

Registro de la Propuesta del Trabajo Fin de Grado (TFG)

NOMBRE DEL ALUMNO: Coleta Múgica Lecuona

PROGRAMA: E-6 (ADE + RRII)

GRUPO: 5º B

FECHA: 01/11/2020

Director Asignado: Fernández Fernández, José Luis
Apellidos Nombre

Título provisional del TFG

Ensayo y Erro: Máquinas Automatizadas y su Vacío Ético

ADJUNTAR PROPUESTA (máximo 2 páginas: objetivo, bibliografía, metodología e índice preliminares)

Firma del estudiante:

Firma del Director:

Fecha:

El **objetivo** de este trabajo es analizar más a fondo el funcionamiento del Machine Learning dentro de la Inteligencia Artificial y los peligros que estos sistemas pueden tener sin un marco ético bien establecido. Además, a través de este trabajo cuestionaremos la implementación de la ética en cada paso de la cadena de aprendizaje de los algoritmos de estas máquinas. Los programas automatizados con el conocimiento aplicado tienen la capacidad de aprender de forma autónoma junto al uso de un gran volumen de datos sin intervención humana. Sin embargo, ¿Son los algoritmos capaces de aprender dentro de ciertos límites éticos? ¿Pueden considerar ciertas situaciones que ocultan los marcos teóricos? Estas preguntas serán estudiadas y se les dará una resolución aproximada en este ensayo.

Sobre la **metodología** para cumplir con el objetivo, hay que destacar que se seguirán y analizarán numerosas investigaciones en torno a estas tecnologías. Se estudiarán casos específicos relacionados con el aprendizaje automático y su ética y teniendo en cuenta las diversas conclusiones que estos ofrecen, se compararán para extraer una nueva conclusión sobre la viabilidad ética de las máquinas automatizadas y para ver cómo se pueden aplicar valores éticos en el proceso de aprendizaje de algoritmos. Además de esto, la investigación también observará en qué medida la ética ya se tiene en cuenta en el aprendizaje no supervisado y en el “**Deep Learning**”.

Bibliografía

1. Benthall S. & Haynes, BD. (2019) Racial categories in machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Atlanta.
2. Berk, R. A., Sorenson, S. B., & Barnes, G. (2016). Forecasting domestic violence: A machine learning approach to help inform arraignment decisions. *Journal of Empirical Legal Studies*, 13(1)
3. Binns, R. (2018) Fairness in Machine Learning: Lessons from Political Philosophy, *Proceedings of Machine Learning Research*
4. Wachter, S.; Mittelstadt, B.; Floridi, L. (2017) Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, *International Data Privacy Law*

5. Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2018). Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? *Philosophy & Technology*

Índice

1. Introducción

- 1.1. Objetivos
- 1.2. Justificación del tema
- 1.3. Metodología

2. Estado de la cuestión

- 2.1. ¿Qué es la Inteligencia Artificial?
- 2.2. ¿Qué es Machine Learning dentro de la IA?
- 2.3. ¿Qué es el Aprendizaje Profundo dentro del Aprendizaje Automático?
- 2.4. ¿Cuál es el papel de la ética en la IA?

3. El Aprendizaje Automatizado y la Ética

- 3.1. Historia de la dimensión ética en Machine Learning
- 3.2. Vacío de un marco legal sobre la Ética de Machine Learning

4. Casos que apoyan la capacidad de Machine Learning para actuar éticamente

- 4.1. Caso 1
- 4.2. Caso 2
- 4.3. Caso 3

5. Casos que socavan la capacidad de Machine Learning para actuar éticamente

- 5.1. Caso 1 → hong kong (se puede ver quien es quien)
- 5.2. Caso 2 → facebook manipulacion
- 5.3. Caso 3

6. Conclusiones

7. Bibliografía