watertight compartments".

The set of activities introduce students to the representation (visually and logically) of ethical implications of machine behaviours, as they are expected to become “moral machines” in the near future. In the meanwhile, students will be able to analyse the implications of programming decision making frameworks and algorithms to these machines and discuss those consequences. Future directions of this study will move forward towards the implication of Computer Sciences subjects, in order to continue this work in Ethics classroom by representing different moral scenarios with programming languages usually studied in these subjects. These scenarios could be perfectly programmed in Arduino, Scratch or any other language that could also be inserted to virtual or real robots, in order to both improve computational thinking abilities and check the algorithms and decision making processes in “real” scenarios, so trying the meaning of roboethics in real robots.

6. ACKNOWLEDGMENTS

With the support of the EU Erasmus+ Programme. KA2 project “TACCLE 3 – Coding” (2015-1-BE02-KA201-012307).

This project has been funded with support from the European Commission. This communication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

7. REFERENCES

- [1] Allen, C., Smit, I. and Wallach, W. 2005. Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and Information Technology*, 7, 149-155. DOI=10.1007/s10676-006-0004-4.
- [2] Anderson, M. and Anderson, S. 2007. Machine Ethics: Creating an Ethical Intelligent Agent. *AI Magazine*, 28(4), 15-26.
- [3] Balanskat, A. and Engelhardt, K. 2015. Computing our future. Computer programming and coding. Priorities, school curricula and initiatives across Europe. European Schoolnet, Brussels.
- [4] Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., and Rumble, M. 2012. Defining twenty-first century skills. In P. Griffin, B. McGaw, and E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Springer, Netherlands.
- [5] Bonnefon, J.-F., Shariff, A. and Rahwan, I. 2016. The social dilemma of autonomous vehicles. *Science*, 352(6293), 1573-1576. DOI=10.1126/science.aaf2654.
- [6] Bratman, M. 1987. *Intention, plans and practical reason*. The University of Chicago Press, Chicago, IL.
- [7] Bundy, A. 2007. Computational thinking is pervasive. *Journal of Scientific and Practical Computing* 1, 67–69.
- [8] Caprile, M., Palmén, R., Sanz, P. and Dente, G. 2015. Encouraging STEM studies. Labour Market Situation and Comparison of Practices Targeted at Young People in Different Member States. Study. European Parliament. Directorate General for Internal Policies. Policy Department A: Economic and Scientific Policy. DOI= 10.2861/519030.
- [9] CEDEFOP, 2015. *EU Skills Panorama (2014) – STEM skills Analytical Highlight*. European Commission. http://skillspanorama.cedefop.europa.eu/sites/default/files/EUSP_AH_STEM_0.pdf.
- [10] Cherubini, M. 2014. Ethical Autonomous Vehicles. Retrieved from <http://research.mchrbn.net/eav>.
- [11] Coleman, K. G. 2001. Android arete. Toward a virtue ethic for computational agents. *Ethics and Information Technology*, 3, 247-265. DOI=10.1023/A:1013805017161.
- [12] Conferencia de Directores y Decanos de Ingeniería Informática, Asociación de Enseñantes Univesritarios de Informática, 2014. Declaración por la inclusión de asignaturas específicas de ciencia y tecnología informática en los estudios básicos de la enseñanza secundaria y bachillerato. CODDII-AENUI. <http://coddii.org/wp-content/uploads/2015/06/declaración-CODDII-AENUI.pdf>.
- [13] Cuny, J., Snyder, L. and Wing, J. M. 2011. Demystifying Computational Thinking for Non-Computer Scientists. Unpublished work.
- [14] European Commission, 2007. *Key Competences for Lifelong Learning. European Reference Framework*. Office for Official Publications of the European Communities, Luxembourg.
- [15] Fleischmann, K. R., Oard, D. W., Cheng, A.-S., Wang, P. and Ishita, E. 2009. Automatic classification of human values: Applying computational thinking to information ethics. *Proceedings of the Association for Information Science and Technology*, 46, 1–4. DOI=10.1002/meet.2009.1450460345.
- [16] García-Peñalvo, F. J. 2016. A brief introduction to TACCLE 3 – Coding European Project. In *2016 International Symposium on Computers in Education (SIIE)*, F.J. García-Peñalvo and J.A. Mendes Eds. IEEE, USA.
- [17] García-Peñalvo, F. J. 2016. Presentation of the TACCLE3 Coding European Project. <http://repositorio.grial.eu/handle/grial/654>.
- [18] García-Peñalvo, F. J. 2016. Proyecto TACCLE3 – Coding. In *XVIII Simposio Internacional de Informática Educativa, SIIE 2016*, F.J. García-Peñalvo and J.A. Mendes Eds. Ediciones Universidad de Salamanca, Salamanca, España, 187-189.
- [19] García-Peñalvo, F. J. 2016. What Computational Thinking Is. *Journal of Information Technology Research*, 9(3), v-viii.

- [20] Goodall, N. 2014. Ethical Decision Making During Automated Vehicle Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2424, 58-65. Transportation Research Board of the National Academies, Washington, D.C. DOI= 10.3141/2424-07.
- [21] Greene, J. D. 2016. Our driverless dilemma. When should your car be willing to kill you? *Science* 352(6293), 1514-1515. DOI= 10.126/science.aaf9534.
- [22] Grover, S. and Pea, R. D. 2013. Computational Thinking in K-12. A Review of the State of the Field. *Educational Researcher*, 42(1), 38-43. DOI= 10.3102/0013189X12463051.
- [23] Justo, D. 2016. España necesitará importar ingenieros si no aumentan los titulados en diez años. *Cadena Ser* (2016-11-10). http://cadenaser.com/ser/2016/10/11/sociedad/1476184570_764354.html.
- [24] Knigh, W. 2015. How to Help Self-Driving Cars Make Ethical Decisions. *MIT Technology Review* (July 29, 2015). Retrieved from <https://www.technologyreview.com/s/539731/how-to-help-self-driving-cars-make-ethical-decisions/>.
- [25] Lye, S. Y. and Koh, J. H. L. 2014. Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior*, 41, 51-61. DOI= 10.1016/j.chb.2014.09.012.
- [26] Ministerio de Educación Cultura y Deporte, 2015. Orden ECD/65/2015, de 21 de enero, por la que se describen las relaciones entre las competencias, los contenidos y los criterios de evaluación de la educación primaria, la educación secundaria obligatoria y el bachillerato. Boletín Oficial del Estado (n. 25, 29th January 2015, pp. 6986-7003). http://www.boe.es/diario_boe/txt.php?id=BOE-A-2015-738.
- [27] Rahwan, I., Bonnefon, J.-F. and Shariff, A. 2016. Moral Machine. Retrieved from <http://moralmachine.mit.edu>.
- [28] Riek, L. D. and Howard, D. 2014. A Code of Ethics for the Human-Robot Interaction Profession. *Proceedings of We Robot* (Coral Gables, FL, April 4-5, 2014). Available at SSRN: <https://ssrn.com/abstract=2757805>.
- [29] Rychen, D. S. and Salganik, L. H. (eds). 2003. *Key Competencies for a Successful Life and a Well-Functioning Society*. Hogrefe & Huber, Göttingen.
- [30] Salas, J. 2016. ¿Comprarías un coche que elegiría matarte para salvar otras vidas? *El País* (June 24, 2016). Retrieved from http://elpais.com/elpais/2016/06/22/ciencia/1466610816_591801.html.
- [31] TACCLE 3 Consortium TACCLE 3: Coding Erasmus + Project website. City, 2016.
- [32] UNESCO. 2015. Level-setting and recognition of learning outcomes. The use of level descriptors in the twenty-first century. United Nations Educational, Scientific and Cultural Organization, Paris.
- [33] Veres, S. M., Molnar, L., Lincoln, N. K. and Morice, C. P. 2010. Autonomous vehicle control systems - A review of decision making. *Proceedings of the Institution of Mechanical Engineers, Part I - Journal of Systems and Control Engineering* 225(3). DOI= 10.1177/2041304110394727.
- [34] Wallach, W., Franklin, S. and Allen, C. 2010. A Conceptual and Computational Model of Moral Decision Making in Human and Artificial Agents. *Topics in Cognitive Science*, 2, 454-485. DOI= 10.1111/j.1756-8765.2010.01095.x
- [35] Wilson, C., Sudol, L. A., Stephenson, C., and Stehlik, M. 2010. *Running on Empty: The Failure to Teach K-12 Computer Science in the Digital Age*. Association for Computing Machinery (ACM), New York.
- [36] Wing, J. 2011. Research notebook: Computational thinking—What and why? *The Link Magazine*, Spring. Carnegie Mellon University, Pittsburgh. Retrieved from <https://www.cs.cmu.edu/link/research-notebook-computational-thinking-what-and-why>.
- [37] Wing, J. M. 2006. Computational Thinking. *Communications of the ACM*, 49(3), 33-35. doi:10.1145/1118178.1118215
- [38] Wing, J. M. 2008. Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society*, 366, 3717-3725. DOI= 10.1098/rsta.2008.0118.
- [39] Wing, J.M., 2011. Computational Thinking. In *2011 IEEE Symposium on Visual Languages and Human-Centric Computing*, G. Costagliola, A. Ko, A. Cypher, J. Nichols, C. Scaffidi, C. Kelleher and B. Myers Eds., 3-3.
- [40] Yadav, A., Good, J., Voogt, J. and Fisser, P. 2017. Computational Thinking as an Emerging Competence Domain. In M. Mulder (ed.), *Competence-based Vocational and Professional Education, Technical and Vocational Education and Training: Issues, Concerns and Prospects*, 23, 1051-1067. DOI= 10.1007/978-3-319-41713-4_49.