MANUAL DE ÉTICA APLICADA EN INTELIGENCIA ARTIFICIAL

MÓNICA VILLAS OLMEDA
JAVIER CAMACHO IBÁÑEZ

ANAYA MULTIMEDIA

ÍNDICE DE CONTENIDOS

2 ÉTICA Y TECNOLOGÍA

DE QUÉ TRATA ESTE CAPÍTULO

En este capítulo vamos a profundizar en la relación entre ética y tecnología, y más concretamente exploraremos la cuestión de por qué tiene sentido hablar de ética con relación a la inteligencia artificial y los datos. El objetivo es comprender cuál es la dimensión ética en los sistemas de IA y de qué herramientas disponemos para poder gestionarla.

La ética se concibe a menudo como un freno, que puede ralentizar la innovación, pero, como veremos, se trata precisamente de lo contrario. La función de los frenos en los automóviles es realmente la de permitir desplazarnos a mayor velocidad, con la confianza de disponer de un mecanismo de seguridad en caso de que ocurriera algún incidente. Esto es precisamente lo que posibilita la ética con relación a la tecnología: innovar con seguridad y confianza.

Revisaremos diferentes perspectivas acerca de la filosofía de la tecnología, sobre si la tecnología es neutral o no, para que el lector vaya construyendo su propia visión. De igual manera, realizaremos una introducción al concepto de ética y a cómo podemos gestionar la dimensión ética en la resolución de problemas y en la toma de decisiones. Para ello, propondremos un proceso sencillo en cuatro fases: ver, pensar, elegir y actuar.

Por último, exploraremos algunas características específicas de los aspectos éticos relativos a los sistemas de IA, y se introducirá el concepto de "ética por diseño".

¿QUÉ SON LOS PRINCIPIOS?

El origen de la palabra "ética" viene del vocablo griego *ethos*. La ética, como se ha mencionado en capítulos anteriores, se define según la RAE como "la parte de la filosofía que trata del bien y del fundamento de sus valores", pero también como "el conjunto de normas morales que rigen en el ámbito de nuestra vida la conducta de las personas". Como ejemplos habituales en los que se habla de ética, nos encontramos con ética profesional, ética periodística, ética deportiva, etc., un sinfín de ellas que reflejan, como ha sido siempre, una preocupación humana desde su primera utilización por Sócrates hace ya más de 2.000 años.

Es muy habitual la aparición de nuevos dilemas éticos con determinadas evoluciones tecnológicas. Estas pueden suponer cambios beneficiosos para toda la sociedad, pero también pueden causar un perjuicio a las personas si no son utilizadas de la manera adecuada o, dicho de otro modo, si no se considera la ética conjuntamente con la evolución tecnológica. Este podría ser el caso de numerosas tecnologías, como la energía nuclear, las cámaras de detección de personas, los coches autónomos o la propia inteligencia artificial, entre otras.

Los principios aplicados a la IA nos van a ayudar a reconocer las consideraciones éticas a tener en cuenta cuando se usa esta tecnología para la toma de decisiones autónomas. No es sencillo definir en qué consiste un principio; por eso, vamos a basarnos en el glosario para sistemas inteligentes y autónomos de la organización IEEE (*Institute of Electrical and Electronics Engineers*, Instituto de Ingenieros Eléctricos y Electrónicos),[1] que incluye diferentes definiciones teniendo en cuenta el contexto en el que se están utilizando (disciplinas computacionales, ingeniería, economía, filosofía o regulaciones internacionales):

- "Un principio es una regla fundamental, un estándar o precepto en materia de moralidad o conducta de una persona".
- "Una proposición que se considera tan fundamental y obvia que no necesita defensa ni apoyo".
- "Y, en las ciencias empíricas, una declaración de una regularidad establecida, similar a una ley".
- "Una regla o norma especialmente de buen comportamiento".

1. https://standards.ieee.org/content/dam/ieeestandards/standards/web/documents/other/ead1e_glossary.pdf.

Revisando estas referencias, se pueden definir los principios éticos aplicados a la IA como aquellos que nos ayudan a preservar los derechos y libertades de las personas, sin frenar la innovación tecnológica. En lo que se refiere a derechos y libertades, está claro que los principios, sean cuales sean, van a estar muy influidos por la Declaración de los Derechos Humanos de las Naciones Unidas de 1948. La consecuencia más importante de este acuerdo es que la mayoría de los derechos que aparecen en este documento forman parte hoy de las leyes constitucionales de los países democráticos. Como para estos principios nos referimos a una tecnología muy concreta, la IA, es también clave el conocimiento de cómo funciona esta tecnología, así como los ámbitos de aplicación de esta, no es lo mismo diseñar aplicaciones de IA para recomendación de música que hacerlo para reconocimiento visual de objetivos militares.

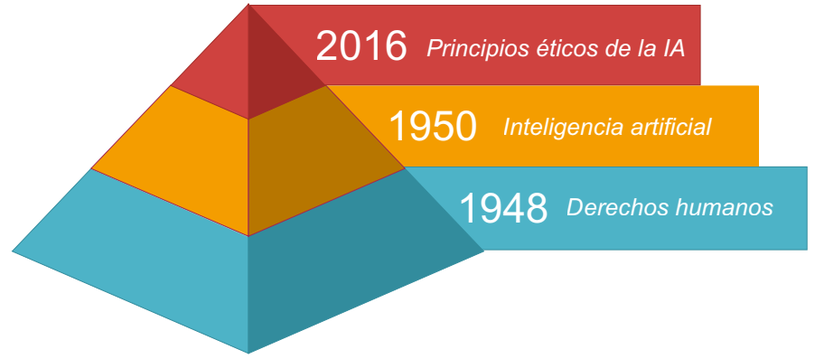

Figura 3.1. Principios, inteligencia artificial y derechos humanos.

Para entender un poco mejor algunos ejemplos de principios de ética para IA, así como la agrupación de estos en distintas categorías, vamos a tomar como base el estudio de Alethicslab.[2] Alethicslab es un laboratorio independiente, creado en 2017 con el objetivo de analizar los retos éticos en IA y proporcionar recomendaciones útiles para los investigadores y desarrolladores. El estudio incluye una recopilación de más de 100 principios, que han sido encontrados en documentos publicados hasta mayo de 2021, y se mantiene actualizado periódicamente, lo que es de gran utilidad para los profesionales que trabajan en ética e IA. Además de proporcionar un detalle de los documentos recopilados, su página web incluye una representación gráfica

2. https://aiethicslab.com/big-picture/.

Responsabilidad en el ejercicio de una profesión

Por definición, toda profesión[3] maneja un conocimiento valioso y unas capacidades, competencias y habilidades como un "grupo" o colectivo en la sociedad. Cada profesión tiene una utilidad propia para la sociedad, ya que de lo contrario no existiría.[4] Esta utilidad podríamos también denominarla como el "bien intrínseco" de la profesión. A través de la formación y del desempeño exitoso de dicho bien intrínseco, cada colectivo de profesionales es reconocido por la sociedad, se gana la aceptación de su profesión, adquiere cierto estatus y reconocimiento y, claro está, recibe un beneficio económico. En este caso, el colectivo de profesionales (sean médicos, científicos, arquitectos, ingenieros, desarrolladores de software, etc.) será responsable de ofrecer sus productos y servicios a la sociedad para cumplir con ese bien intrínseco y para "responder"[5] a la legitimación recibida de la sociedad y dar el uso esperado a sus conocimientos y capacidades específicas.

Por otra parte, estos colectivos o grupos de profesionales desarrollan mecanismos de asociación (como es el caso de los colegios o asociaciones profesionales) con el fin de supervisar el ejercicio de su profesión, armonizar criterios y desarrollar mecanismos de autorregulación. En muchos casos, por el impacto que una profesión tiene en un país, también los estados desarrollan una regulación específica para ciertas profesiones, tanto para el "acceso" a dicha profesión (por ejemplo, en el caso de oposiciones públicas) como para su ejercicio. Como veremos más adelante en el capítulo, la responsabilidad en el ejercicio de una profesión tiene una correlación alta con la valoración de riesgos asociados a dicha profesión y por tanto con la previsible regulación para gestionar dichos riesgos. Un ejemplo claro sería el sector financiero, que, por su impacto, es de los más regulados y donde más regulación nueva se desarrolla. No es sorprendente, por tanto, que existan ya iniciativas para regular el campo de la inteligencia artificial, por el previsible impacto que va a suponer a nivel global.

3. Aludimos a la definición de profesión como servicio a la sociedad único, definido y esencial, considerado como una vocación, basado en conocimientos y técnicas de carácter intelectual, que requiere un período previo de preparación especializada y que demanda un amplio campo de autonomía (Bilbao, Fuertes y Guibert, 2006).
4. Dejamos al lector la reflexión e investigación sobre algunas profesiones y oficios ya extinguidos u otros en riesgo de desaparición. Al mismo tiempo, surgen nuevas profesiones como consecuencia de los cambios tecnológicos.
5. Recordemos la interpretación de la "respons-abilidad" como la capacidad de dar respuesta frente a determinadas circunstancias.

SOBRE LA RESPONSABILIDAD Y LA REPUTACIÓN PROFESIONAL

Observemos por un momento la imagen de la figura 4.1.

Figura 4.1. Carretera "pintada".

Pensemos ahora que, en vez de tratarse de "cierta" negligencia a la hora de pintar las líneas de una carretera, pudiera tratarse de un informe, un software, una subrutina, una propuesta de especificaciones, una respuesta a un pliego de condiciones, etc. Es casi seguro que podamos pensar en alguien de nuestro entorno (centro educativo, empresa, institución, departamento) que, con mucha probabilidad, pudiera haber procedido de esa manera, pero también es seguro que podemos pensar en alguien del mismo entorno por quien pondríamos la mano en el fuego y no creeríamos que esa persona es capaz de tal cosa.

Este sencillo ejemplo nos puede ayudar a reflexionar sobre nuestro comportamiento responsable como profesionales, y cómo nuestras decisiones y acciones afectan a nuestra reputación profesional (y personal). De ahí que la responsabilidad sea un valor tan apreciado por todas las empresas y organizaciones.

Responsabilidad y tecnología

En el capítulo anterior, tratamos la cuestión de la ética y la tecnología y cómo la tecnología cobra un especial significado ético. Las circunstancias actuales, en las que el *homo faber* se sitúa por encima del *homo sapiens* e incluso se perfilan

5 PRIVACIDAD

DE QUÉ TRATA ESTE CAPÍTULO

En este capítulo vamos a tratar una de las cuestiones ineludibles, y más actuales, de la ética con relación a la gestión de los datos, la información y los sistemas de inteligencia artificial. Hoy en día, la privacidad es la preocupación principal de los CIO, CISO, DPO y demás rangos en el escalafón técnico, pero también en las áreas no técnicas, como, por ejemplo, de los responsables de cumplimiento normativo, los gabinetes jurídicos o de muchos responsables de unidades de negocio.

Las cuestiones de privacidad de los datos ocupan páginas y pantallas en prácticamente todos los medios de comunicación. El Reglamento General de Protección de Datos (RGPD) se ha convertido en conversación (y preocupación) habitual, tanto dentro como fuera de España y de Europa, pero hasta hace poco no había asumido un papel tan protagonista. Efectivamente, hasta ahora, los pilares de la ciencia económica y de la competitividad eran el capital, la tierra, el trabajo o la tecnología. Pero un quinto factor, los datos, ha alcanzado la misma relevancia que los anteriores y, por tanto, es considerado como un activo esencial y una fuente de ventaja competitiva por parte de las compañías y corporaciones, las instituciones, los gobiernos y los países.

En este capítulo abordaremos el concepto de privacidad y sus principales características. Exploraremos asimismo los términos de integridad contextual, transparencia estructurada, privacidad por diseño y algunas de las técnicas más

datos reales. La calidad de la comparativa se utiliza de manera iterativa en la primera red para generar otro modelo de datos sintéticos, hasta que el índice de comparación supera el umbral establecido.

Ingeniería de la privacidad

En base a los conceptos y marcos presentados, tales como la privacidad por diseño, la integridad contextual o la transparencia estructurada, la ingeniería de la privacidad va a consistir en el proceso de traducir y aplicar estos para cada caso de uso, mediante una metodología específica y escogiendo el set de herramientas y técnicas más apropiadas para cada caso. Esto implica adaptar las tareas clásicas de planificación, desarrollo, control y mejora, al terreno de la privacidad, comprendiendo los objetivos que se persiguen, las políticas y requisitos de privacidad específicos a tener en consideración, desarrollar los correspondientes procesos y procedimientos, impartir formación y disponer de indicadores y mecanismos para asegurar el cumplimiento de dichos procesos y los niveles de servicio o calidad efectivos.

Las diferentes herramientas y técnicas se deberán utilizar de manera selectiva o combinada, en función del caso de uso, para conseguir garantizar los niveles de privacidad objetivo. Como se ha visto anteriormente, entre estas herramientas destacan aquellas encaminadas a garantizar la privacidad de entrada, como la anonimización (y sus limitaciones), la encriptación homomórfica, la computación multiparte, el aprendizaje federado o el uso de datos sintéticos. Además, la privacidad diferencial juega un papel importante a la hora de garantizar la privacidad de salida, es decir, no poder inferir datos individuales a partir de los resultados del modelo. En todo caso, habrá que valorar los requisitos, ventajas e inconvenientes, aplicados a cada caso de uso.

A diferencia de lo que ocurre en otras áreas de *machine learning* o *deep learning*, donde el objetivo es disponer de un modelo rápidamente y luego ir ajustando sucesivamente los parámetros, cuando se trata de la privacidad, hay que ser más prudentes y responsables, ya que la pérdida o filtrado de datos, en cualquier etapa del desarrollo y explotación del modelo, puede dejar datos personales expuestos. Recientemente se ha extendido el uso de cuestionarios y auditorías de impacto en la privacidad, a la hora de desarrollar o valorar desarrollos o aplicaciones existentes.

EJEMPLO DE PREGUNTAS PARA CUESTIONARIO DE IMPACTO EN PRIVACIDAD[31]

Los siguientes son solo algunos ejemplos de preguntas extraídos del cuestionario de evaluación de impacto desarrollado por DataEthics.eu. Para cada una de las preguntas hay que responder SÍ (y explicar el cómo) o NO (y explicar el por qué):

- ¿Garantiza que los derechos del usuario son prioritarios, en lugar de los intereses comerciales o intereses institucionales?
- ¿Garantiza que, principalmente, los usuarios se beneficien de sus propios datos, y no solo la organización?
- ¿Utiliza los principios de privacidad por diseño y puede describirlos de forma clara y transparente?
- ¿Garantiza que los datos de los usuarios (en la medida de lo posible) se procesan directamente en los propios dispositivos de los usuarios?
- Cuando el tratamiento de los datos es necesario en un lugar distinto al de los propios dispositivos del usuario, como su servidor o una solución en la nube, ¿los datos recogidos no están relacionados con una persona identificable?
- ¿Utiliza los datos para predecir comportamientos a nivel individual o solo patrones?

RESUMEN DEL CAPÍTULO

- En este capítulo, hemos abordado uno de los conceptos claves y más actuales con relación a la ética y la IA: la privacidad.
- El derecho a la privacidad (aquello particular y personal de cada individuo) es un derecho universal, que no puede transferirse ni se puede renunciar a él.

31. `https://dataethics.eu/wp-content/uploads/dataethics-impact-assessment-2021.pdf`.

A continuación, en la tabla 6.1, se describen las métricas y fórmulas más usadas.

Tabla 6.1. Métricas y fórmulas más usadas.

Métrica	Fórmula	Nombre completo	Descripción
TPR	$\frac{TP}{TP+FN}$	**Tasa de verdaderos positivos (*True positive rate, Recall, Sensitivity)***	Este valor, junto con el siguiente, nos ayuda a medir cómo de bueno es el algoritmo discriminando entre casos positivos y negativos. En este caso, es el **porcentaje de positivos detectado correctamente** por el algoritmo.
TNR	$\frac{TN}{TN+FP}$	**Tasa de verdaderos negativos (*True negative rate, Especificity)***	Se llama especificidad, al contrario que el anterior, es el porcentaje de negativos detectado correctamente por el algoritmo.
ACC	$\frac{TP+TN}{TP+FN+FP+TN}$	**Exactitud (*Accuracy*)**	Es la **exactitud** y trata de medir lo cerca que ha estado una medición de su valor verdadero, es decir, es la cantidad de predicciones positivas que fueron correctas.
PPV	$\frac{TP}{TP+FP}$	**Porcentaje de casos positivos (*Positive predictive value, Precision)***	Precisión, que lo que indica es el porcentaje de casos positivos.
F1	$2*(\frac{PPV*TPR}{PPV+TPR})$	**F1 score**	Esta métrica nos resume la precisión y la sensibilidad en una sola métrica y es muy útil cuando las clases son desiguales.

EJEMPLO DE CONTRATACIÓN

Suponemos que tenemos una lista de 10 valores sobre la variable que queremos predecir, que es contratación de personas. Esta variable es la etiqueta de los datos y tiene el valor de contratado (en verde) y no contratado (en rojo). Para hacer la matriz de confusión, necesitamos comparar los valores predichos por el algoritmo con los valores reales. Con esto vamos calculando los valores de la matriz de confusión; una vez obtenida la matriz de confusión, se pueden ir obteniendo el resto de las métricas. En la figura 6.5, vemos un ejemplo.

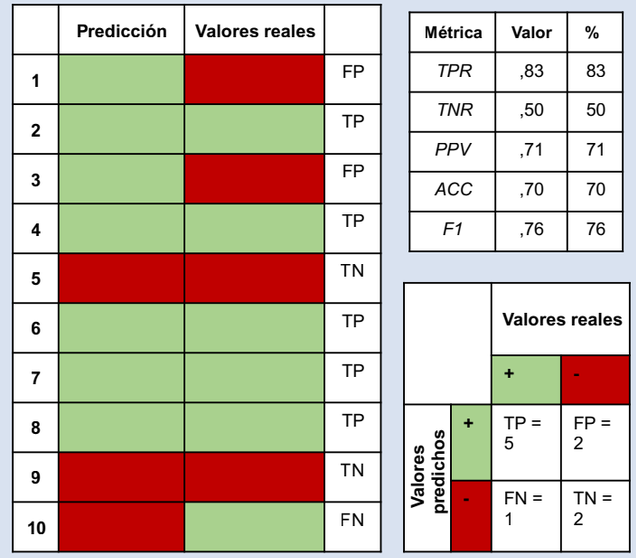

Figura 6.5. Ejemplo de contratación.

Además de las métricas explicadas con el ejemplo, es importante entender también el funcionamiento de la **curva ROC**, que se usa, además de la exactitud (*accuracy*), cuando los datos no están balanceados, y que va a permitir tomar decisiones sobre la idoneidad del modelo. Además de la curva ROC, en la figura 6.6, para representar TPR y FPR, se utiliza mucho la métrica AUC.

figura 7.23. En este caso, 5,93 es el valor predicho que haría que el vino se clasificara en una categoría concreta, teniendo en cuenta la aportación de los valores SHAP de cada uno de los atributos. Si no se tuviera esa información, el valor predicho sería el valor base 5,65, que aparece como **base value** en la gráfica.

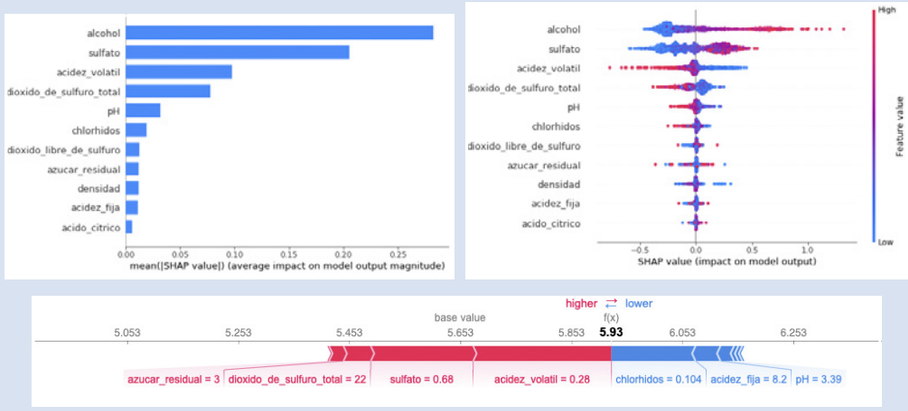

Figura 7.23. Valores de SHAP.

Saliency Map es un tipo de técnica que proviene de la rama de la visión artificial[44] y está muy orientada a proporcionar explicaciones para redes neuronales que trabajan con imágenes. "*Saliency*" se refiere a valores únicos de la imagen (píxeles, resolución o intensidad) en un contexto de procesamiento de imágenes. En este tipo de modelos, lo que se intenta conocer es qué nodos, conexiones o atributos dentro de una red neuronal han llevado a una predicción concreta, o bien qué pasaría si ciertos nodos o conexiones no existieran. En la figura 7.24, la parte de la derecha sería el "*saliency map*", que es básicamente la parte de esa foto que ha llevado a que la clasificación sea correcta, en este caso de una "bicicleta". En una red neuronal, el clasificador genera un vector de puntuaciones [0,01, 0,02, 0,3...,0,8...0,04] para las distintas clases a reconocer ['gato', 'perro', 'persona', 'bicicleta'...'ratón'] y la predicción final es la clase con la puntuación más alta. Lo que trata de hacer esta aproximación es calcular el gradiente de las puntuaciones con respecto a todos los píxeles de la imagen.

44. `https://arxiv.org/pdf/1312.6034.pdf`.

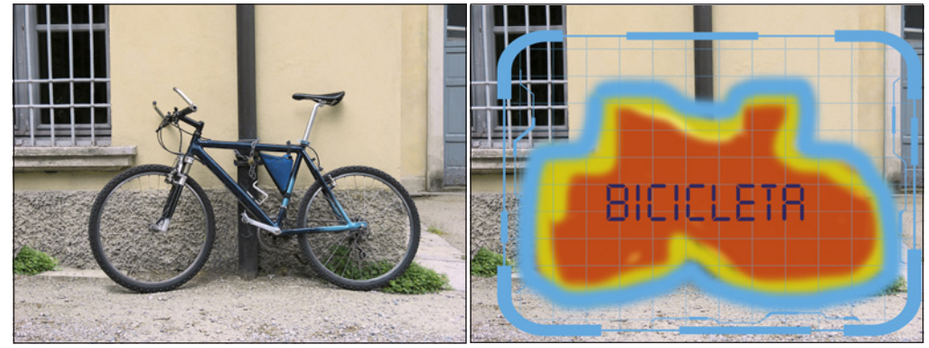

Figura 7.24. Saliency Map.

La ventaja de este método es que es muy visual y, comparado con métodos como LIME o SHAP que pueden también explicar imágenes, es un método más rápido de computación. Este tipo de métodos que trabajan atribución de píxeles pueden ser bastante frágiles porque cambios mínimos en los datos de entrada pueden resultar en explicaciones muy diferentes.[45]

DIFERENTES HERRAMIENTAS DEL MERCADO

Hemos seleccionado 11 herramientas, incluyendo LIME y SHAP. Todas las soluciones son de código abierto en Python, algunas permiten su implementación en R, y cinco de ellas están impulsadas por empresas tecnológicas: **IBM, Google, Microsoft, H2O** y **Oracle**.

Tanto en el caso de Microsoft como en el caso de IBM se añaden métodos desarrollados por sus equipos de investigación utilizando sus propios algoritmos. Todas ellas trabajan con modelos de "caja blanca" y de "caja negra". En el caso de Microsoft, se incluye una metodología a través de una nueva herramienta llamada **EBM** (***Explainable Boosting Machine***) y, en el caso de IBM, se incluyen ocho modelos diferentes[46] con algoritmos específicos desarrollados por sus equipos de investigación. La herramienta de Google tiene un mayor desarrollo en explicabilidad para los temas relacionados con sesgos o equidad, cuyo detalle se mostró en el capítulo de equidad. No obstante, también incluye una implementación para el análisis con

45. `https://christophm.github.io/interpretable-ml-book/`.

46. `https://arxiv.org/abs/1909.03012`.

en cuatro niveles: inaceptables o aplicaciones prohibidas, riesgo alto, riesgo medio y riesgo mínimo. Como hemos indicado, esta cuestión es muy relevante de cara a la tercera etapa, de adopción y escalabilidad de los sistemas de IA, donde el papel de las auditorías éticas y los organismos evaluadores independientes cobra especial protagonismo (de hecho, sobre esta cuestión gira el único artículo que está en vigor desde la publicación de la propuesta de regulación en abril de 2021). Sin duda, los códigos, estándares, certificaciones y auditorías éticas en el ámbito de los sistemas de IA están en pleno impulso y son un complemento y una alternativa a los desarrollos regulatorios.

Pero, más allá de las iniciativas para regular las aplicaciones y sistemas de IA, debe predominar el concepto de **responsabilidad**, entendido no solo como una rendición de cuentas de nuestras acciones como ingenieros, directivos, profesionales o empresas, sino como la determinación previa de lo que se ha de hacer, tomando en consideración a las personas. Efectivamente, como se ha ido viendo a lo largo del manual, en el caso de los sistemas de IA, a veces es difícil identificar todas las partes que se pueden ver afectadas por nuestras decisiones, así como prever todas sus implicaciones y consecuencias. De ahí que el análisis de riesgos, como apunta la regulación, sea una de las tareas clave a la hora de diseñar, desarrollar e implantar sistemas de IA. Este análisis de riesgos e impactos abarca las áreas fundamentales tratadas en el manual: privacidad, equidad y explicabilidad. Por otra parte, emerge la necesidad de gobernar los sistemas de IA y la propia cultura de la organización con una clara orientación a los datos, pero centrada en las personas, durante todo el ciclo de vida de estos sistemas. La gobernanza y la construcción de una determinada cultura son, como la ética, un proceso en sí mismo, que requiere de reflexión, tiempo, participación, compromiso, procedimientos formales, comunicación, códigos, formación, controles, retroalimentación, creatividad, liderazgo, transparencia y honestidad.

Cabe realizar otra reflexión relacionada con la responsabilidad. Toda actividad humana está sujeta a la condición moral, en tanto en cuanto la persona dispone de libertad, consciencia y voluntad, y precisamente de ese control moral se deriva la responsabilidad moral. La responsabilidad moral debe atribuirse por tanto a una persona, o grupo de personas, en cuyo caso, dicha responsabilidad sería compartida, pero no se vería diluida ni mucho menos desaparecería. Obviamente, surge un debate con respecto a dicha agencia moral, en el momento en que incluimos sistemas de IA en la toma de decisiones. Mientras que se trate de sistemas de apoyo a la toma de decisiones, proporcionando sugerencias, datos o comparativas, y la decisión la tome siempre una persona, no supone ningún gran cambio. Ahora bien, tanto en el caso de sistemas semiautónomos *(human in the loop)*, como en el caso de los sistemas autónomos (*human on the loop*), puede surgir una cierta controversia o riesgo de pérdida de dicho control moral, o al menos volver más compleja la atribución de la responsabilidad, no solo de cara a la rendición de cuentas, sino de cara a esa responsabilidad previa y ejercicio consciente de previsión de consecuencias. En el caso de los sistemas autónomos, esta situación es clara en el momento en que dichos sistemas van a tomar la decisión sin intervención humana (aunque la participación humana se requiera para definición y ajuste de los parámetros para un conjunto de decisiones). Pero también puede producirse en el caso de los sistemas semiautónomos, en la transferencia de control entre el sistema y la persona, si las decisiones se toman en intervalos de tiempo de décimas de segundo. En esos casos, hay que prestar de nuevo especial atención al mapa de riesgos y la consideración de los principios fundamentales de la acción responsable.

Durante los últimos años, el concepto de **privacidad** se ha tensionado y actualmente está rodeado de paradojas, falsos enunciados o, cuando menos, cierta confusión. Lo primero que conviene aclarar es que el derecho a la privacidad, es decir, a mantener "privado" lo que cada individuo determine como particular y personal suyo, forma parte de la Declaración Universal de los Derechos Humanos, y es un derecho irrenunciable. En segundo lugar, conviene realizar alguna aclaración también sobre el concepto de "propiedad" de los datos, en concreto, de los datos personales. En tanto en cuanto estos datos personales definen la identidad de un individuo, corresponden a la esfera de lo privado, y más que de "propiedad", conviene hablar de derechos, o, dicho de otra manera, la "protección de datos" no debe interpretarse como protección sobre la "propiedad" de los datos, sino más bien como **protección del derecho de cada individuo a su privacidad**. En tercer lugar, hay que considerar que, si se pudiera utilizar de manera correcta la gran cantidad de datos disponibles hoy en día, se podrían abordar problemas reales y globales, por ejemplo, en el área de salud, que resultarían beneficiosos para la humanidad en su conjunto. Entonces, la pregunta es: ¿cómo poder utilizar datos personales, respetando el derecho de las personas a su privacidad? La respuesta parcial, hasta la fecha, ha sido el uso de técnicas de anonimización, que, como se ha visto en el capítulo correspondiente, no están exentas de riesgos. Otra respuesta parcial consiste en generar un cierto falso debate entre privacidad y transparencia. El verdadero problema a ese respecto son aquellas situaciones en las que, por preservar la privacidad, no se tiene acceso a suficiente información para tomar una decisión con precisión.

¿Conoces la regulación europea sobre IA?
¿Trabajas con IA y estás empezando a implementar principios éticos?
¿Tomas decisiones de negocio con IA y entiendes sus implicaciones éticas?
¿Desarrollas IA y quieres entender los conceptos fundamentales de ética?
Este es tu libro.

El uso de aplicaciones y sistemas de inteligencia artificial (IA) se extiende a todos los sectores y actividades. Por ello se hace necesario comprender los conceptos fundamentales y los nuevos retos con relación a la dimensión ética de la IA.

En este manual, partiendo de una introducción al contexto actual, se presentan los principios generales propuestos por diferentes organismos internacionales, y se propone una ética de inteligencia artificial, en la que se abordan, de una manera resumida y práctica, cuatro principios fundamentales: responsabilidad, privacidad, equidad y explicabilidad. El texto persigue ofrecer al lector un manual de comprensión y consulta con ejemplos, herramientas y referencias.

El libro está dirigido tanto a estudiantes de grado y posgrado como a profesionales de áreas relacionadas con la IA, con perfiles técnicos y no técnicos, que aborden proyectos relacionados con ingeniería informática, ciencia de datos, analítica de datos, *machine learning*, transformación digital, desarrollo de aplicaciones y algoritmos o marketing digital. En definitiva, todas las áreas de negocio en las que está presente actualmente la IA.

La ética es una dimensión ineludible a cualquier actividad profesional, y este manual pretende dotar de los conocimientos y capacidades de comprensión necesarias para las personas que en su día a día trabajan con inteligencia artificial.

ANAYA MULTIMEDIA
www.anayamultimedia.es
ISBN 978-84-415-4595-3
2315177
9 788441 545953