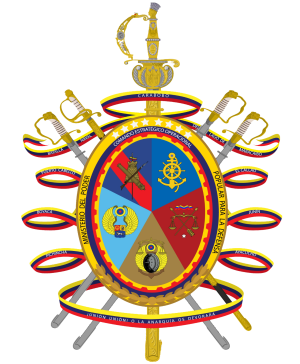
**REPÚBLICA BOLIVARIANA DE VENEZUELA**



**MINISTERIO DEL PODER POPULAR PARA LA DEFENSA**

**UNIVERSIDAD NACIONAL EXPERIMENTAL POLITÉCNICA**

**DE LA FUERZA ARMADA NACIONAL BOLIVARIANA**

**NÚCLEO MIRANDA – EXTENSIÓN SANTA TERESA DEL TUY**

**Los Principios Éticos Y Los Desafíos Asociados Al Desarrollo De Sistemas Autónomos Con Inteligencia Artificial**

**ESTUDIANTE:**

Sergio Ricaflor

C.I: 30.162.461

**TUTORA:**

Chacón Nelly

Santa Teresa del Tuy, Mayo de 2025

**EL PROBLEMA**

**PLANTEAMIENTO DEL PROBLEMA**

El desarrollo de sistemas autónomos con inteligencia artificial (IA) plantea profundos dilemas éticos a nivel global, cuestionando los fundamentos mismos de la responsabilidad humana en la toma de decisiones automatizadas. Como señala Bostrom (2014), *"la IA avanzada podría actuar de formas que no fueron anticipadas por sus creadores, generando consecuencias irreversibles"* (p. 22). Esto es especialmente crítico en ámbitos como la conducción autónoma, la medicina o la defensa, donde errores algorítmicos pueden costar vidas.

La falta de marcos normativos internacionales coherentes agrava esta incertidumbre, generando riesgos de asimetrías regulatorias entre países con distintos estándares éticos. Además, la IA autónoma desafía nociones tradicionales de agencia moral, ya que los sistemas pueden operar como *"cajas negras"*, dificultando la rendición de cuentas. Según Mittelstadt et al. (2016), *"la opacidad algorítmica socava la transparencia, un pilar fundamental de la ética en sociedades democráticas"* (p. 5). Esto tensiona principios universales como la justicia y la autonomía individual, especialmente cuando estos sistemas replican sesgos históricos en áreas como el reclutamiento laboral o la vigilancia masiva.

Finalmente, la escalada en la adopción de IA autónoma podría concentrar poder en actores con mayor acceso a datos y capacidad computacional, exacerbando desigualdades globales. Como advierte Zuboff (2019), *"el capitalismo de vigilancia utiliza la IA para extraer valor de los individuos sin su consentimiento genuino"* (p. 105). Este escenario exige replantear cómo distribuir los beneficios y riesgos de la IA de manera equitativa, evitando que su desarrollo profundice brechas socioeconómicas.

En el ámbito institucional, empresas y gobiernos enfrentan el desafío de implementar sistemas autónomos éticamente robustos sin sacrificar innovación. Como destacan Jobin et al. (2019), *"aunque el 84% de las organizaciones declaran adherirse a principios éticos en IA, solo el 18% los implementa en la práctica"* (p. 389). Esta brecha entre teoría y acción es evidente en algoritmos de alto impacto social, como los usados en préstamos bancarios o diagnósticos médicos, donde la opacidad genera desconfianza pública.

La rendición de cuentas también sigue siendo ambigua. Según Bryson y Winfield (2017), *"sin mecanismos claros de atribución de responsabilidad, las instituciones pueden eludir su papel en los daños causados por sistemas autónomos"* (p. 2). Casos como los accidentes de vehículos autónomos revelan vacíos legales que exigen protocolos de gobernanza internos, incluyendo comités éticos y mecanismos de reparación para víctimas.

Diagnóstico del investigador (Observaciones en startups tecnológicas venezolanas):  
En empresas que utilizan plataformas como LinkedIn o herramientas automatizadas de reclutamiento, se identificaron tres problemas clave:

1. Filtros ocultos en ATS: Los candidatos desconocen los criterios automatizados (ej.: palabras clave o puntajes) que descartan sus postulaciones.
2. Sesgos no auditados: Los algoritmos replican discriminaciones inadvertidas (ej.: preferencia por universidades o género).
3. Asimetría informativa: Las startups priorizan eficiencia sobre transparencia, sin explicar cómo se toman decisiones que afectan la empleabilidad.

#### **Interrogantes**

* ¿Cuáles son los principales dilemas éticos en el diseño y uso de sistemas autónomos con IA?
* ¿Qué desafíos técnicos, legales y sociales surgen con la adopción de estas tecnologías?
* ¿Cómo puede diseñarse un modelo de ética en IA que garantice transparencia, justicia y alineación con el bienestar social?

**Objetivo General**  
 Proponer un modelo de ética que asegure la implementación transparente, justa y alineada con el bienestar social de sistemas autónomos con inteligencia artificial.

**Objetivos Específicos**

* Diagnosticar los dilemas éticos fundamentales en el diseño y uso de sistemas autónomos con IA.
* Determinar los desafíos técnicos, legales y sociales en la adopción de estas tecnologías.
* Diseñar un modelo de ética aplicable a sistemas autónomos, integrando principios de transparencia, equidad y rendición de cuentas.

#### 

#### **Justificación**

El desarrollo de esta investigación sobre los dilemas éticos en sistemas autónomos con inteligencia artificial se justifica desde múltiples dimensiones, dada la urgencia y trascendencia del tema en el contexto tecnológico actual:

**Relevancia académica y teórica:** Este estudio contribuye a cerrar la brecha identificada por Jobin et al. (2019) entre los principios éticos declarados y su implementación real en sistemas de IA. Al analizar los fundamentos filosóficos de la responsabilidad en sistemas autónomos, la investigación enriquece el debate sobre agencia moral en la era algorítmica, abordando lo que Mittelstadt et al. (2016) denominan "la crisis de transparencia" en sociedades digitales. El trabajo sistematiza los marcos éticos existentes y propone adaptaciones para contextos específicos, avanzando en lo que Bryson y Winfield (2017) identifican como una necesidad crítica: mecanismos claros de atribución de responsabilidad.

**Aporte práctico e institucional:**Para organizaciones y gobiernos, la investigación provee herramientas concretas para:

1. Diseñar protocolos de auditoría algorítmica que mitiguen los sesgos identificados en procesos de reclutamiento
2. Establecer estándares de transparencia para sistemas de toma de decisiones automatizadas
3. Desarrollar marcos de gobernanza que equilibren innovación y protección de derechos fundamentales

El diagnóstico en startups venezolanas revela necesidades específicas en economías emergentes, donde la adopción acelerada de IA ocurre frecuentemente sin salvaguardas éticas adecuadas. Los hallazgos permitirán a estas empresas implementar sistemas más justos en áreas críticas como reclutamiento laboral.

**Impacto social:** La investigación aborda directamente lo que Zuboff (2019) denuncia como "capitalismo de vigilancia", proponiendo mecanismos para:

* Proteger la autonomía individual frente a sistemas opacos
* Prevenir la discriminación algorítmica en acceso a oportunidades
* Reducir asimetrías de poder en la economía digital

En el contexto venezolano, el estudio es particularmente pertinente porque:

1. La creciente adopción de herramientas automatizadas en procesos laborales carece de supervisión ética
2. Los sesgos en algoritmos pueden exacerbar desigualdades estructurales
3. No existen marcos regulatorios específicos para IA

**Innovación metodológica:**La investigación integra:

* Análisis filosófico de la responsabilidad moral
* Evaluación técnica de sistemas algorítmicos
* Propuestas normativas basadas en estudios de caso reales

Este enfoque multidisciplinario responde a lo que Bostrom (2014) identifica como necesidad fundamental: abordar los desafíos éticos de la IA desde perspectivas complementarias.

**CAPÍTULO II**

**MARCO TEÓRICO**

**ANTECEDENTES**

**Antecedente Internacional**

El estudio de Degli-Esposti (2023), publicado por CSIC y Catarata en el libro "La ética de la inteligencia artificial", tiene como objetivo general examinar las controversias emergentes en torno a las implicaciones sociales y los retos morales de la IA.

Su metodología se basa en el análisis de las implicaciones sociales y los retos morales de la IA, con un tipo de investigación cualitativo, presentando como foco de estudio las reflexiones sobre la confianza que depositamos en esta tecnología y en quienes la desarrollan.

Este estudio ofrece un contexto amplio sobre los debates éticos generados por la IA en la sociedad. Destaca la necesidad de reflexionar sobre la confianza que depositamos en esta tecnología y sus desarrolladores, lo cual puede informar la comprensión de las dimensiones éticas y sociales de la IA.

**Antecedente Internacional**

El estudio de Gómez Cárdenas, Fuentes Penna y Castro Rascón (2024), publicado en Ciencia Latina Revista Científica Multidisciplinar, bajo el título "El Uso Ético Y Moral De La Inteligencia Artificial En Educación E Investigación", tuvo como propósito central analizar los desafíos éticos y morales que plantea la adopción de la IA en los ámbitos educativo y de investigación.

La estrategia metodológica de este trabajo se basó en una revisión bibliográfica y el análisis de los desafíos éticos y morales que plantea la adopción de la IA en los ámbitos educativo y de investigación. El tipo de investigación se definió como investigación documental. La población y muestra comprendió artículos científicos y análisis sobre la implementación de IA en la educación e investigación.

Este antecedente examina específicamente los desafíos éticos y morales de la IA en la educación e investigación. Subraya la importancia de abordar aspectos como la transparencia, los sesgos, la privacidad de los datos y el plagio académico, lo cual es crucial para analizar el uso ético de la IA en estos campos.

**Antecedente Nacional**

El estudio de Jiménez (2024), realizado en la Universidad Pedagógica Experimental Libertador, Venezuela, y titulado "Deontología en el uso de la inteligencia artificial: Hacia un futuro responsable", se propuso explorar la importancia de establecer una deontología o código de conducta ética para el desarrollo y la aplicación responsable de la IA.

La metodología empleada fue de revisión documental, tipificado como un análisis teórico. El foco del estudio se centró en la exploración de la importancia de establecer una deontología para el desarrollo y la aplicación de la IA. El tipo de investigación fue teórica y la fuente de información comprendió documentos académicos y reflexiones teóricas sobre la ética en el uso de la IA.

El estudio concluye que es urgente crear un marco ético sólido que guíe el uso de la IA en la sociedad, permitiendo aprovechar sus beneficios y mitigar sus riesgos. Estas ideas sugieren que, aunque la IA ofrece muchas ventajas, también plantea desafíos importantes relacionados con la responsabilidad y la ética en su aplicación.

**BASES TEÓRICAS**

**Principios Éticos:**

John Stuart Mill, en su obra "El Utilitarismo", aborda la moralidad y los principios que deben regir las acciones. Si bien no define "principios éticos" como un término, su enfoque se centra en el principio de la mayor felicidad.

"El credo que acepta como fundamento de la moral la Utilidad, o el Principio de la Mayor Felicidad, sostiene que las acciones son correctas en la medida en que tienden a promover la felicidad, e incorrectas en cuanto tienden a producir lo contrario de la felicidad. Por 'felicidad' se entiende el placer y la ausencia de dolor; por 'infelicidad' el dolor y la privación de placer." — John Stuart Mill, *El Utilitarismo*

### **El Desarrollo de Sistemas Autónomos con IA: Un Desafío Ético Fundamental**

El desarrollo de sistemas autónomos con IA plantea desafíos éticos que van más allá de las aplicaciones educativas. A medida que estos sistemas se integran en diversos aspectos de la sociedad, como el transporte, la atención médica y la toma de decisiones, surgen preguntas críticas sobre la responsabilidad, la rendición de cuentas y los valores que deben guiar su diseño y uso (Bostrom, 2014).

### **La Necesidad de Marcos Regulatorios y Éticos**

La falta de marcos regulatorios y éticos claros para el desarrollo de sistemas autónomos con IA agrava la incertidumbre y aumenta el riesgo de consecuencias no deseadas. Es fundamental establecer principios y directrices que aborden cuestiones como la transparencia, la explicabilidad, la justicia y la equidad en el diseño y la implementación de estos sistemas.

### 

### **Implicaciones Éticas en Diferentes Contextos**

Las implicaciones éticas del desarrollo de sistemas autónomos con IA varían según el contexto en el que se apliquen. En el ámbito educativo, por ejemplo, el uso de IA para la toma de decisiones sobre el aprendizaje de los estudiantes plantea preguntas sobre la autonomía, la equidad y el papel del docente. En otros contextos, como el transporte o la atención médica, las decisiones automatizadas pueden tener implicaciones aún más graves para la seguridad y el bienestar humano.

### **El Contexto Venezolano y los Sistemas Autónomos con IA**

En Venezuela, al igual que en otros países en desarrollo, la adopción de sistemas autónomos con IA se enfrenta a desafíos particulares relacionados con las limitaciones infraestructurales, la desigualdad en el acceso a la tecnología y la falta de marcos regulatorios adecuados. Es crucial considerar estos factores al analizar las implicaciones éticas del desarrollo de estos sistemas en el contexto venezolano.

### **Planes de Acción para un Desarrollo Ético de Sistemas Autónomos con IA**

Para abordar los desafíos éticos planteados por el desarrollo de sistemas autónomos con IA, se propone el desarrollo de planes de acción que promuevan un enfoque ético y responsable, que incluyan:

* **Fomento de la investigación interdisciplinaria:** Promover la colaboración entre expertos de diferentes campos, como la ética, la informática, el derecho y las ciencias sociales, para abordar las complejas cuestiones éticas relacionadas con los sistemas autónomos con IA.
* **Desarrollo de marcos éticos y regulatorios:** Establecer principios, directrices y regulaciones claras que guíen el diseño, el desarrollo y la implementación de sistemas autónomos con IA, garantizando que se respeten los valores humanos fundamentales.
* **Promoción de la transparencia y la explicabilidad:** Exigir que los sistemas autónomos con IA sean transparentes en su funcionamiento y que sus decisiones puedan ser explicadas y justificadas.
* **Inclusión de la perspectiva de los usuarios y las comunidades afectadas:** Garantizar que el desarrollo de sistemas autónomos con IA tenga en cuenta las necesidades, los valores y las preocupaciones de los usuarios y las comunidades que se verán afectadas por su uso.
* **Educación y concientización:** Promover la educación y la concientización sobre las implicaciones éticas de la IA entre los desarrolladores, los responsables de la toma de decisiones y el público en general.

Estos planes de acción buscan garantizar que el desarrollo de sistemas autónomos con IA se realice de manera ética y responsable, maximizando sus beneficios potenciales y minimizando sus riesgos.

**BASES LEGALES**

En Venezuela, el marco legal para el desarrollo y la aplicación de la Inteligencia Artificial se fundamenta en varios instrumentos jurídicos, aunque carece de una legislación específica y exhaustiva sobre la IA. A continuación, se presentan las principales bases legales que, de manera directa o indirecta, regulan el uso de esta tecnología:

### **Constitución de la República Bolivariana de Venezuela (1999)**

La Carta Magna establece los principios fundamentales que deben guiar el desarrollo tecnológico en el país. En particular, los siguientes artículos son relevantes:

* **Artículo 102:** Derecho a la educación, que debe promover el desarrollo del potencial creativo de cada ser humano y el pleno ejercicio de su personalidad en una sociedad democrática. La IA debe ser utilizada como herramienta para mejorar la calidad de la educación y garantizar la igualdad de oportunidades.
* **Artículo 110:** Reconoce el interés público de la ciencia, la tecnología y la innovación, y establece la obligación del Estado de promoverlas y apoyarlas. El Estado debe garantizar que el desarrollo de la IA se realice en beneficio del interés general y respetando los derechos humanos.

### **Leyes que regulan aspectos relacionados con la IA**

Aunque no existe una ley específica sobre la IA, diversas leyes regulan aspectos que son relevantes para su desarrollo y aplicación:

* **Ley Orgánica de Protección de Datos Personales:** Establece los principios, derechos y obligaciones que rigen el tratamiento de datos personales, incluyendo los datos utilizados por sistemas de IA. Es fundamental garantizar la privacidad y seguridad de los datos en el desarrollo y uso de la IA.
* **Ley Especial contra los Delitos Informáticos:** Tipifica y sanciona los delitos cometidos a través de medios informáticos, incluyendo aquellos que puedan ser perpetrados mediante el uso de la IA. Esta ley es relevante para prevenir y sancionar el uso malicioso de la IA.
* **Ley de Propiedad Intelectual:** Protege las creaciones intelectuales, incluyendo el software y los algoritmos utilizados en sistemas de IA. Es importante garantizar la protección de los derechos de propiedad intelectual en el desarrollo de la IA.
* **Código Civil:** Establece los principios generales de la responsabilidad civil, que pueden ser aplicables en casos de daños causados por sistemas de IA autónomos. Se requiere una adaptación de estos principios para abordar los desafíos específicos de la IA.

### **Marcos regulatorios y éticos**

Además de las leyes, existen iniciativas y propuestas para desarrollar marcos regulatorios y éticos específicos para la IA. Estos marcos buscan establecer principios y directrices que orienten el desarrollo y uso responsable de la IA, abordando cuestiones como:

* **Transparencia y explicabilidad:** Los sistemas de IA deben ser transparentes en su funcionamiento y sus decisiones deben ser explicables.
* **Responsabilidad y rendición de cuentas:** Se deben establecer mecanismos para determinar la responsabilidad en caso de daños causados por sistemas de IA.
* **No discriminación e igualdad:** Los sistemas de IA deben ser diseñados y utilizados de manera que no perpetúen ni amplifiquen la discriminación.
* **Seguridad y privacidad:** Se deben garantizar la seguridad de los datos y la privacidad de los individuos en el desarrollo y uso de la IA.

**Definición de Términos Básicos**

**Algoritmo:** "Un conjunto finito de reglas bien definidas para resolver un problema en un número finito de pasos." - **Donald Knuth**, en su obra seminal "The Art of Computer Programming".

**Autonomía:** "La capacidad de un agente para actuar por sí mismo, tomar decisiones y controlar sus acciones sin intervención externa directa en cada paso." - **Luciano Floridi**, en "The Ethics of Information".

**IA (Inteligencia Artificial):** "La ciencia e ingeniería de hacer máquinas inteligentes, especialmente programas de computadora inteligentes. La IA está relacionada con la tarea similar de usar computadoras para comprender la inteligencia humana, pero la IA no tiene que limitarse a métodos que son biológicamente observables." - **John McCarthy**, en un memorando de 1955 considerado el documento fundacional del campo.

**Marco Normativo:** "Un conjunto de principios, leyes, regulaciones y directrices que establecen los límites y las expectativas para el comportamiento y las prácticas dentro de un dominio específico." - Esta definición es más funcional y comúnmente entendida en el ámbito legal y de políticas públicas, aunque autores como **Ronald Dworkin** en "Law's Empire" exploran la naturaleza de los principios y reglas en los sistemas legales, que son componentes clave de un marco normativo.

**Opacidad Algorítmica:** "La dificultad o imposibilidad práctica de comprender cómo un algoritmo llega a una decisión o resultado particular." - **Frank Pasquale**, en "The Black Box Society: The Secret Algorithms That Control Money and Information".

**Responsabilidad:** "La obligación de responder por una acción, decisión o resultado, incluyendo la disposición a ser culpado o elogiado por ello." - Si bien la responsabilidad es un concepto filosófico y ético amplio, **H.L.A. Hart** en "Punishment and Responsibility" ofrece un análisis profundo de sus implicaciones legales y morales. En el contexto de la tecnología, se busca adaptar estas ideas a los sistemas automatizados.

**Sistema Autónomo:** "Un sistema capaz de realizar tareas y tomar decisiones sin intervención humana directa y continua." - **Peter Asaro**, en su trabajo sobre ética robótica, explora las implicaciones de la creciente autonomía en los sistemas tecnológicos.

**Sesgo:** "Una desviación sistemática de la norma o la verdad, que puede llevar a resultados injustos o inexactos." - En el contexto de la IA, este concepto se ha popularizado, y autoras como **Cathy O'Neil** en "Weapons of Math Destruction" ilustran cómo los sesgos en los datos pueden perpetuar desigualdades.

**Transparencia:** "La cualidad de ser abierto, comprensible y fácilmente accesible en cuanto a la naturaleza, el funcionamiento y la lógica subyacente de un sistema o proceso." - En el ámbito de la IA, la transparencia es un tema de debate crucial, y autores como **Tim Miller** han investigado sobre la interpretabilidad de los modelos de aprendizaje automático como un aspecto de la transpa

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Ítems | Totalmente de acuerdo | De acuerdo | Indiferente | En desacuerdo | Totalmente en desacuerdo |
| Variable: Los Principios Éticos | | | | | |
| Dimensión: | | | | | |
| 1 - Es fundamental establecer principios y directrices claras que guíen el diseño, desarrollo e implementación de sistemas autónomos con IA para garantizar el respeto de los valores humanos fundamentales. |  |  |  |  |  |
| 2 - Considero que la transparencia y la explicabilidad del funcionamiento de los algoritmos son pilares fundamentales de la ética en las sociedades democráticas. |  |  |  |  |  |
| 3 - Es necesario que los marcos éticos para la IA aborden la atribución de responsabilidad en caso de daños causados por sistemas autónomos, incluso cuando los sistemas operan como "cajas negras". |  |  |  |  |  |
| Variable: Inteligencia Artificial | | | | | |
| Dimensión: | | | | | |
| 4 - El desarrollo de la IA avanzada podría generar consecuencias irreversibles si no se anticipan adecuadamente sus acciones. |  |  |  |  |  |
| 5 - La concentración de poder en actores con mayor acceso a datos y capacidad computacional debido a la adopción de la IA autónoma podría exacerbar las desigualdades globales. |  |  |  |  |  |
| 6 - En el contexto de la IA, garantizar la seguridad de los datos y la privacidad de los individuos es fundamental. |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Objetivo General: Proponer un modelo de ética que asegure la implementación transparente, justa y alineada con el bienestar social de sistemas autónomos con inteligencia artificial.** | | | | | | |
| Objetivos Específicos | Variable | Variable Conceptual | Dimensión | Indicadores | Técnica e Instrumento | Items |
| Diagnosticar los dilemas éticos fundamentales en el diseño y uso de sistemas autónomos con IA. | Principios Éticos | *"Los principios éticos son aquellas normas básicas que permiten dirimir conflictos y orientar la acción en una sociedad pluralista"* **(Cortina, *Ética mínima*, 1986, p. 52)**. | Responsabilidad y Rendición de Cuentas | Atribución de responsabilidad | Revisión Documental, Encuesta (Cuestionario) | 1, 2 y 3 |
| Determinar los desafíos técnicos, legales y sociales en la adopción de estas tecnologías. | Inteligencia Artificial | *"La ciencia e ingeniería de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes."* **(McCarthy, *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955)**. | Establecimiento de Principios | Sesgos algorítmicos, desigualdad | Revisión Documental, Encuesta (Cuestionario) | 4, 5 y 6 |
| **No Se Operacionaliza** | | | | | | |

# **REFERENCIAS BIBLIOGRÁFICAS**

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Floridi, L., et al. (2018). AI4People—An Ethical Framework for a Good AI Society. *Nature Human Behaviour*.

Mittelstadt, B., et al. (2016). The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*.

Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs.

Jobin, A., et al. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*.

Wachter, S., et al. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist. *Harvard Journal of Law & Technology*.

Degli-Esposti, S. (2023). *La ética de la inteligencia artificial*. CSIC; Catarata.

Gómez Cárdenas, R., Fuentes Penna, A., & Castro Rascón, A. (2024). El Uso Ético Y Moral De La Inteligencia Artificial En Educación E Investigación. *Ciencia Latina Revista Científica Multidisciplinar*, *8*(5), 3243-3262.<https://doi.org/10.37811/cl>\_rcm.v8i5.13801

Jiménez, Y. (2024). Deontología en el uso de la inteligencia artificial: Hacia un futuro responsable. *Dialógica, Revista Multidisciplinaria*, *21*(1), 66-83.

Asaro, P. M. (2011). *Robotics and ethics: A naviagation through the philosophical landscape*. AI and Society, *26*(4), 353-372.

Dworkin, R. (1986). *Law's empire*. Belknap Press of Harvard University Press.

Floridi, L. (2014). *The ethics of information*. Oxford University Press.

Hart, H. L. A. (2008). *Punishment and responsibility: Essays in the philosophy of law* (2nd ed.). Oxford University Press.

Knuth, D. E. (1997). *The art of computer programming, volume 1: Fundamental algorithms* (3rd ed.). Addison-Wesley.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1-38.

O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.