

Pensamiento Computacional entre Filosofía y STEM. Programación de Toma de Decisiones aplicada al Comportamiento de “Máquinas Morales” en Clase de Valores Éticos

Antonio Miguel Seoane Pardo

CÓMO REFERENCIAR ESTE ARTÍCULO:

Antonio Miguel Seoane Pardo Computational Thinking between Philosophy and STEM. Programming Decision Making applied to the Behaviour of “Moral Machines” in Ethical Values Classroom”, en IEEE-RITA, 2018
Doi: <https://doi.org/10.1109/RITA.2018.2809940>

Title— Computational Thinking between Philosophy and STEM. Programming Decision Making applied to the Behaviour of “Moral Machines” in Ethical Values Classroom.

Abstract— This article describes a learning activity on computational thinking in Ethics classroom with compulsory secondary school students (14-16 years old). 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 also have to deal 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.

Index Terms— Computational thinking, decision making, game-based learning, Ethics, Logic, moral machines, self-driving car.

I. INTRODUCCIÓN

EN las últimas décadas, tanto los investigadores como las autoridades educativas han venido discutiendo sobre las competencias y destrezas clave en las que debe incidir el sistema educativo, desde la Educación Infantil hasta la finalización de los estudios universitarios, tanto para la formación de ciudadanos como para la cualificación de

trabajadores que encajen en las que (se espera) sean las habilidades clave de los empleos del siglo XXI [1]. Se trata de una discusión que dista mucho de estar cerrada, por más que se tienda a definir sistemas de competencias y destrezas más o menos estandarizados. Por ejemplo, el proyecto DeSeCo, de la OCDE, define un marco de competencias clave estructuradas en tres categorías bastante amplias: *usar herramientas de manera interactiva, interactuar en grupos heterogéneos y actuar de forma autónoma* [2]. Esta taxonomía constituye el fundamento para el desarrollo de las competencias de PISA. Por otra parte, el Marco europeo de competencias clave para el aprendizaje permanente [3] define un total de ocho competencias clave, a saber: *comunicación en la lengua materna, comunicación en lenguas extranjeras, competencia matemática y competencias básicas en ciencia y tecnología, competencia digital, aprender a aprender, competencias sociales y cívicas, sentido de la iniciativa y espíritu de empresa, y conciencia y expresión culturales*. Estas competencias, que adoptaron la forma jurídica de Recomendación del Parlamento y del Consejo Europeo, de 18 de diciembre de 2006, alcanzan el rango de Ley en las sucesivas reformas educativas españolas y, de manera más específica, en la LOMCE, la ley educativa actualmente en vigor [4]. Otros estudios [5] dividen las competencias en aspectos personales y éticos (*vivir en el mundo*), competencias laborales (entre las que se distinguen las *herramientas de trabajo* y las *formas de trabajar*) y de pensamiento (*formas de pensar*). El *pensamiento crítico, la resolución de problemas y la toma de decisiones* es una de las competencias (en este caso multifactorial) que se señalan como esenciales dentro del conjunto de las asociadas con el pensamiento. Se resalta aquí esta competencia en particular porque encaja perfectamente con el enfoque principal de este trabajo, y pone de manifiesto hasta qué punto son una única competencia: el pensamiento crítico es crucial para definir estrategias adecuadas conducentes a la resolución de problemas y debe ser la guía para la toma de decisiones de los individuos.

Además del problema suscitado por la definición de un marco competencial general para la formación académica y el aprendizaje permanente, en los últimos años se ha venido

Manuscrito recibido el 1 de junio de 2017; revisado 1 de julio; aceptado 26 de octubre.

English version received January, 29th, 2018. Revised February, 6th. Accepted February, 16th.

A. M. Seoane Pardo es Profesor Asociado del Área de Métodos de Investigación y Diagnóstico en Educación en la Universidad de Salamanca y miembro del GRUPO de Investigación en InterAcción y eLearning (GRIAL). (e-mail: anton.seoane@usal.es)
(<https://orcid.org/0000-0001-8887-3954>).

discutiendo el papel que han de desempeñar las competencias STEM (acrónimo formado por las iniciales de las áreas de Ciencias, Tecnología, Ingeniería y Matemáticas, en inglés) en la educación. En este contexto, se asiste a una extraña paradoja: la demanda creciente de puestos de trabajo relacionados con la tecnología y el conocimiento científico (particularmente en el ámbito de la Ingeniería, pero también en otras áreas) no se refleja en el incremento de estudiantes en los grados universitarios que proporcionan dicha cualificación. Antes al contrario, todo parece indicar que algunos países (entre los cuales se encuentra España) podrían empezar a demandar ingenieros extranjeros en la próxima década si esta tendencia no cambia drásticamente [6]. Esto es algo que preocupa tanto en contextos académicos como políticos, puesto que se trata de una situación que puede afectar (y mucho) al desarrollo económico del país, en un contexto, además, en el que hay una elevada tasa de universitarios sin empleo, subempleados o recurriendo a la emigración como única forma de encontrar perspectivas laborales. Si bien la situación de los universitarios españoles no es comparable actualmente a la de otros países del entorno europeo, la falta de estudiantes en las carreras universitarias relacionadas con STEM no es un fenómeno que se produzca únicamente en España, sino que afecta también a otros países europeos [7].

Así pues, cabe plantearse qué ocurre en un sistema educativo que no es capaz de orientar a sus estudiantes hacia estudios que conducen a puestos de trabajo con las mayores y mejores condiciones de empleabilidad, especialmente cuando esta es una tendencia creciente, sostenida y estructural, no una circunstancia coyuntural del sistema productivo. Según algunos estudios realizados en Estados Unidos [8], el problema radica en la forma en que los docentes enseñan Informática actualmente, puesto que los centros educativos han hecho un enorme esfuerzo por incrementar la utilización, integración y enseñanza de las TIC en las aulas, al tiempo que confunden “el uso de la tecnología con la enseñanza de la informática como disciplina académica fundamental en el ámbito de las áreas STEM”. Dichos estudios recomiendan, entre otras medidas, “la creación de un conjunto bien definido de estándares para la enseñanza de la informática en educación preuniversitaria, basados en conceptos de pensamiento algorítmico/computacional”. Por otra parte, en la Unión Europea, donde se estima un déficit de 800.000 profesionales relacionados con los ámbitos TIC e Informática para 2020, “muchos docentes, padres, economistas y políticos empiezan a pensar que los estudiantes necesitan habilidades de informática y programación”, según un reciente estudio de European Schoolnet [9]. En esta misma línea, la Conferencia de Directores y Decanos de Ingeniería Informática reivindicaba en 2014 la inclusión de materias específicas de ciencia y tecnología informática en la Educación Secundaria Obligatoria y el Bachillerato [10]. Además, apuntan, es necesario reforzar las competencias STEM y abordarlas de manera transversal y genérica, sin que queden reducidas solo a las “asignaturas STEM”. Para ello, es necesario incrementar y reforzar las acciones formativas destinadas al profesorado, al objeto de garantizar su efectiva preparación.

Por otra parte, el informe *EU Skills Panorama 2014* (publicado en abril de 2015) [11], dedicado a las competencias STEM, establece que, en la mayoría de los países, el número de estudiantes y graduados en titulaciones STEM se ha incrementado. Sin embargo, muchos empleadores manifiestan que los egresados no están preparados para el trabajo, puesto que no poseen las competencias adecuadas, particularmente en lo concerniente a las competencias transversales (*soft skills*), a saber: competencias comunicativas, trabajo en equipo, pensamiento creativo, etc. En consecuencia, todo parece indicar que, además de la falta de competencias STEM, estas tampoco son suficientes, y otras competencias tan necesarias deben acudir “en auxilio” de las anteriores. Ni unas ni otras, además, deben ser solo objeto de estudio de las materias específicas: al igual que la competencia digital no debe circunscribirse a la asignatura de Informática, las competencias comunicativas no son solo atribución de los docentes de Lengua, y así sucesivamente. Se necesita un enfoque mucho más global para abordar las competencias, destrezas y habilidades que se deben impartir a los estudiantes, y no se pueden reducir las competencias clave (ni siquiera el pensamiento algorítmico/computacional) a un modelo de currículum STEM. El Proyecto Europeo TACCLE 3 – Coding [12-15] tiene como finalidad orientar a los docentes y proporcionar la formación necesaria para promover un nuevo *currículum de informática* que incluya, entre otros aspectos, la programación [9] y el pensamiento computacional [16-20]. La creatividad, el pensamiento crítico, la discusión ética y moral y la conciencia cultural, por mencionar solo algunos ejemplos, son competencias fundamentales para lograr una educación de calidad adaptada a las necesidades del siglo XXI. Por eso este trabajo se centra en el pensamiento computacional “más allá” del modelo STEM, precisamente en un ámbito (la Ética y los Valores Éticos) en el que los problemas a abordar deben ser analizados lógicamente (“computacionalmente”), pero las conclusiones están muy lejos de poder ser consideradas *soluciones* y, antes al contrario, suelen ser el punto de partida para nuevas discusiones y desafíos para el razonamiento y el pensamiento crítico. En pocas palabras, se propone el uso del pensamiento computacional como medio para la discusión ética y moral, en lugar de promover el pensamiento computacional como fin en sí mismo.

En las siguientes páginas se explicará el desarrollo de una actividad de aprendizaje que tiene como finalidad la toma de decisiones en la asignatura de Valores Éticos, con un enfoque basado en el pensamiento computacional. Para ello, en primer lugar se definirán con claridad los conceptos de “Pensamiento Computacional” y “Ética de las máquinas” mediante una sucinta revisión de la literatura científica; a continuación, se presentarán las actividades desarrolladas con estudiantes de Enseñanza Secundaria Obligatoria relacionadas con las máquinas morales; finalmente, se establecerá una comparativa entre esta actividad formativa y otras experiencias, para extraer algunas conclusiones a este respecto.

Este trabajo constituye la versión extendida y traducida de una comunicación presentada en la *Fourth International*

Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 16) [21]. Respecto a la comunicación original, se ha realizado un análisis más detallado en la mayor parte de las secciones, lo que se explicita a continuación. En primer lugar, se ha hecho un análisis detallado de los diferentes estudios de marcos competenciales que sirven de referencia para los actuales marcos legislativos, en lo que a competencias y destrezas se refiere, y se han aportado más evidencias del problema que se plantea en este estudio, tanto a nivel europeo como global. Además, se ha enfatizado el papel que proyectos como TACCLE 3 – Coding pueden desempeñar en la mejora de las competencias STEM desde edades tempranas. En la Fundamentación Teórica, se ha abordado un análisis más pormenorizado del marco jurídico y legislativo en materia de educación en España, desde los años 70 hasta la actualidad, y se ha explicado en qué medida el fomento de determinadas competencias no está logrando el impacto deseado, en términos de mejora de las habilidades STEM. En la Descripción de la Acción Formativa se han analizado con mayor profusión los diferentes modelos de toma de decisiones que han servido como referencia para el desarrollo del marco propuesto en esta experiencia. Finalmente, se han extendido las secciones de Discusión y Conclusiones y Trabajo Futuro, y se han añadido referencias respecto al trabajo original.

II. FUNDAMENTACIÓN TEÓRICA: PENSAMIENTO COMPUTACIONAL Y ÉTICA DE LAS MÁQUINAS

Jeannette M. Wing afirma con rotundidad que “el pensamiento computacional expresa cómo piensan los humanos, no los ordenadores” [19], porque representa la manera en que los seres humanos resuelven problemas y toman decisiones cuando actúan de manera racional. Obviamente, tomar decisiones y resolver problemas no son algo que afecte únicamente al razonamiento lógico o “STEM”. Por el contrario, se trata de estrategias que empleamos en nuestra vida cotidiana, ya sea de forma consciente o inconsciente, y el propósito de estas páginas es precisamente poner de manifiesto cómo podemos hacer uso de tales estrategias de forma estructurada y algorítmica (computacional), a fin de abordar, explicar y tratar de resolver una instancia concreta de los dilemas morales, aquellos que se aplican a un tipo específico de artefactos denominados *máquinas morales*, esto es, máquinas programadas para tomar decisiones con implicaciones de carácter ético. Un ejemplo de dichas máquinas morales, cuyo “comportamiento” es relativamente sencillo de estudiar y analizar en Valores Éticos con estudiantes, es el del coche autónomo.

A. Pensamiento Computacional y Valores Éticos

Cuando se pretende promover estrategias, herramientas o contextos de innovación en los procesos formativos, es casi inevitable tener que enfrentarse a la *resistencia al cambio* que ejercen los docentes, por una parte, pero también el sistema educativo en su conjunto. Esta amenaza potencial dificulta cualquier proceso de innovación, ya sea didáctica, metodológica o tecnológica. Por el contrario, en el sistema educativo, como en cualquier colectivo humano, hay

profesionales que sienten una poderosa atracción hacia cualquier nueva moda o tendencia, pues son naturalmente innovadores, perciben con entusiasmo cualquier novedad y tratan de experimentarla. Estos, curiosamente, en lugar de ser percibidos como catalizadores de la innovación, lamentablemente (y por supuesto sin pretenderlo), con frecuencia alimentan todavía más, si cabe, la resistencia al cambio de los primeros, que se ven incapaces de “seguir el ritmo” de los más avanzados. De este modo, sienten que la innovación es algo parecido al castigo divino del titán Sísifo, una tarea tan interminable como inútil que se les impone sin que saber muy bien por qué debe llevarse a cabo. He ahí otro de los riesgos de la innovación. Por último, citando a Descartes en el *Discurso del método*, están aquellos que, “sin ser llamados ni por su alcurnia ni por su fortuna al manejo de los negocios públicos, no dejan de hacer siempre, en idea, una nueva reforma” [22]. No es fácil plantear un marco estable y duradero para la innovación donde, desde 1970, se han vivido ocho leyes educativas diferentes en Primaria y Secundaria (LGE, LOECE, LODE, LOGSE, LOPEG, LOCE, LOE y LOMCE), una en Formación Profesional (Ley Orgánica de Cualificaciones) y cuatro en enseñanzas universitarias (LGE, LRU, LOU y LOMLOU). La “innovación legislativa”, cuando no produce cambios significativos en los procesos de enseñanza-aprendizaje (tal es el caso de las reformas antedichas en España, que no han definido un marco para la innovación ni la mejora de los procesos docentes) deviene una especie de gatopardismo: “si queremos que todo siga como está, es necesario que todo cambie”, decía uno de los personajes de Giuseppe Tomasi di Lampedusa en su novela.

Las reformas educativas no parecen suponer un estímulo para Las últimas reformas educativas han favorecido el estudio de disciplinas relacionadas con la Economía y la necesidad de incluir contenidos económicos y de emprendimiento entre los contenidos básicos de todas las materias, en detrimento de materias como Ética, Educación para la Ciudadanía e Historia de la Filosofía, incluso cuando muchos autores (entre ellos afamados economistas) ponen el acento en la importancia del pensamiento crítico que, si bien no es territorio exclusivo de las materias del área de Filosofía, sí tienen en ellas un terreno fértil para su estímulo. Recuérdese que el pensamiento crítico es clave para la resolución de problemas, según un estudio anteriormente citado [5].

Junto con el incremento de los contenidos de Economía, la presencia creciente de las materias sobre TIC e Informática es otro de los elementos que llama la atención. En este último caso, el refuerzo de estas áreas les confiere la responsabilidad de incrementar la competencia digital y, en cierto modo, de hacer que los estudiantes tomen conciencia de la relevancia de las materias STEM, atrayéndoles hacia ellas, dada la baja demanda de este tipo de estudios y su alto potencial de empleabilidad. Sin embargo, con toda probabilidad habría sido deseable realizar un análisis en profundidad sobre las causas que subyacen a esta falta de “vocación y formación en STEM” para reflexionar sobre cómo el sistema educativo puede contribuir a resolver este problema, porque incrementar “sin más” la presencia de materias relacionadas con las TIC y la Informática es inútil

si no se fomenta la creatividad, la abstracción, el razonamiento lógico, el pensamiento crítico, etc.; y estas competencias pueden y deben ser estimuladas en todas las materias, a condición de que se haga con los enfoques metodológicos y didácticos adecuados.

Entre las competencias que un estudiante debe poseer destaca el pensamiento lógico o algorítmico, últimamente denominado también *pensamiento computacional*. Este concepto surge precisamente en el contexto de la discusión sobre las competencias en alfabetización digital y la relevancia que actualmente se atribuye al desarrollo de código informático como parte de la adquisición de dichas competencias. Pues bien, el desarrollo de código informático está estrechamente relacionado con las tareas de programación, pero el pensamiento computacional comporta un uso más amplio, como indica García Peñalvo: “aunque el desarrollo de código informático es interesante, resulta más útil enfatizar en la idea del pensamiento computacional como aplicación de alto nivel de abstracción y aproximación algorítmica a la resolución de cualquier tipo de problemas” [16]. En consecuencia, es importante distinguir entre pensamiento computacional y desarrollo de código informático o lenguajes de programación. La programación constituye un modo particular de pensamiento computacional [23].

El pensamiento computacional, tal como ha sido definido por J. M. Wing, es un tipo de pensamiento analítico [19] que “describe la actividad mental de formular un problema que admite una solución computacional” [20]. De este modo, su influencia se extiende a áreas de investigación tanto científicas como humanísticas [24], y puede ponerse en práctica en cualquier materia del currículo escolar. Otra definición (más sistemática y comprehensiva) de Wing y otros autores, establece que “el pensamiento computacional es el conjunto de procesos de pensamiento implicados en la formulación de problemas y sus soluciones para que dichas soluciones puedan ser representadas de tal modo que puedan ser ejecutadas por un agente de procesamiento de la información” [25]. Así pues, el pensamiento computacional no puede restringirse a materias relacionadas con la computación o la programación, y la robótica no puede ser el único modo de poner en práctica esta competencia. Las llamadas competencias STEM no son solo responsabilidad de las materias STEM.

El pensamiento computacional es una competencia emergente tanto en la educación primaria como secundaria. Numerosos estudios (como por ejemplo los desarrollados en Reino Unido, Holanda y Estados Unidos) ponen de manifiesto que se están realizando importantes esfuerzos por mejorar esta competencia a través de las materias relacionadas con la Informática [26]. Sin embargo, otros análisis sobre el estado de la cuestión [27] reconocen que, a pesar de las experiencias innovadoras de introducción a la computación (utilizando diferentes entornos y herramientas, tales como programación visual, kits de robótica, videojuegos, etc.), persisten ciertas dificultades para definir con claridad las competencias computacionales e incluso para evaluar su impacto efectivo en términos de mejora del rendimiento académico. Estos estudios revelan la necesidad de seguir investigando en el ámbito de las ciencias de la

educación aplicadas a este contexto, incluido el impacto de nuevos enfoques como el aprendizaje situado y distribuido, el uso de agentes inteligentes, el análisis de la interacción, la actividad y el discurso en contextos digitales, aspectos cognitivos implicados en el aprendizaje de conceptos computacionales, etc. Además, y esto es si cabe más relevante, se ha investigado muy poco sobre el impacto de la informática como medio para la enseñanza de otras disciplinas; por ejemplo, en qué medida las estrategias de resolución de problemas, tal como se desarrollan en tareas de programación informática, pueden aplicarse en otras materias.

Probablemente este último aspecto no se esté enfocando de forma adecuada. La cuestión no radica en cómo *exportar* el pensamiento computacional de la informática a otras materias. La mejor manera de promover eficazmente el pensamiento computacional y, en consecuencia, incrementar la capacidad de razonamiento lógico de nuestros estudiantes, consiste precisamente en su introducción como estrategia pedagógica para la formulación y resolución de problemas, tanto en asignaturas científicas como humanísticas, no solo en las materias de informática. Y es precisamente así como este estudio pretende contribuir proponiendo un ejemplo de desarrollo de pensamiento computacional desde el área de Filosofía y, más concretamente, desde la asignatura de Valores Éticos en Enseñanza Secundaria Obligatoria. El desarrollo de un marco para los procesos de toma de decisiones aplicado al análisis de dilemas morales, dadas ciertas variables y condiciones específicas, y conforme a ciertos principios morales, constituye un ejemplo de pensamiento computacional aplicado a esta materia mediante el que los estudiantes pueden poner en práctica el razonamiento abstracto sin recurrir al uso de lenguaje simbólico ni de programación, pero siguiendo los mismos esquemas de abstracción, resolución de problemas y toma de decisiones del pensamiento computacional en el contexto de la inteligencia artificial.

B. Ética de las Máquinas y Máquinas Morales

Anderson y Anderson explican con claridad [28] un concepto relativamente nuevo que consiste en dotar de dimensión ética a las máquinas y que difiere del que comúnmente se conoce como *ética computacional*. Se trata del concepto de *ética de las máquinas*. A diferencia de la ética computacional, que hace referencia al comportamiento humano y a las implicaciones morales del uso de las máquinas por los humanos, la ética de las máquinas alude a la conducta moral de los propios artefactos dotados de inteligencia artificial.

Este principio de ética de las máquinas, según el cual es posible dotar de conducta moral a seres de inteligencia artificial podría parecer, en primera instancia, simple ciencia ficción. De hecho, el origen de la ética de las máquinas se remonta a las Tres Leyes de la Robótica, conocidas también como “Leyes de Asimov”, quien definió este conjunto de reglas por vez primera en un relato corto, “Runaround”, de 1942. No obstante, con el posterior desarrollo de la robótica, por una parte, y de la inteligencia artificial, por otra, surge el problema de la toma de decisiones por parte de máquinas

cada vez más complejas que, como parte de su proceso de sofisticación, pronto estarán en condiciones de actuar de forma autónoma. En consecuencia, algunas de las decisiones que puedan tomar comportarán consecuencias éticas y morales.

La ética de las máquinas debe tomarse en consideración, al menos, por tres razones: “en primer lugar, se derivan consecuencias éticas de lo que las máquinas son capaces de hacer actualmente y de lo que se espera que hagan en el futuro [...]. En segundo lugar, podría argumentarse que el temor de los humanos a la posibilidad de máquinas autónomas inteligentes se debe a la preocupación acerca de si estas máquinas se comportarán éticamente, por lo que el futuro de la inteligencia artificial puede estar en juego [...]. Finalmente, creemos que es posible que la investigación en ética de las máquinas permita avanzar en el estudio de la teoría ética [28]. Se trata, pues, de un problema con implicaciones de ingeniería y filosóficas y, para ser honestos, estas últimas son con diferencia mucho más complejas que las primeras, a tenor de las discusiones de la Filosofía Occidental durante más de 2.500 años. Sea como sea, cualquier intento por definir un conjunto de principios o un marco para el comportamiento ético de las máquinas constituye un buen principio para enfrentarse a este desafío, incluso si nos vemos obligados a discutir metafóricamente sobre “virtudes de las máquinas” [29], a definir principios reguladores de carácter moral para la inteligencia artificial, o a crear incluso una *directiva primera* para el código ético de la interacción humano-robot (IHR) [30].

Algunos autores sugieren que, antes de tratar de *enseñar* a las máquinas a tomar decisiones éticas, quizá sería interesante tratar de aplicar el pensamiento computacional a la propia Ética y, de este modo, tratar de automatizar el proceso de detección y clasificación de los valores expresados a través de la comunicación humana [31]. Así pues, suponiendo que fuera técnica y filosóficamente posible convertir máquinas en sujetos capaces de producir un razonamiento moral explícito, el debate radicaría en la selección del enfoque moral más adecuado [32]. Los enfoques deductivos (*top-down approaches*) se basan en la idea según la cual los principios morales o teorías éticas constituyen la base para la creación de reglas que permitan desarrollar acciones éticamente correctas. La Ética deontológica de Kant o el Utilitarismo constituyen teorías éticas susceptibles de encajar en este tipo de enfoque. Por otra parte, los enfoques inductivos (*bottom-up approaches*) no se basan en teorías éticas concretas sino en principios evolutivos y dinámicos, partiendo del principio según el cual la inteligencia artificial “imita” el desarrollo infantil (tomando como base las teorías de Turing y, posteriormente, de Piaget, Kohlberg, etc.), y de este modo la máquina “aprenderá”. Finalmente, y puesto que ninguno de los modelos anteriores satisface los criterios para el diseño de una entidad artificial como agente moral, se requiere el desarrollo de enfoques híbridos que combinen un conjunto de reglas de carácter deductivo con la experiencia y el aprendizaje recopilado por la máquina.

Aunque esta discusión resulta extremadamente interesante y merecería sin duda un análisis más pormenorizado, excede con mucho el propósito de estas páginas y deberá dejarse

para estudios posteriores. En este caso, nos centraremos en diferentes enfoques morales (en su mayoría deductivos, de acuerdo con la clasificación anterior), con el propósito de proponer y analizar escenarios en los que los estudiantes puedan representar, responder y discutir diferentes posibilidades en relación con una serie de dilemas morales que se experimentarán en contextos en los que ciertas máquinas interactúan con otras máquinas y/o seres humanos.

C. ¿Por qué el Vehículo Autónomo?

Los vehículos autónomos serán pronto una realidad cotidiana en nuestras vidas. El primer test de conducción autónoma se llevó a cabo en 2007 y algunos prototipos, como el coche autónomo de Google, han recorrido miles de kilómetros de conducción en carreteras convencionales [33]. Son numerosos los estudios sobre algoritmos de navegación y modelos de toma de decisiones y sistemas de control para la programación de estos vehículos [34]. Pero probablemente algunas de las principales dificultades para el desarrollo de los vehículos autónomos sean más de índole filosófica (concretamente ética) y sociológica que propiamente técnica. En 2015, la Universidad de Stanford realizó un seminario que reunió a filósofos e ingenieros para discutir e implementar principios éticos que pudieran ser formalizados y codificados para su posterior desarrollo y experimentación en simulaciones e incluso en vehículos reales [35]. El debate en la prensa general y las revistas de divulgación científica mostraba titulares como “¿Compraría un coche que elegiría matarte para salvar otras vidas?” [36] o “Nuestro dilema sin conductor. ¿Cuándo tu coche debe estar dispuesto a matarte?” [37]. Estas cabeceras subrayan el hecho de que no existe un consenso ético (ni sociológico) sobre el enfoque moral del que debe dotarse a estos artefactos. Algunos estudios muestran que existe una cierta tendencia a aceptar, con carácter general, normas y principios de carácter utilitarista (como en el caso del conocido dilema del tranvía [38-39], en el sentido de minimizar el número de víctimas en la carretera y optar por el menor daño posible en términos cuantitativos). Sin embargo, los mismos sujetos que manifiestan estas actitudes morales preferirían viajar en y poseer un vehículo que protegiera sus vidas a toda costa, con independencia de las consecuencias utilitaristas que se pudieran derivar de las acciones de la máquina [33]. Ciertamente, la mayoría de nosotros accionaría la palanca que desviaría al tranvía de su trayectoria para que “solo” se produzca una víctima en lugar de las cinco que fallecerían si el tranvía siguiera el rumbo establecido. Eso, claro está, si la persona que está sola en la segunda vía no es un niño al que conocemos y que, cosas de la vida, es nuestro propio hijo.

Los principios morales de carácter utilitarista serían quizá aceptables para los fabricantes de automóviles, especialmente si por “utilidad” se entiende “menor coste”. En este caso, probablemente, las compañías de seguros también se sumarían a este tipo de directrices, solo que lo que unos y otros entenderían por “menor coste” no sería necesariamente lo mismo. Y algo similar ocurriría si sumáramos a esta ecuación a los gobiernos. Por otro lado, muchos de los demás enfoques (deontológico, por ejemplo,

o incluso el comportamiento aleatorio en ciertos casos, por mencionar únicamente algunos enfoques deductivos) serían perfectamente aceptables y éticamente defendibles. Esto pone sobre la mesa la cuestión (no menor) sobre quién debe decidir los principios morales en función de los cuales se programa a la máquina: ¿los estados?; ¿las compañías de seguros?; ¿las marcas?...

En resumen, “el problema, al parecer, es más filosófico que técnico. Antes de que podamos imponer nuestros valores a las máquinas, debemos resolver cómo hacer nuestros valores claros y consistentes” [37]. Así pues, el vehículo autónomo proporciona un campo de experimentación ideal para el pensamiento computacional en relación con la Ética. La otra razón para centrar el problema en los coches autónomos es que existe un número significativo de estudios y simuladores que nos permite visualizar y formalizar dilemas morales. Los *comportamientos éticos* del vehículo autónomo se representan habitualmente como trayectorias y movimiento, lo que resulta muy interesante por la claridad con la que se pueden analizar visualmente causas y consecuencias, así como tomar decisiones basándose en determinados principios éticos. Algunos de estos escenarios se han concebido específicamente a partir de enfoques de carácter ético [40], y facilitan la posibilidad de analizar y simular diferentes opciones para el mismo escenario adoptando varios principios éticos o comportamientos [41]. Otros, en cambio, han sido definidos con un enfoque sociológico, y permiten decidir cuál de las posibles opciones es la que se considera más adecuada, comparando la respuesta con la que proporcionan otros usuarios [42]. El estudio de estos diferentes enfoques para la toma de decisiones constituye un ejercicio de “calentamiento” muy interesante para el desarrollo de la acción formativa que se describe en el próximo apartado.

III. DESCRIPCIÓN DE LA ACCIÓN FORMATIVA

El propósito de este estudio es describir un conjunto de actividades sobre pensamiento computacional aplicadas a la asignatura de Valores Éticos con estudiantes de Educación Secundaria Obligatoria (14-16 años). Los estudiantes objeto de estas actividades podrían haber cursado previamente Valores Éticos en los cursos inferiores, puesto que en el actual currículo educativo español se trata de una asignatura optativa, alternativa a los estudios de religión confesional. Por tanto, los estudiantes de 3º y 4º de E. S. O. deben estar familiarizados con la resolución de dilemas morales y la resolución de problemas en contextos éticos. Por el contrario, se trata de estudiantes que carecen de conocimiento de lógica simbólica (que se estudia en 1º de Bachillerato, el primer curso de educación secundaria no obligatoria, en la asignatura de Filosofía), de modo que esta actividad será también útil como introducción al razonamiento lógico, si bien se desarrollará sin recurrir al uso de estructuras de lenguaje formal.

En las siguientes subsecciones se mostrará la estructura y las líneas maestras de la actividad formativa. Previamente, sin embargo, se destinarán algunas líneas a explicar las dificultades que se derivarían del uso de otros marcos de toma de decisiones éticas existentes, ya sea por su

complejidad o por la falta de un enfoque “puramente ético” para la experimentación en el contexto aquí descrito.

A. Modelo de Toma de Decisiones Éticas

Aunque no sea el propósito de este trabajo, se ha considerado oportuno realizar una breve revisión de la literatura sobre modelos y marcos de toma de decisiones éticas, con el objeto de analizar su aplicabilidad en este contexto. Básicamente se han estudiado dos tipos de marcos: unos, relacionados con modelos computacionales de toma de decisiones en inteligencia artificial (o aplicados específicamente a máquinas morales), y otros, procedentes directamente de la investigación filosófica. Ambos deberán ser descartados para el propósito de este estudio, por las razones que se indicarán a continuación.

Los modelos más utilizados aplicables a la programación de inteligencia artificial, tales como BDI o LIDA (por mencionar solo un par de ejemplos), son demasiado complicados para satisfacer los requerimientos de una actividad simple para jóvenes estudiantes de secundaria. El modelo BDI (*Belief-Desire-Intention model of human practical reasoning*), del filósofo Michael Bratman [43], constituye la base para el desarrollo del modelo de *software* BDI para la programación de agentes inteligentes, así que, en principio, parece resultar interesante para trabajar el pensamiento computacional en procesos de toma de decisiones en Ética debido a la potencialidad de añadir principios morales a seres artificiales. Por otra parte, LIDA es una arquitectura cognitiva que incluye procesos de decisión moral, tanto para humanos como para agentes artificiales [44]. El proceso cognitivo comienza a partir de una colección inductiva (*bottom-up*) de datos sensoriales, valores y experiencias, y a continuación “actúa” mediante procesos deductivos (*top-down*) para dar sentido a todo el comportamiento. LIDA, BDI y otras arquitecturas cognitivas (como por ejemplo GWT, Clarion, ACT-R, etc.), incluso si poseen la capacidad de adoptar decisiones éticas “reales”, no satisfacen las condiciones requeridas para el desarrollo de procesos de toma de decisiones sencillos como los requeridos en este contexto.

En relación con los modelos para el análisis de dilemas morales, existe una amplia lista de modelos, desde deontológicos a utilitaristas, humanistas y otros muchos enfoques. Tales modelos, que se han aplicado con éxito en Medicina, Enfermería, Marketing, Economía, Psicología, etc., no son apropiados en este caso debido a su falta de “estructura computacional”. Permiten el desarrollo de análisis racionales de “pros” y “contras” y el estudio de las implicaciones de diferentes decisiones, pero son todavía demasiado complicados para el propósito de esta actividad formativa.

Así pues, ¿qué tipo de modelo de toma de decisiones éticas debe emplearse? En primer lugar, es necesario centrarse únicamente en enfoques deductivos (*top-down*) para la toma de decisiones, puesto que el propósito de esta actividad es comprender las implicaciones y consecuencias de la toma de decisiones a partir de diferentes principios éticos, y no definir un modelo de “aprendizaje” de un agente de inteligencia artificial, en el que se combinaran valores morales (evolutivos) con experiencias de aprendizaje

“reales”, lo cual constituiría un escenario ideal pero excesivamente complejo para este caso. En segundo lugar, es necesario reducir cada uno de los diferentes principios éticos a una regla (o a un conjunto reducido de reglas), de forma que se puedan programar y ejecutar comportamientos específicos en función de determinados principios prestablecidos en una máquina que habrá de analizar una serie de *inputs* y ejecutar un *output* en función de dichos principios previamente definidos y programados. En tercer lugar, el modelo resultante debe ser aplicable a escenarios en los que existan dos o más opciones en situaciones de colisiones inevitables.

Así pues, se ha definido un conjunto de enfoques éticos para guiar los procesos de toma de decisiones (aunque se pueden definir y añadir otros), a saber:

A. Enfoques Consecuencialistas

- A1. Utilitarismo: “la mejor acción es la que proporciona el mayor beneficio o causa el menor daño posible”.
- A2. Egoísmo (auto-protección): “la mejor acción es la que me protege a mí y a los que están conmigo”.
- A3. Ética basada en el beneficio: “la mejor acción es la que produce el menor coste económico”.

B. Enfoques No Consecuencialistas

- B1. Deontología: “la mejor acción es la que protege a quienes actúan conforme a las normas”.
- B2. No determinismo: “el comportamiento es el resultado del azar”.

B. Acción Formativa: Toma de Decisiones Éticas por parte de Vehículos Autónomos en Colisiones Inevitables

Al objeto de describir las actividades de aprendizaje a desarrollar en la clase de Valores Éticos, se propone la siguiente plantilla. De este modo, la planificación didáctica puede utilizarse separadamente del resto del presente trabajo.

1) Título

Introducción a las máquinas morales y la toma de decisiones en clase de Valores Éticos

2) Presentación

El objetivo de esta actividad consiste en formar a los estudiantes de E. S. O. en pensamiento computacional aplicado al ámbito de la Ética. Para ello, la planificación de la lección se centra en las implicaciones éticas de la programación de vehículos autónomos para que estos ejecuten determinados comportamientos con consecuencias morales en situaciones en las que una colisión es inevitable y se producirán daños personales y materiales con toda seguridad. Los estudiantes deben aprender a analizar, representar, estudiar posibles soluciones (*outcomes*) y reflexionar sobre los comportamientos éticos de las máquinas, así como “programar” dichos comportamientos como respuesta a determinados estímulos (*inputs*), siguiendo determinados principios morales.

Edad: 14-16 años.

Nivel: Medio.

Competencias: pensamiento computacional, toma de decisiones, razonamiento lógico, discusión ética, dilemas morales.

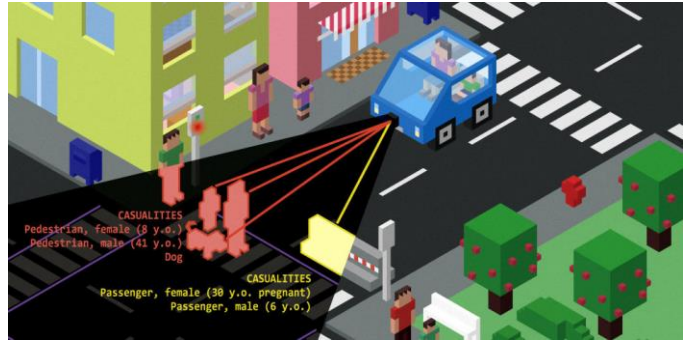


Figura 2. Ejemplo de escenario predefinido para discutir en grupos (Iyad Rahwan. <http://www.popularmechanics.com>)

3) Finalidad de la Actividad

Esta lección está orientada al desarrollo de competencias relacionadas con el pensamiento computacional, el razonamiento lógico y la toma de decisiones basada en algoritmos. Para ello, se analizarán y estudiarán diferentes enfoques éticos y se formalizarán sus principios morales esenciales al objeto de programar, teóricamente, máquinas inteligentes para que tomen decisiones morales de manera autónoma. Además de estas competencias “computacionales”, los estudiantes adquirirán consciencia de la relevancia de los valores y de los enfoques éticos, no solo como principios que sirven como guía de las decisiones cotidianas, sino también para discutir y tratar de alcanzar un consenso en relación con determinados principios morales “socialmente aceptados” para decidir cómo deberían comportarse unas máquinas “inteligentes”, tanto en las interacciones con humanos como con otras máquinas y el entorno.

4) Herramientas y Recursos

- ✓ MIT Media Lab: Moral Machine. <http://moralmachine.mit.edu>.
- ✓ ClaimMS: AccidentSketch.com. <http://draw.accidentsketch.com>.
- ✓ AV-DMEC Framework (véase plan de actividades).
- ✓ Matthieu Cherubini: Ethical autonomous vehicles. <https://vimeo.com/85939744>.
- ✓ Kahoot: <http://kahoot.it>.

5) Plan de Actividades

El presente plan de actividades se estructura en cinco sesiones, de la siguiente manera:

Sesión 1. Introducción y Plataforma Moral Machine

La primera sesión se destinará a presentar a los estudiantes los desafíos inherentes a la programación de vehículos autónomos, no solo desde un punto de vista técnico sino también (y quizá principalmente) debido a los problemas de carácter ético.

La primera actividad consistirá en la lectura de un artículo de periódico [2] y la posterior discusión en grupos sobre las diferencias entre los coches autónomos y los actuales vehículos, así como sobre las consecuencias de permitir que sean los vehículos quienes tomen las decisiones. ¿Quién es responsable del daño que provoca el coche? ¿Cómo te sentirías si supieras que, en determinadas circunstancias, el coche estaría dispuesto a matarte en lugar de matar a otros?

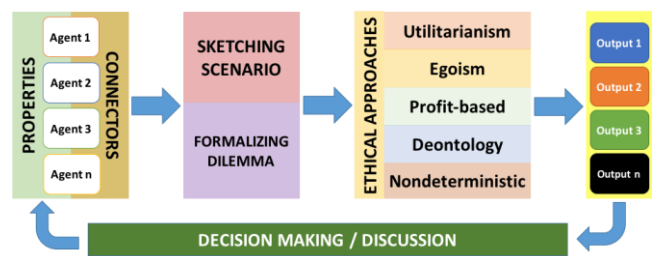


Figura 3. Marco AV-DMEC

En la segunda actividad, los estudiantes habrán de decidir cómo debería comportarse un vehículo en determinadas condiciones. Para ello, se les pedirá que visiten el portal del MIT Lab *Moral Machine*, <http://moralmachine.mit.edu> (véase figura 1), en el que pueden ver, diseñar y juzgar diferentes escenarios, para posteriormente comparar sus respuestas con los de otros usuarios y sus propios compañeros.

Sesión 2. Análisis de Escenarios Predefinidos

Durante la segunda sesión, el profesor dividirá la clase en grupos y proporcionará a los estudiantes algunos escenarios predefinidos (como se muestra en la Figura 2, por ejemplo). A continuación, los grupos analizarán, discutirán y decidirán cómo debería comportarse en vehículo según determinados enfoques éticos estudiados en temas anteriores, adecuadamente redefinidos para que encajen en las situaciones que se les presenta para su análisis. A continuación, cada grupo explicará al resto de compañeros el escenario que le ha correspondido para su análisis, así como las posibles opciones de acuerdo con diferentes enfoques éticos. Finalmente, se discutirá cada escenario en clase, tratando de alcanzar un acuerdo sobre el comportamiento y el enfoque ético que implica la “mejor” solución, si tal cosa es posible.

Sesión 3. Escenarios diseñados por los estudiantes (i)

Durante la tercera sesión, se pedirá a los estudiantes que diseñen en grupos sus propios escenarios para la toma de decisiones éticas por parte de máquinas inteligentes, recopilando datos de la matriz presente en la Tabla 1 y que formalicen los procesos de toma de decisiones conforme al marco *Autonomous Vehicle – Decision Making in Ethics*

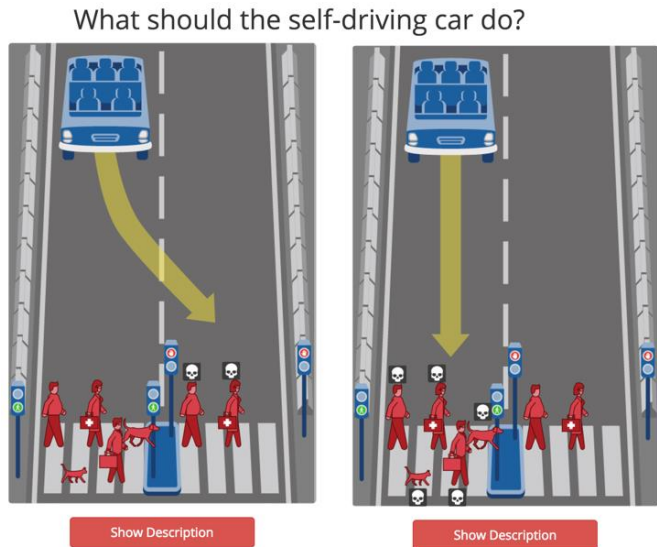


Figura 1. Portal *Moral Machine*. Funcionalidad de juicio

Classroom (AV-DMEC), tal como se muestra en la Figura 3. Para ello, deberán comenzar por seleccionar el número y naturaleza de *agentes* involucrados en el escenario. A continuación, dotarán a los agentes de determinadas *propiedades* al objeto de completar y clarificar las acciones a analizar. Seguidamente, realizarán un boceto de la escena utilizando algún tipo de herramienta gratuita como AccidentSketch.com, <http://draw.accidentsketck.com>. Deben describir la escena con proposiciones lógicas utilizando *conectores* y lenguaje natural (por ejemplo: El coche A se cruza en la trayectoria del vehículo autónomo; el vehículo autónomo no puede detenerse a tiempo para impedir la colisión Y atropellará al motociclista O chocará contra la barrera O atropellará a dos peatones en la acera).

Sesión 4. Escenarios diseñados por los estudiantes (ii)

Tras verificar que el borrador y las proposiciones lógicas encajan y describen con claridad el escenario propuesto por cada grupo, se pedirá a los estudiantes que analicen las posibles respuestas tomando en consideración diferentes *enfoques éticos*. Tras estudiar y discutir tales enfoques, se les pedirá que seleccionen el mejor enfoque ético en función del comportamiento que consideran “más deseable”, fundamentando éticamente su decisión.

Sesión 5. Discusión y Feedback

La última sesión se destinará a discutir sobre algunos problemas surgidos de las sesiones 3 y 4. Por ejemplo, ¿existe algún enfoque ético que resulte preferible *por norma general*? ¿Existen escenarios en los que resulta imposible determinar una respuesta “mejor”? ¿Existen escenarios en los que ninguno de los enfoques éticos parece proporcionar una solución razonable? Con el objetivo de mostrar a los estudiantes la complejidad del pensamiento computacional y los procesos de toma de decisiones, se les invitará a ver el vídeo *Ethical autonomous vehicles*, <https://vimeo.com/85939744>, en el que Matthieu Cherubini muestra dos casos de estudio (escenarios) bajo tres enfoques éticos diferentes. Se invitará a los estudiantes a analizar los algoritmos éticos presentados y la representación formal, tal como se muestra en <http://research.mchrbn.net/eav>. Finalmente, se pedirá a los estudiantes que participen en una actividad de competición de aprendizaje basado en juego (ludificación), al objeto de evaluar lo que han aprendido. Para ello, se realizará un test con Kahoot! (<http://kahoot.it>), previamente elaborado por el profesor.

TABLA I
MATRIZ CON ELEMENTOS DE EJEMPLO PARA EL DESARROLLO DE
ESCENARIOS DE ACCIDENTES CON VEHÍCULOS AUTÓNOMOS

Agentes	Coche autónomo, Bus, Bus escolar, Moto, Ciclista, Peatón, Obstáculo, Semáforo, [...]
Propiedades	Luz roja/verde, Niño, Bebé, Mujer embarazada, Anciano/a, Sentido contrario, Mismo sentido, Más rápido/Más lento, Con/Sin casco, Cruza correcta/incorrectamente, Impacto, Detención, Atropello [...]
Conectores	Y, O, SI, ENTONCES, O BIEN
Enfoques éticos	Utilitarismo, Egoísmo, Ética basada en el beneficio, Deontología, No determinismo

6) Evaluación

El objetivo de evaluación de esta experiencia es doble: por una parte, se pretende evaluar la capacidad de los estudiantes para comprender, definir (tanto visual como lingüísticamente) y tomar decisiones éticas sobre los problemas propuestos. Para ello, el docente tomará nota de las experiencias y resultados, discusiones y argumentaciones de los estudiantes, y les guiará para que alcancen un mejor y más correcto razonamiento lógico y ético, en el caso de que no se estén desarrollando de la manera adecuada. Por otro lado, es necesario evaluar el nivel de satisfacción de los estudiantes con la acción formativa en su conjunto, saber cómo se han sentido y hasta qué punto creen que han mejorado su capacidad para analizar lógicamente dilemas morales y para aplicar diferentes enfoques éticos a diversos escenarios predefinidos o creados por ellos mismos. Esto se llevará a cabo mediante un conjunto de cuestionarios desarrollados con Kahoot! (<http://kahoot.it>) y un proceso de reflexión final, al que seguirá un breve ensayo individual para su entrega al docente. De este modo, la actividad puede ser evaluada y mejorada para sucesivas iteraciones.

IV. DISCUSIÓN

Los modelos actualmente disponibles para describir y formalizar procesos de toma de decisiones aplicables a las máquinas morales son todavía demasiado complicados como para que puedan ser utilizados en clases de Ética con estudiantes de enseñanza obligatoria. Dichos modelos se centran en cómo la máquina “actúa” para recopilar información del entorno, cómo su inteligencia artificial procesa la información y cómo es capaz de ejecutar comportamientos programados (incluso con enfoques éticos). En lugar de esto, la actividad formativa aquí descrita está dirigida a reflexionar en la moralidad de la decisión misma más que a tratar de explicar cómo debería comportarse la máquina para actuar éticamente. Interesa más conocer *qué* debería hacer una máquina moral y *por qué* un vehículo autónomo debería actuar de cierta forma adoptando determinadas decisiones, en lugar de explicar *cómo* la máquina sería capaz de llevar a cabo tales tareas, porque nuestros estudiantes desempeñan una función de “filósofos”, no de “informáticos”. Además, es necesario que los estudiantes sean conscientes de la relevancia de formalizar los dilemas morales a resolver por parte de estas máquinas morales y que sean capaces de analizar las posibles conductas (*outputs*) bajo ciertos presupuestos éticos o enfoques éticos.

Plataformas como *Moral Machine*, del MIT Media Lab, constituyen una herramienta muy interesante para visualizar y juzgar escenarios predefinidos, o incluso para el diseño de nuevas situaciones en contextos similares. Sin embargo, *Moral Machine* se ha diseñado para analizar múltiples actitudes hacia las decisiones morales de vehículos autónomos y, por tanto, las implicaciones éticas se descubren *a posteriori*, en tanto que conjuntos de valores que se conforman en comparación con las respuestas mayoritarias del resto de usuarios. Para el propósito de esta experiencia de aprendizaje, por el contrario, es necesario analizar y tomar decisiones éticas desde enfoques éticos *a priori*, de forma que se puedan comparar diferentes comportamientos que podrían ser preferibles en función del algoritmo moral que fundamente los procesos de toma de

decisiones en cada caso. Además, este instrumento del MIT permite diseñar, visualizar y juzgar en escenarios donde solo dos opciones son posibles, lo cual es interesante para el análisis y discusión de un dilema, pero la experiencia resulta más interesante si los escenarios permiten contar con tres o más respuestas o comportamientos posibles, y también, por qué no, si estos comportan un cierto grado de controversia moral.

Esta propuesta no alcanza el nivel de complejidad técnica de un marco para la toma de decisiones aplicable a inteligencia artificial ni la simplicidad (todavía) de una herramienta en la que los usuarios pueden añadir variables, definir *outputs* y evaluar dichas respuestas en función de diferentes enfoques éticos. Sin embargo, posee las virtudes de permitir a los estudiantes definir un conjunto potencialmente infinito de escenarios a partir de un conjunto finito de variables, analizar posibles respuestas mediante el boceto y formalización de los escenarios utilizando el lenguaje natural (pero con estructuras lógicas y conectores) y, finalmente, comparar y discutir las consecuencias de las decisiones adoptadas tomando como referencia los enfoques éticos que los estudiantes conocen perfectamente. Durante la experiencia ponen en práctica competencias de pensamiento computacional, procesos de toma de decisiones, análisis filosófico de comportamientos morales y otras competencias transversales: trabajo en equipo, comunicación oral, debate y argumentación, etc.

Esta planificación didáctica debería favorecer, en consecuencia, que los estudiantes tomaran conciencia de la relevancia de las decisiones éticas y supieran hasta qué punto el desarrollo de la inteligencia artificial presupone la definición de un conjunto de principios éticos (que tanto filósofos como científicos y políticos están muy lejos de consensuar todavía) o un marco ético para la toma de decisiones por parte de las denominadas máquinas morales. El enfoque del pensamiento computacional aplicado a la Ética ayudará a los estudiantes a comprender la necesidad de combinar reflexión ética y computación para formalizar algoritmos capaces de llevar a cabo, cuanto menos, decisiones éticas “débiles”, en tanto en cuanto la tecnología, la ciencia y la ética no sean capaces de desarrollar auténticas máquinas morales.

V. CONCLUSIONES Y TRABAJO FUTURO

De todas las ramas de la Filosofía, la Ética es sin duda alguna una de las que se enfrenta a mayores y más importantes desafíos y que comporta implicaciones de mayor relevancia. Aparte de la importancia intrínseca de enseñar a los estudiantes a adquirir competencias para comportarse de manera éticamente responsable, tanto como personas como en su faceta de ciudadanos, es evidente que los enfoques éticos permean y afectan a las ciencias hasta el punto de determinar y condicionar su desarrollo futuro, tal como este trabajo ha demostrado en el caso de las máquinas morales y la inteligencia artificial.

Además, vivir en el siglo XXI comporta la necesidad de adquirir una serie de competencias que todavía se siguen discutiendo, pero entre las cuales se encuentran, sin duda alguna, el pensamiento computacional, la resolución de problemas y las competencias STEM. La educación obligatoria debe garantizar la adquisición efectiva de dichas competencias mediante el desarrollo de planes formativos

que ayuden a nuestros estudiantes a adquirirlas en el seno del currículum escolar. Sin embargo, el pensamiento computacional no es solo cuestión de asignaturas de informática, como tampoco el comportamiento ético y moral es dominio exclusivo de los estudios de carácter filosófico.

El desarrollo de actividades formativas con un enfoque interdisciplinar, que abarcan materias del ámbito de la Informática, Ética, Matemáticas, Artes Plásticas, Lengua, etc., permitirá alcanzar el objetivo de preparar a los estudiantes para converger con su futuro como trabajadores y como ciudadanos.

Esta iniciativa permite a los estudiantes comprender y experimentar las implicaciones de las decisiones morales, al tiempo que ponen en práctica competencias filosóficas como el razonamiento lógico, resolución de dilemas morales, etc. Además, conectar constantemente Ciencias y Humanidades contribuye a comprender el conocimiento humano como un *continuum* en lugar de presentarse como compartimentos estancos, lo cual es un error demasiado frecuente en nuestras sociedades actuales y, por ende, también en nuestros sistemas educativos.

Este conjunto de actividades introduce a los estudiantes en la representación (tanto gráfica como mediante formalización lógicas) de las implicaciones de los comportamientos de las máquinas, en la medida en que se espera que lleguen a convertirse en “máquinas morales” en un futuro próximo. Mientras tanto, los estudiantes pueden analizar las implicaciones de la programación de marcos de toma de decisiones y algoritmos en esas máquinas y discutir sus consecuencias.

Futuros desarrollos de esta investigación avanzarán hacia la implicación de las materias de Informática, de forma que las tareas iniciadas durante la clase de Valores Éticos puedan seguir representándose, mediante el desarrollo de iniciativas de programación o *coding*, en las clases de Informática o TIC. Para ello, y empleando lenguajes de programación como Arduino o Scratch, los escenarios definidos podrían insertarse en robots virtuales o reales, al objeto de mejorar al tiempo las habilidades de pensamiento computacional y comprobar los algoritmos y procesos de toma de decisiones en escenarios “físicos”, de forma que se pueda poner en práctica el significado de la *roboética* en auténticos robots.

VI. AGRADECIMIENTOS

Con el apoyo del Programa de la Unión Europea Erasmus+ Programme. KA2 project “TACCLE 3 – Coding” (2015-1-BE02-KA201-012307).

Este proyecto se ha financiado con el apoyo de la Comisión Europea. Esta comunicación refleja únicamente las opiniones del autor, y la Comisión no puede ser considerada responsable de cualquier uso que se pudiera hacer de la información contenida en este documento.

REFERENCIAS

- [1] UNESCO. Level-setting and recognition of learning outcomes. The use of level descriptors in the twenty-first century. Paris, United Nations Educational, Scientific and Cultural Organization, 2015.
- [2] D. S. Rychen and L. H. Salganik (eds). *Key Competencies for a Successful Life and a Well-Functioning Society*. Göttingen, Hogrefe & Huber, 2003.
- [3] European Commission. *Key Competences for Lifelong Learning. European Reference Framework*. Luxembourg, Office for Official Publications of the European Communities, 2007.
- [4] Ministerio de Educación Cultura y Deporte. 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.
- [5] M. Binkley, O. Erstad, J. Herman, S. Raizen, M. Ripley, M. Miller-Ricci and M. Rumble, “Defining twenty-first century skills” In P. Griffin, B. McGaw, and E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Netherlands, Springer, 2012.
- [6] D. Justo, “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.
- [7] M. Caprile, R. Palmén, P. Sanz and G. Dente. *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, 2015. DOI= 10.2861/519030.
- [8] C. Wilson, L. A., Sudol, C. Stephenson and M. Stehlik. *Running on Empty: The Failure to Teach K-12 Computer Science in the Digital Age*. New York, Association for Computing Machinery (ACM), 2010.
- [9] A. Balanskat and K. Engelhardt. *Computing our future. Computer programming and coding. Priorities, school curricula and initiatives across Europe*. Brussels, European Schoolnet, 2015.
- [10] Conferencia de Directores y Decanos de Ingeniería Informática, Asociación de Enseñantes Universitarios de Informática, *Declaración por la inclusión de asignaturas específicas de ciencia y tecnología informática en los estudios básicos de la enseñanza secundaria y bachillerato*. CODDII-AENUI, 2014. <http://coddii.org/wp-content/uploads/2015/06/declaración-CODDII-AENUI.pdf>.
- [11] CEDEFOP. *EU Skills Panorama (2014) STEM skills Analytical Highlight*. European Commission, 2015. http://skillspanorama.cedefop.europa.eu/sites/default/files/EUSP_AH_STEM_0.pdf.
- [12] F. J. García-Peñalvo. “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, 2016.
- [13] F. J. García-Peñalvo, D. Reimann, M. Tuul, A. Rees and I. Jormanainen, “An overview of the most relevant literature on coding and computational thinking with emphasis on the relevant issues for teachers,” TACCLE3 Consortium, Belgium, 2016. doi: 10.5281/zenodo.165123
- [14] F. J. García-Peñalvo. “Proyecto TACCLE3 – Coding”, in *XVIII Simposio Internacional de Informática Educativa, SIIE 2016*, F. J. García-Peñalvo and J. A. Mendes (eds.). Salamanca, Ediciones Universidad de Salamanca, pp. 187-189, 2016.
- [15] TACCLE 3 Consortium. TACCLE 3: Coding Erasmus + Project website. City, 2016. <http://www.taccle3.eu/en/>.
- [16] F. J. García-Peñalvo. “What Computational Thinking Is”. *Journal of Information Technology Research*, 9(3), pp. v-viii, 2016.
- [17] J. M. Wing. “Research notebook: Computational thinking—What and why?” *The Link Magazine*, Spring. Pittsburgh, Carnegie Mellon University, 2011. Retrieved from <https://www.cs.cmu.edu/link/research-notebook-computational-thinking-what-and-why>.
- [18] J. M. Wing. “Computational Thinking”. *Communications of the ACM*, 49(3), pp. 33-35, 2006. doi:10.1145/1118178.1118215.
- [19] J. M. Wing. “Computational thinking and thinking about computing”. *Philosophical Transactions of the Royal Society*, 366, pp. 3717-3725, 2008. DOI= 10.1098/rsta.2008.0118.
- [20] J. M. Wing. “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, 2011.
- [21] A. M. Seoane Pardo. “Computational thinking beyond STEM: an introduction to “moral machines” and programming decision making in ethics classroom”. *Proceedings of the Fourth International Conference on Technological Ecosystems for*

- Enhancing Multiculturalism*, pp. 37-44, 2016. DOI= <https://doi.org/10.1145/3012430.3012494>.
- [22] R. Descartes. *Discurso del método*. Madrid, Alianza Editorial, 2011 (original, 1637).
- [23] S. Y. Lye and J. H. L. Koh. "Review on teaching and learning of computational thinking through programming: What is next for K-12?" *Computers in Human Behavior*, 41, pp. 51-61, 2014. DOI= 10.1016/j.chb.2014.09.012.
- [24] A. Bundy. "Computational thinking is pervasive". *Journal of Scientific and Practical Computing* 1, pp. 67-69, 2007.
- [25] J. Cuny, L. Snyder and J. M. Wing. Demystifying Computational Thinking for Non-Computer Scientists. Unpublished work, 2011.
- [26] A. Yadav, J. Good, J. Voogt and P. Fisser. "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, pp. 1051-1067, 2017. DOI= 10.1007/978-3-319-41713-4_49.
- [27] S. Grover and R. D. Pea. "Computational Thinking in K-12. A Review of the State of the Field". *Educational Researcher*, 42(1), pp. 38-43, 2013. DOI= 10.3102/0013189X12463051.
- [28] M. Anderson and S. Anderson. "Machine Ethics: Creating an Ethical Intelligent Agent". *AI Magazine*, 28(4), pp. 15-26, 2007.
- [29] K. G. Coleman. "Android arete. Toward a virtue ethic for computational agents". *Ethics and Information Technology*, 3, pp. 247-265, 2001. DOI=10.1023/A:1013805017161.
- [30] L. D. Riek and D. Howard. "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>.
- [31] K. R. Fleischmann, D. W. Oard, A.-S. Cheng, P. Wang and E. Ishita. "Automatic classification of human values: Applying computational thinking to information ethics". *Proceedings of the Association for Information Science and Technology*, 46, pp. 1-4, 2009. DOI= 10.1002/meet.2009.1450460345.
- [32] C. Allen, I. Smit and W. Wallach. "Artificial morality: Top-down, bottom-up, and hybrid approaches". *Ethics and Information Technology*, 7, pp. 149-155, 2005. DOI=10.1007/s10676-006-0004-4.
- [33] J.-F. Bonnefon, A. Shariff and I. Rahwan. "The social dilemma of autonomous vehicles". *Science*, 352(6293), pp. 1573-1576, 2016. DOI= 10.1126/science.aaf2654.
- [34] S. M. Veres, L. Molnar, N. K. Lincoln and C. P. Morice. "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), 2010. DOI= 10.1177/2041304110394727.
- [35] W. Knight. "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/>.
- [36] J. Salas. "¿Compraría 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.
- [37] J. D. Greene. "Our driverless dilemma. When should your car be willing to kill you?". *Science* 352(6293), pp. 1514-1515, 2016. DOI= 10.126/science.aaf9534.
- [38] P. Foot. *The Problem of Abortion and the Doctrine of the Double Effect in Virtues and Vices*. Oxford, Basil Blackwell, 1978.
- [39] J. J. Thomson. "Judith Jarvis Thomson, The Trolley Problem", *Yale Law Journal*(94), pp. 1395-1415, 1975.
- [40] N. Goodall. "Ethical Decision Making During Automated Vehicle Crashes". *Transportation Research Record: Journal of the Transportation Research Board*, 2424, pp. 58-65. Washington, D. C. Transportation Research Board of the National Academies, 2014. DOI= 10.3141/2424-07.
- [41] M. Cherubini. "Ethical Autonomous Vehicles". Retrieved from <http://research.mchrbn.net/eav>, 2014.
- [42] I. Rahwan, J.-F. Bonnefon and A. Shariff. "Moral Machine". Retrieved from <http://moralmachine.mit.edu>, 2016.
- [43] M. Bratman. *Intention, plans and practical reason*. Chicago, IL, The University of Chicago Press, 1987.
- [44] W. Wallach, S. Franklin and C. Allen. "A Conceptual and Computational Model of Moral Decision Making in Human and Artificial Agents". *Topics in Cognitive Science*, 2, pp. 454-485, 2010. DOI= 10.1111/j.1756-8765.2010.01095.x.



Antonio Miguel Seoane Pardo (A Coruña, 1971) es Doctor por la Universidad de Salamanca (2014), Licenciado en Filosofía y Ciencias de la Educación (1994, Universidad de Salamanca) y funcionario en activo del Cuerpo de Profesores de Enseñanza Secundaria de la Junta de Castilla y León (España). Imparte docencia de Filosofía y Geografía e Historia desde 1998, actualmente en el IES Venancio Blanco, de Salamanca. Es Profesor Asociado en el Departamento de Didáctica, Organización y Métodos de Investigación de la Universidad de

Salamanca. Ha sido Profesor Colaborador en el Máster Universitario en Informática Pluridisciplinar (Especialidad de Enseñanza y Aprendizaje Electrónico) de la UAH (España) y es docente del Máster en eLearning: Tecnologías y Métodos de Formación en Red, de la USAL (España), y en el Máster TIC en Educación, de esta misma universidad.

Es Coordinador Académico e investigador activo del Grupo de Investigación en Interacción y eLearning (GRIAL). Es investigador reconocido en metodología de la formación online y diseño estratégico de instituciones virtuales. Fue el primer director de la Universidad Virtual de la USAL hasta diciembre de 2009, a cargo de desarrollar la metodología de formación para los profesores de la Universidad de Salamanca y la estrategia de apoyo a los más de 30.000 usuarios del campus virtual. Colabora habitualmente en iniciativas de formación del profesorado (ha impartido más de 75 cursos, tanto en ámbito nacional como internacional) y es autor de casi un centenar de publicaciones entre libros, capítulos y artículos de impacto sobre eLearning, la metodología de la formación online, la teoría de la comunicación y filosofía y retórica griega antigua. Ha participado y participa en más de una veintena de proyectos de investigación, tanto nacionales como internacionales, y 16 contratos de transferencia I+D+i. Ha presentado numerosas ponencias en congresos nacionales e internacionales y es autor de un buen número materiales didácticos disponibles en línea sobre herramientas y recursos TIC para la formación en línea.