

Mayo 2022

HACIA UN PROSPECTO ALGORÍTMICO

EN EL MARCO REGULATORIO LABORAL EN ARGENTINA

Análisis tecnológico, marco regulatorio
y buenas prácticas

Autores:

Gemma Galdón Clavell, Antoni Lorente Martínez.

Editores:

Erika Molina, Mara Balestrini.

Si desea citar esta publicación: Galdón Clavell, G., Lorente Martínez, A. (2022). Hacia un prospecto algorítmico en el marco regulatorio laboral en Argentina. Análisis tecnológico, marco regulatorio y buenas prácticas, CIPPEC – Éticas – BID Lab. Buenos Aires, 2022.

La responsabilidad de las opiniones expresadas en esta publicación es entera responsabilidad de sus autores y no refleja necesariamente la posición de CIPPEC, BID Lab, Éticas y de todos sus miembros. Las publicaciones de BID, CIPPEC y Éticas se encuentran disponibles en sus respectivos sitios web www.iadb.org/es, www.cippec.org y www.eticasconsulting.com/resources/ [OT1]

Las opiniones expresadas en esta obra son de los autores y no necesariamente reflejan el punto de vista del BID, de su Directorio Ejecutivo ni de los países que representa, así como tampoco del Comité de Donantes del FOMIN (BID Lab) ni de los países que representa.

Copyright © 2022 Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<https://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial. No se permiten obras derivadas.[OT1] Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.



CONTENIDO

Informe ejecutivo	5
Qué es, para qué sirve y cómo se crea el prospecto algorítmico	6
Introducción	7
Definiendo la "revolución digital" en el trabajo	8
El uso de algoritmos en procesos laborales	11
Algoritmos de contratación: búsqueda, evaluación y entrevista	11
Post-contratación: gestión del trabajo, control y evaluación del rendimiento	16
Coordinación algorítmica: el caso de la economía de plataforma	22
Marco regulador: leyes, sentencias y decisiones de las Autoridades de Protección de Datos (DPAs)	25
Fallos judiciales y decisiones de DPAs	27
Buenas prácticas y avances en transparencia algorítmica	31
Registros algorítmicos	31
Etiquetas y calificaciones	34
"Model cards"	37
Hacia un Prospecto Algorítmico	38
Bibliografía	45
Anexo	54
Guía de uso del cuestionario para la obtención de un prospecto algorítmico	54

Informe ejecutivo

El informe "Hacia un prospecto algorítmico en el marco regulatorio laboral en Argentina. Análisis tecnológico, marco regulatorio y buenas prácticas" se lleva a cabo en el marco del proyecto "Desarrollo de la Economía de Plataformas en ciudades como instrumentos para promover la inclusión social, el emprendimiento y la innovación", implementado por CIPPEC (Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento) con el apoyo de BID Lab (el laboratorio de innovación del Banco Interamericano de Desarrollo), y en articulación con la iniciativa FAIR LAC del Grupo BID, que promueve el uso responsable de la tecnología y, en particular, de la IA. El propósito de este informe, elaborado por el equipo de investigación de Éticas, es desarrollar el marco para el desarrollo de herramientas innovadoras para la regulación de las soluciones de inteligencia artificial (IA), específicamente de la transparencia en el ámbito de las relaciones laborales y el contexto de América Latina.

Lo primero que analiza este estudio es el contexto histórico de los avances tecnológicos. Para ello proporcionamos definiciones que ayudan a comprender dichos avances, así como una explicación detallada de los modos y dinámicas mediante los cuales la inteligencia artificial (IA) y los algoritmos impactan sobre las relaciones laborales.

Posteriormente el informe se centra en cuáles son las consecuencias de las transformaciones que las herramientas de datos realizan sobre los entornos y relaciones laborales, desde el proceso de contratación (análisis de los efectos de las herramientas algorítmicas en la búsqueda, evaluación, entrevista y selección de trabajadores) hasta la evaluación del rendimiento algorítmico y su gestión (dirección, evaluación y corrección algorítmica) y el uso de sistemas de IA para organizar los procesos de producción y gestión de recursos, incluyendo trabajadores y trabajadoras, con un énfasis específico en el uso de las herramientas de esta categoría en el contexto de la economía de plataformas. En este sentido, tomamos a la Unión Europea y a España como referencia en términos de regulación, dado que es en éstos territorios dónde la regulación tanto de las políticas de privacidad en relación a los datos como de las nuevas tecnologías está siendo impulsada con mayor fuerza, ofreciendo un camino a seguir para otros países.

Existe un creciente número de litigios que abordan explícitamente el rol de los algoritmos en general y aquellos usados en el empleo y el entorno laboral en particular. Los casos contra los algoritmos en el ámbito laboral varían mucho dependiendo de la jurisdicción nacional donde los juicios tienen lugar, la forma en la que se formulan las cuestiones legales, qué disposición de la legislación se está vulnerando y el tipo de derechos fundamentales que se evocan. En este aspecto, analizamos el marco regulador a través de las decisiones judiciales y de Agencias de Protección de Datos para explorar cómo el sistema legal gestiona los distintos conflictos emergentes en contextos laborales, proporcionando conocimiento específico sobre los avances tecnológicos actuales, la formación de la regulación y estableciendo patrones claros de qué se puede y no se puede hacer.

Para finalizar, revisamos los intentos recientes de promover la transparencia algorítmica a través de metodologías genéricas. Dada la complejidad técnica de los sistemas algorítmicos, y con el fin de facilitar dicha transparencia, presentamos una propuesta de "prospecto algorítmico" como la herramienta específica para la incorporación de la transparencia algorítmica. A pesar de que el formato propuesto podría y debería ser válido también para otros sectores, lo centramos en la legislación laboral, dado que el uso de sistemas de IA

integrados en plataformas laborales está afectando notoriamente las relaciones laborales y el acceso al trabajo.

Qué es, para qué sirve y cómo se crea el prospecto algorítmico

El prospecto algorítmico es una herramienta pensada para que las personas impactadas por los sistemas de toma de decisiones automatizadas conozcan, de una manera resumida y simplificada, la información básica acerca de ese sistema algorítmico.

Este prospecto aborda:

- Aquellos algoritmos que mediante el uso intensivo de datos – ya sea previo a su aplicación o durante la misma–, permiten automatizar decisiones y procesos.
- Aquellos algoritmos flexibles que sean capaces de percibir su entorno y transformar dichas observaciones en acciones que maximicen su probabilidad de éxito en alguna tarea.

El prospecto está diseñado de manera que los desarrolladores de los algoritmos puedan rellenar una serie de formularios y sus respuestas se traduzcan en un despliegue informativo similar al de los prospectos farmacológicos. Es altamente técnico para asegurar que sean los desarrolladores los responsables de proporcionar una información que es necesaria e imprescindible para la rendición de cuentas de los sistemas que deciden sobre aspectos clave de la vida de las personas. El cuestionario en el que se basa está compuesto por más de 40 preguntas agrupadas en torno a tres ejes: información general del algoritmo, análisis técnico, y determinación del impacto que dicho algoritmo tiene. Pretende servir no sólo para facilitar el registro y seguimiento de sistemas algorítmicos por parte de los organismos reguladores correspondientes, sino también para incrementar la transparencia en el uso de modelos de inteligencia artificial y machine learning. Del mismo modo, servirá para su uso frente a demandas legales por falta de transparencia o presencia de sesgos, y otras vulneraciones a derechos laborales y personales básicos.

Es, en definitiva, una llamada a la transparencia algorítmica y a hacer estos sistemas accesibles a aquellas personas que se ven afectadas por ellos.

Introducción

Este informe es el resultado del trabajo del equipo de investigación de Éticas, y su propósito es desarrollar el marco para el desarrollo de herramientas innovadoras para la regulación de las soluciones de inteligencia artificial (IA), específicamente en el ámbito de las relaciones laborales y el contexto de América Latina. La investigación se lleva a cabo en el marco del proyecto "Desarrollo de la Economía de Plataformas en ciudades como instrumentos para promover la inclusión social, el emprendimiento y la innovación", implementado por CIPPEC (Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento) con el apoyo de BID Lab (el laboratorio de innovación del Banco Interamericano de Desarrollo), y en articulación con la iniciativa FAIR LAC del Grupo BID, que promueve el uso responsable de la tecnología y, en particular, de la IA.

El informe comienza posicionando los avances tecnológicos actuales en su respectivo contexto histórico, proporcionando definiciones que ayudan a comprender dichos avances. Sigue proporcionando una explicación detallada de los modos y dinámicas mediante los cuales la inteligencia artificial (IA) y los algoritmos impactan sobre las relaciones laborales.

El segundo capítulo explora cómo y cuáles son las consecuencias de las transformaciones que las herramientas de datos realizan sobre los entornos y relaciones laborales, desde el proceso de contratación (análisis de los efectos de las herramientas algorítmicas en la búsqueda, evaluación, entrevista y selección de trabajadores) hasta la evaluación del rendimiento algorítmico y su gestión (dirección, evaluación y corrección algorítmica) y el uso de sistemas de IA para organizar los procesos de producción y gestión de recursos, incluyendo trabajadores y trabajadoras, con un énfasis específico en el uso de las herramientas de esta categoría en el contexto de la economía de plataformas.

En el capítulo 3, se revisan las decisiones judiciales y de Agencias de Protección de Datos para explorar cómo el sistema legal gestiona los distintos conflictos emergentes en contextos laborales, proporcionando conocimiento específico sobre los avances tecnológicos actuales, la formación de la regulación y estableciendo patrones claros de qué se puede y no se puede hacer. En este sentido, se ha tomado a la Unión Europea y a España como referencia en términos de regulación, dado que es en éstos territorios dónde la regulación tanto de las políticas de privacidad en relación a los datos como de las nuevas tecnologías está siendo impulsada con mayor fuerza, ofreciendo un camino a seguir para otros países.

El capítulo 4 revisa los intentos recientes de promover la transparencia algorítmica a través de metodologías genéricas. Dada la complejidad técnica de estos sistemas, y con el fin de facilitar dicha transparencia, se presenta una propuesta de "prospecto algorítmico" como la herramienta específica para la incorporación de la transparencia algorítmica. A pesar de que el formato propuesto podría y debería ser válido también para otros sectores, se ha discutido en relación a la legislación laboral, dado que el uso de sistemas de IA integrados en plataformas laborales está afectando notoriamente las relaciones laborales y el acceso al trabajo.

Definiendo la “revolución digital” en el trabajo

La naturaleza del trabajo se ha visto afectada durante siglos por los avances tecnológicos. Desde la Revolución Industrial, las nuevas tecnologías han remodelado la producción, el trabajo, las habilidades y las relaciones laborales. El progreso y las innovaciones tecnológicas han sido factores clave del crecimiento económico, la productividad, la riqueza y la prosperidad (Broughel and Thierer 2019). A la par de la transformación de estructuras económicas, sociales y políticas existentes, los cambios tecnológicos también han jugado un papel revolucionario convirtiendo en obsoletos algunos sectores o trabajos, a menudo dando lugar a inquietud y agitación social (Caprettini and Voth 2017). Mientras que la proliferación de nuevas tecnologías digitales puede percibirse como parte de esta vasta historia del desarrollo tecnológico, su escala, ámbito, velocidad e impacto en los procesos laborales a menudo se ha considerado como algo sin precedentes (McKinsey Global Institute 2017). Es por eso que la llamada revolución digital a veces se identifica como la Cuarta Revolución Industrial¹ (Schwab 2016) o la quinta revolución tecnológica del capitalismo² (Freeman and Louçã 2002; Perez 2003). No obstante, otros argumentan que la “revolución digital” ha tenido un menor impacto en la productividad comparada con la Revolución Industrial (Gordon 2016), y/o que en realidad las tecnologías del siglo XXI se encuadran en la Tercera Revolución Industrial, centrada en el uso de la electrónica y la informática para la automatización de procesos. Evaluar la escala y el impacto de estos cambios es por tanto un asunto urgente si se desea que las políticas públicas puedan anticiparse y atajar las dinámicas negativas que son consecuencia de los recientes avances tecnológicos. Este informe intenta abarcar precisamente eso.

La innovación tecnológica ha pasado de ser un campo nicho con un impacto limitado a una explosión de soluciones y herramientas que penetran en el día a día, revolucionan procesos y tienen un profundo impacto en derechos y garantías ya establecidas. Este cambio masivo se ha visto alimentado por los datos y, específicamente, por la habilidad de generar, recopilar y procesar una gran cantidad de datos. A medida que la gente se desenvuelve con dispositivos móviles y usa internet, que los procesos humanos y los negocios van incorporando tecnologías basadas en datos, la mayoría de las actividades humanas han pasado a ser “*datificadas*” o “*datificables*”. A día de hoy, la mayoría de las actividades que hacen las personas se traducen en puntos de datos y muy pocas cosas escapan de los dispositivos de recopilación de datos. Despertadores, documentos y conversaciones generan datos que son capturados, así como las rutinas, interacciones y las vidas sociales, amorosas y, especialmente, laborales de las personas. Por ejemplo, cuando alguien desaparece, la cantidad de fuentes de datos que se pueden explorar para trazar sus últimos movimientos es inmensa: en entornos urbanos, cámaras públicas, información sobre sus tarjetas de crédito y datos telefónicos que pueden trazar las acciones que no se han podido ver pero sí grabar. Lo

¹ Como manifiesta Schwab (2016): “La Primera Revolución Industrial usaba agua y generación mediante vapor para mecanizar la producción. La Segunda usaba potencia eléctrica para crear la producción en masa. La Tercera usaba la electrónica y la informática para automatizar la producción. Ahora una Cuarta Revolución Industrial se está construyendo sobre la Tercera, la revolución digital que ha venido teniendo lugar desde mediados del último siglo. Se caracteriza por la fusión de tecnologías que desdibujan la línea entre las esferas física, digital y biológica”.

² “Las cuatro revoluciones tecnológicas anteriores fueron: la Revolución Industrial inicial (alrededor de 1771); la revolución del vapor y las vías férreas (sobre 1892); la revolución del acero, la electricidad y la ingeniería pesada (alrededor de 1876); y la revolución del aceite, el automóvil y la producción en masa (sobre 1908)”. (Eurofound 2018)

mismo sucede con los procesos empresariales, logísticos y movimientos de dinero o materiales, o con las administraciones públicas. La transformación digital penetra en todos los sectores: todo queda grabado y, con las herramientas adecuadas, puede ser trazado y analizado.

La *datificación* de todo ha ido a la par del desarrollo de nuevas herramientas para recopilar toda esa información generada y así analizarla. El incremento de la potencia de procesamiento ha posibilitado el almacenaje y procesamiento de toda esta información y el desarrollo de lo que se ha llamado Big Data. El Big Data define la habilidad para recopilar y procesar gran cantidad de datos que pueden ser usados para analizar, identificar patrones y observar fenómenos que pueden resultar evidentes en esos datos pero no a simple vista. El Big Data significa que es posible conocer, por ejemplo, cómo evoluciona la COVID19 recopilando datos de distintas fuentes y proporcionando indicadores relevantes como índices de contagio, ocupación de la Unidad de Cuidados Intensivos o número de fallecimientos. Por supuesto, la fiabilidad de estos indicadores dependerá de la calidad de los datos recibidos y de la habilidad de los distintos actores para traducir sus operaciones diarias en entradas digitales actualizadas, pero los datos diarios de casos de COVID son un buen ejemplo de cómo funciona el Big Data, así como sus posibilidades y limitaciones. De este modo, Eurofound (2018) define la revolución digital como "una aceleración general del ritmo de cambio tecnológico en la economía, impulsado por una expansión masiva de nuestra capacidad de almacenar, procesar y comunicar información usando dispositivos electrónicos".

De cualquier modo, el Big Data y las mejoras en la capacidad de procesamiento son solo una de las dinámicas que están moldeando el momento tecnológico actual. El desarrollo de algoritmos de aprendizaje automático (machine learning en inglés) ha acelerado la maduración de soluciones de Inteligencia Artificial. La IA representa un paso más allá del Big Data, ya que incorpora la automatización de los procesos de datos. La habilidad de los procesos de datos para analizar y derivar información y análisis complejos a partir de grandes conjuntos de datos (y los patrones que estos revelan) evoluciona hasta convertirse en procesos mediante los cuales los datos pueden dar un paso más y los algoritmos se programan para implementar de forma autónoma y sin supervisión humana algunos de los resultados del análisis de datos. Estos son los desarrollos que tienen un impacto claro en el mundo laboral, y que se cubren en este informe.

De acuerdo al diccionario Oxford, los algoritmos son "procesos o conjuntos de normas a seguir en cálculos u otras operaciones de resolución de problemas, especialmente las realizadas por ordenadores". En Wikipedia, son una "secuencia finita de instrucciones bien definidas, normalmente usadas para resolver una clase de problemas específicos o realizar un cálculo". Son normas e instrucciones que nos ayudan a organizar una gran cantidad de información (Big Data) e identificar patrones en grandes bases de datos. Esta rápida identificación de patrones en datos históricos o pasados es lo que ha dado lugar a la promesa de la "predicción" algorítmica, la cual en realidad describe un proceso por el cual el análisis de dinámicas anteriores produce una estimación (una probabilidad) de que una dinámica específica vuelva a ocurrir.

Estos procesos algorítmicos tienen una amplia cantidad de aplicaciones e implicaciones: tener acceso a datos sobre conductas de consumo durante un gran periodo de tiempo puede ayudar a los proveedores a identificar tendencias asociadas a la edad, localización geográfica o cualquier otra información, y así calcular la probabilidad de que los clientes actuales y futuros reproduzcan las conductas de consumo de clientes anteriores. Lo mismo aplica a la habilidad de procesar gran cantidad de datos médicos: los algoritmos pueden procesar un gran volumen de imágenes de rayos X y encontrar patrones que puedan contribuir a abordar problemas médicos antes que mediante la observación humana, la cual puede resultar más creativa pero menos eficiente. En el contexto laboral, el tratamiento de datos de rendimiento a escala se utiliza para realizar predicciones sobre el rendimiento o la empleabilidad de individuos y grupos, por ejemplo.³

Para poner ejemplos concretos: los algoritmos de IA son los que permiten que las aplicaciones de movilidad (GPS) puedan sugerir rutas eficientes en tiempo real. Mediante el procesamiento a gran escala de conjuntos de datos sobre cómo se ha comportado la movilidad en una localización específica en el pasado (recopilando datos de automóviles, teléfonos móviles, imágenes por satélite, etc.) e información actual sobre eventos recientes (accidentes, obras, etc.), estos sistemas pueden sugerir la ruta más eficiente, siguiendo los parámetros definidos en el momento de la programación del sistema, de forma autónoma y automática: la ruta más eficiente será la que realizaron más personas en el pasado, a no ser que los datos del presente indiquen alguna interrupción. Siguiendo el ejemplo del COVID usado más arriba, el aprendizaje automático y la IA son el resultado de usar el Big Data no solo para capturar un momento de la pandemia (defunciones de ese día) sino también para hacer una estimación (lo que algunos llaman "predicción") de cuántas defunciones habrá el día de mañana basándose en los patrones pasados y desarrollos actuales. Esta es la contribución de los algoritmos y la IA a las tecnologías que nos rodean.

Los algoritmos pueden ir de muy simples a muy complejos. Un algoritmo simple puede estar programado para encontrar palabras específicas en una gran base de datos de texto (cientos o miles de CVs, por ejemplo), o para identificar formas anómalas en un gran conjunto de imágenes de rayos X, como en el ejemplo mostrado más arriba. Un algoritmo más complejo puede tener que actuar como un árbol de decisiones: si se encuentra x, entonces hacer y. Cuando se replican los árboles de decisiones, emergen algoritmos más complejos que a menudo se definen como "redes neuronales". En estos sistemas, en los que distintos conjuntos de datos interactúan y los algoritmos son programados para seguir normas de un modo que se podría considerar autónomo, la causa de una decisión algorítmica puede resultar difícil de entender, creando lo que Frank Pascuale (2016) ha definido como "Cajas Negras". Un ejemplo de dichos sistemas es el algoritmo usado en la banca para identificar el fraude de tarjetas de crédito: los algoritmos son programados para identificar un comportamiento que se salga de la norma y para producir alertas acordes con esas anomalías (un comportamiento anómalo supone la cancelación de la tarjeta), pero la razón por la cual se accionan estas alertas puede no estar clara, por lo que el "comportamiento" del algoritmo resulta difícil de descifrar o explicar.

³ [IBM using AI to predict employee performance | BenefitsPRO](#)

La mayoría de los algoritmos que nos rodean a día de hoy son algoritmos simples o algoritmos de árboles de decisiones, las redes neuronales continúan teniendo usos muy limitados y en contextos que no tienen un gran impacto en la toma de decisiones cotidianas o laborales. No obstante, la posibilidad de que los algoritmos actúen de forma autónoma es lo que más revuelo levanta en la era de la "Inteligencia Artificial". Entender qué es realidad y qué es promesa tecnológica resulta clave para desarrollar políticas que aborden retos actuales y a la vez sean capaces de anticipar estratégicamente posibles evoluciones futuras y sus impactos.

El uso de algoritmos en procesos laborales

Para capturar el momento tecnológico actual, centramos el análisis en la comprensión de cómo intervienen los algoritmos en distintos "momentos" o niveles de las relaciones laborales. Siguiendo el trabajo de Monedero y Denzic (2019), cubrimos las tecnologías de datos desplegadas en contextos laborales de una forma "secuencial", desde la contratación a la post-contratación y los procesos de evaluación y organización del trabajo. No pretendemos proporcionar una lista exhaustiva de tecnologías digitales disponibles usadas en estos contextos, puesto que semejante esfuerzo tendría un alcance limitado teniendo en cuenta el rápido crecimiento y desarrollo de este campo. Nuestro objetivo es destacar las tendencias y debates clave que conforman los procesos de trabajo y empleo en la era digital, y presentarlos de forma que sean comprensibles y abordables desde una perspectiva regulatoria. Este ejercicio proporciona un marco de análisis sólido basado en dinámicas estructurales que puede resultar útil a corto y medio plazo para categorizar, analizar y regular nuevos productos y procesos tecnológicos.

Algoritmos de contratación: búsqueda, evaluación y entrevista

La contratación es un punto de entrada crítico en las actividades económicas. Determina quién tiene acceso a qué tipo de trabajos. La contratación ha sido históricamente un proceso manual, pero hoy las nuevas tecnologías han creado un floreciente mercado de servicios y aplicaciones que prometen mejorar cómo se evalúa a los profesionales usando Big Data e IA. Desde la aparición de los primeros portales de trabajo a finales de los años 90 como Monster, CareerBuilder y Craigslist, los cuales no eran más que tableros de mensajes, la tecnología digital ha remodelado todos los pasos del proceso de contratación, tanto para las empresas como para los que buscan trabajo. Según Glassdoor (2019), la empresa o empleador medio estadounidense gasta alrededor de 4.000 USD y 24 días en contratar un nuevo trabajador o trabajadora. Aunque este dato en particular no se conoce para Europa, el proceso de contratación es, en general, costoso, laborioso y arduo. Y, lo que es más, el coste de una mala contratación puede ser muy elevado (Human Resources – Northwestern

University 2019). Por lo tanto, hay un interés claro en la mejora de los procesos de contratación y en minimizar la posibilidad de malas contrataciones. Pero aunque no está claro que los datos y la IA pueden ayudar en este proceso, cientos de startups prometen precisamente acabar con las malas decisiones de contratación.

Algunas de las tecnologías que se han desarrollado en este espacio son el Sistema de Seguimiento de Candidatos (ATS por sus siglas en inglés), que ayuda a las organizaciones a gestionar y hacer seguimiento del flujo de aplicaciones recibidas tras la publicación de una oferta de trabajo, y el Sistema de Gestión de Contratación (RMS por sus siglas en inglés), el cual complementa al ATS y apoya a los encargados de la selección de personal a la hora de promocionar puestos, en la búsqueda de talento, en crear reservas de personas cualificadas, conseguir candidatos y organizar las entrevistas. De acuerdo a un informe reciente, el mercado global tecnológico de contratación ha crecido hasta 1,75 mil millones de dólares en 2017 y se espera que se duplique hasta 3,1 mil millones de dólares en 2025 (Fuller et al. 2021). Las tecnologías digitales destinadas a la contratación de personal prometen mejorar la eficiencia mediante la reducción del tiempo y coste de contratación, maximizando la calidad de dichas contrataciones, evitando los sesgos propios de los humanos durante el proceso de reclutamiento, reduciendo los tiempos de sustitución mediante el aumento de la retención y manteniendo la diversidad en el entorno de trabajo (Bogen and Rieke 2018; aJunwa and Greene 2019).

Los sesgos humanos, conscientes o inconscientes, en la selección de personal han sido ampliamente documentados. Por ejemplo, Kübler et. al. (2019) demostró la existencia de discriminación de género en distintas ocupaciones pidiendo a propietarios de empresas y gerentes de recursos humanos que evaluaran currículums ficticios. Reynolds et. al. (2021) examinaron cómo la orientación ideológica de los responsables de contratar (como la inclinación hacia el elitismo frente al igualitarismo) llevaba a la discriminación racial durante la contratación. Leckcivillize and Straub (2018) demostraron que llevar un pañuelo en la cabeza reduce las probabilidades de ser seleccionado para un trabajo, especialmente cuando el encargado de la contratación es una persona más mayor.

En este contexto, la promesa de una contratación sin sesgos mediante la eliminación del criterio humano y creando procesos algorítmicos de selección tiene mucho sentido. Pero, ¿pueden las máquinas eliminar por completo las preferencias o inclinaciones propias de los humanos y proporcionar una evaluación más objetiva de los candidatos/as durante los procesos de contratación? La evidencia muestra que este no es necesariamente el caso. Los algoritmos reflejan y perpetúan los sesgos, desigualdades y discriminación de distintas formas. El uso de datos de entrenamiento no representativos es a menudo el culpable de los sesgos de los algoritmos, que analizando datos de un mundo que discrimina aprenden, lógicamente, a discriminar. Como muchos estudios han mostrado,⁴ cuando un cierto grupo de mujeres o personas de color están pobremente representadas en estos datos de entrenamiento del algoritmo, las herramientas de predicción tienen un peor desempeño en dichos grupos. Aunque estas fuentes técnicas de sesgo pueden ser fácilmente detectadas y

⁴ Un famoso ejemplo de cómo unos datos de aprendizaje o entrenamiento poco representativos dan lugar a sesgos en un algoritmo es el estudio *Gender Shades*, el cual muestra que los sistemas de reconocimiento facial tienen un pobre rendimiento con las mujeres de piel más oscura, debido a la poca presencia de éstas en los datos de entrenamiento (Buolamwini and Gebru 2018).

mitigadas, hay otros modos de funcionamiento del sesgo algorítmico que resultan más sutiles. Los datos de entrenamiento, independientemente de que sean representativos o no, pueden también reflejar sesgos, discriminaciones y desigualdades históricas, como el racismo y sexismo estructurales (Turner Lee, Resnick, and Barton 2019). Deshacerse de datos como el género o la etnia en el proceso de programación de los algoritmos no previene del todo esos sesgos, ya que estos atributos están a menudo correlacionados con otros datos (residencia, nivel de estudios, renta, etc.). Por ejemplo, que alguien trabaje muchas horas puede ser un dato útil para determinar si un trabajador o trabajadora puede acceder a una promoción laboral, pero la posibilidad de trabajar más horas está condicionada por la distancia del trabajo a casa, lo que está relacionado con la etnia (Walker 2012), o la existencia de obligaciones familiares o de cuidados, lo que se relaciona con el género. En otros casos, las decisiones de promoción pueden estar relacionadas con el tiempo de permanencia en un mismo lugar de trabajo, lo que de nuevo puede discriminar a mujeres que se han visto obligadas a interrumpir su trayectoria laboral por motivos familiares (Barocas and Selbst 2016). Todos estos factores sociales e históricos son incorporados en las bases de datos, y recogidos y reproducidos por algoritmos.

La contratación no consiste en un solo paso, sino en una serie de acciones que siguen unas a otras en cadena (Bogen and Rieke 2018), y que las tecnologías pretenden reproducir. El proceso empieza con la búsqueda, invitando a los candidatos potenciales a presentarse para las vacantes abiertas, habitualmente mediante portales en internet. Durante la evaluación, los encargados de la selección de personal revisan las calificaciones de los candidatos que han aplicado a una vacante o que han presentado su perfil de forma pasiva a través de distintas plataformas. La etapa de entrevista incluye una evaluación directa e individual de los candidatos o candidatas seleccionados. Finalmente, la etapa de selección tiene lugar cuando se toma una decisión final sobre quién es contratado y quién rechazado. Todos estos pasos de la contratación usan un amplio abanico de herramientas digitales, y cada uno de ellos plantea cuestiones relacionadas con su legalidad y el impacto de los algoritmos en la posibilidad de optar a un empleo en condiciones de igualdad y no discriminación.

Por ejemplo, hay distintas plataformas online disponibles para anunciar empleos y buscar candidatos potenciales. Estas empresas pueden atraer distintos públicos para sus anuncios mediante la creación de grupos de usuarios con atributos comunes como pueden ser la localización, años de experiencia, antigüedad y habilidades. Pero la selección de los grupos de usuarios objetivo no es el único filtro que determina la visibilidad de las ofertas de empleo. Los algoritmos de la plataforma también juegan un papel crucial a la hora de determinar qué anuncios son visibles para qué personas. Puesto que el espacio del anuncio es limitado, la visibilidad de los anuncios depende también de las predicciones algorítmicas sobre quién tiene más posibilidades de interesarse (por ejemplo, pulsar en el anuncio) o tomar la acción deseada (por ejemplo, aplicar para ese puesto) (Bogen and Rieke 2018, 18).

Aquí es donde se introduce el sesgo del algoritmo en las plataformas de empleo, haciendo que algunos anuncios sean sistemáticamente invisibles para ciertos grupos de personas, independientemente de sus habilidades o de la relevancia de su experiencia

laboral. ¿Cómo predicen los algoritmos qué candidato es una buena opción para qué trabajo? Hay dos formas de hacerlo, por un lado, mediante el *filtrado basado en el contenido*, que proporciona recomendaciones basándose en las acciones anteriores de los usuarios (gustos, clics, contenido guardado, trabajos a los que ha aplicado, etc.). La ventaja de ese filtrado basado en el contenido es que no usa información proporcionada por terceros; el inconveniente es que es un obstáculo para nuevos usuarios, ya que el sistema no posee datos sobre ellos/as, o para personas que quieran cambiar de carrera, explorar cosas diferentes, etc. que no se correspondan con sus acciones pasadas (Lakshmi and Lakshmi 2014), pero puede convertirse en una camisa de fuerza al acentuar los sesgos cognitivos y las acciones anteriores de los usuarios. Si una mujer joven elige una serie de trabajos peor pagados durante un periodo de tiempo específico en su vida, después de un tiempo los empleos mejor pagados serán menos visibles para ella, a pesar de sus cualificaciones. El *filtrado colaborativo*, a su vez, funciona creando recomendaciones en base a la identificación de patrones similares de comportamientos y preferencias. En estos casos, el sistema identifica usuarios con intereses similares y toma decisiones en base a afinidades, asegurando que las personas candidatas compartan características. Mientras que el filtrado colaborativo tiene la misma problemática para los nuevos usuarios (no hay suficientes datos anteriores para realizar una correcta asignación a un grupo), tiene mejor capacidad de realizar nuevas recomendaciones que estén alejadas de las acciones anteriores de las personas usuarias. Pero cae también en el riesgo de estereotipar a los usuarios. Por ejemplo, una mujer puede querer buscar un trabajo con mayor responsabilidad y mejor pagado, pero debido a que otras mujeres en una situación laboral similar buscan trabajos no tan bien pagados, puede que sólo acabe recibiendo anuncios para trabajos peor pagados debido a esta evaluación de su perfil. Estos sistemas "predictivos" juegan un papel crucial en la contratación, ya que pueden influenciar considerablemente las decisiones de contratación de las empresas, además de las de las personas candidatas (Almalis, Tsihrintzis and Karagiannis 2014).

Tras recibir las candidaturas, el siguiente paso crítico de la contratación es revisarlas, rechazar aquellas que no cumplan los requerimientos o las que resulten poco convincentes y realizar una selección para pasar a la fase de entrevista. Teniendo en cuenta el gran número de aplicaciones para ciertos puestos, las decisiones automatizadas pueden reducir el tiempo de revisión de las candidaturas.⁵ La mayoría de las candidaturas se rechazan en una etapa temprana de este proceso, y lo suelen hacer los sistemas automatizados. Los sistemas algorítmicos pueden revisar aquellos CVs que no cumplen con los requerimientos mínimos de la descripción de la vacante. Los sistemas de filtrado de candidatos más básicos funcionan a través de búsqueda de palabras clave y preguntas pre-diseñadas llamadas "knockout questions" o preguntas de eliminación (Bogen and Rieke 2018, 26). Las herramientas más avanzadas ofrecen sistemas interactivos como chatbots, entrevistas virtuales, test de personalidad, etc. Por ejemplo, Mya⁶ es un asistente de contratación tipo chatbot que usa IA conversacional que promete acortar el tiempo entre las fases de

⁵ De acuerdo a una publicación reciente, mientras que al inicio de la década de 2010 la media de aplicaciones para un empleo corporativo era de 120, a finales de esta década, este número ha aumentado hasta los 250 candidatos por trabajo (Vincent 2021).

⁶ <https://www.mya.com/>

aplicación y entrevista en un 79%, mejorar la experiencia del candidato/a tras la aplicación manteniendo el contacto y conectar a los encargados de selección de personal con las mejores personas candidatas. Otra herramienta disponible es Ideal,⁷ que usa técnicas de aprendizaje automático para evaluar qué tipo de decisiones ha tomado un empleador en el pasado, y valora el encaje entre esas preferencias y dinámicas pasadas y las personas candidatas en base a datos clave, proporcionando una letra a modo de puntuación, de la A a la D, para cada persona candidata. A pesar de las promesas comerciales de estas soluciones, la automatización del proceso de contratación tiene también costes. Según un reciente informe de la Escuela de Negocios de Harvard, las tecnologías de contratación como ATS (Sistema de Seguimiento de Candidatos) y RMS (Sistema de Gestión de Contratación) agravan una creciente y peligrosa dicotomía entre la población activa, distinguiendo entre candidatos deseables y aquellos a quienes los algoritmos rechazan sistemáticamente al considerarles no cualificados. Estos segundos quedan "ocultos", independientemente de su capacidad para ejecutar los requerimientos de ese empleo (Fuller et al. 2021). Además, los algoritmos de selección laboral usan unos filtros tan simples como estrictos que llevan a la eliminación de candidatos con diferencias marginales, o aquellos que no poseen las cualificaciones secundarias "deseables". Un filtro común es la preferencia de candidatos que no tengan huecos temporales en su vida laboral. A consecuencia de esto, candidatas que dejaron su trabajo durante más de seis meses por distintos motivos, como un embarazo complicado, enfermedad de un miembro de la familia o un traslado son directamente eliminadas del proceso. Otro filtro común es exigir unas cualificaciones de formación muy concretas, lo que excluye a candidatos motivados que tienen experiencia y habilidades para cumplir los requisitos del puesto, pero no una formación específica. En general, y por sus características técnicas, los algoritmos solo pueden valorar aptitudes concretas que se presenten de forma estandarizada (año de nacimiento, universidad de graduación, palabras clave, etc.) por lo que toda la información que aporte un candidato/a y que no haya sido programada en el sistema para ser reconocida, es ignorada por los sistemas de IA.

Sistemas automatizados prometen asistir también en la fase de entrevista de los candidatos. La tecnología de datos y las herramientas de IA ofrecen analizar las entrevistas grabadas en video y evaluar las capacidades, personalidades, motivaciones y la forma de encajar de un candidato/a a una empresa a través de "detectores de emoción mediante IA". Por ejemplo HireVue⁸ proporciona IA que a través de las llamadas "entrevistas estructuradas" promete sistematizar el proceso de entrevistado estableciendo relaciones entre las emociones de las personas candidatas y las "habilidades requeridas para el trabajo", "en lugar de cualidades como su carisma, género, edad, etnia o afinidad personal con el entrevistador"⁹. En una entrevista estructurada, se realizan las mismas preguntas a cada candidato en el mismo orden, y cada candidato es evaluado del mismo modo. HireVue analiza el video mediante sistemas de aprendizaje automático que analizan la expresión facial, el contacto visual, el tono de voz, la elección de palabras, la complejidad de las mismas y los temas discutidos, y las compara con los datos de aprendizaje recogidos en la

⁷ <https://ideal.com/>

⁸ <https://www.hirevue.com/>

⁹ <https://webapi.hirevue.com/wp-content/uploads/2021/06/hirevue-structured-interviews-whitepaper-2021.pdf>

empresa, entrevistando a empleados ya existentes, para producir una escala de puntuación de 1-100 para cada candidato. A pesar de las promesas de HireVue, no existe evidencia científica de que el reconocimiento de las emociones funcione, ni de que las expresiones faciales sean un indicador del futuro desempeño laboral. Además, se ha documentado que los sistemas de reconocimiento facial no tienen buenos resultados cuando se aplican a las mujeres, por ejemplo, y que aún son peores en el caso de las mujeres no blancas (Buolamwini and Gebru 2018). Los sistemas de reconocimiento automático del habla también muestran un sesgo frente a la población negra ([Koenecke et al. 2020](#)) y hacia acentos diferentes por cuestiones de género, edad, diferencias regionales o uso de lenguas no maternas (Feng et al. 2021).

Para resumir, la datificación del proceso de contratación da lugar a riesgos por la falta de transparencia, responsabilidad sobre la toma de decisiones, el potencial sesgo algorítmico y la asimetría de información entre empresarios y candidatos que promueven estos sistemas (Dencik and Stevens 2021). Si no se monitorizan, los algoritmos en el proceso de contratación conllevan el riesgo de reforzar y ampliar desigualdades y la discriminación de ciertos grupos de personas, algo que las leyes llevan mucho tiempo persiguiendo eliminar en los procesos de contratación no digitales. Además de la importante brecha entre la aplicación de provisiones legales en los dominios digitales y no digitales, existe la preocupación de que no siempre es posible conocer si estas prácticas sesgadas existen, y tampoco lo es determinar quién es el responsable final de las mismas y sus consecuencias. Los algoritmos, especialmente aquellos que usan aprendizaje automático y soluciones de IA, funcionan de forma opaca y son muy sensibles a la retroalimentación. Las ofertas de trabajo para cubrir puestos de ingeniería redactadas en tono sesgado pueden atraer a más candidatos hombres que mujeres, independientemente de sus habilidades y cualificaciones. Cuanto más pulse un empleador sobre los perfiles masculinos, más perfiles masculinos recibirán otros empleadores, y menos perfiles de mujeres serán visibles para las empresas que ofrecen puestos de alta cualificación, lo que desincentivará a las mujeres de aplicar a estas ofertas y hará que menos reciban buenas ofertas de trabajo o que, incluso si lo consiguen, su salario será menor.

Post-contratación: gestión del trabajo, control y evaluación del rendimiento

Las tecnologías algorítmicas en el ámbito de la post-contratación dan forma a los procesos de gestión laboral, control y observación del empleado y la evaluación del rendimiento. El mundo laboral ha estado históricamente estructurado en base a estructuras jerárquicas que casi siempre han requerido de la existencia de posiciones de poder para monitorizar, controlar y evaluar otros puestos de arriba a abajo (Ball 2010). Lo que no tiene precedentes en el ámbito de control de los entornos laborales mediante algoritmos es la extensión e intensificación de la monitorización a tiempo real de los trabajadores y trabajadoras, que es posible gracias a los múltiples sistemas de recopilación de datos que rodean toda actividad

laboral. El control algorítmico actual, pues, es más “amplio”, “instantáneo”, “interactivo” y “opaco” (Kellogg, Valentine and Christin 2020, 371) que otros sistemas de control anteriores.

Las tecnologías de control algorítmico en el trabajo son *amplias*, ya que recopilan una gran variedad de datos sobre trabajadores a partir de una amplia gama de sensores y dispositivos. Las cámaras y los dispositivos de audio son capaces de grabar movimientos y lenguaje; los dispositivos móviles pueden registrar la movilidad; los sensores pueden controlar la presencia, los dispositivos biométricos pueden confirmar la identidad y los softwares de control de contenido como el Procesamiento del Lenguaje Natural (PLN) pueden monitorizar los emails y los mensajes de chat para evaluar el estado de humor, productividad y la intención de renunciar de un trabajador. Estas tecnologías son *instantáneas* ya que se retroalimentan en tiempo real. La información se graba, analiza y comunica de forma inmediata y los datos se incorporan al proceso de producción en tiempo real. Son *interactivas* debido a que tanto los empleados y empleadas como el empleador pueden iniciar sesión e interactuar con el sistema. Los algoritmos pueden monitorizar los chats o llamadas y pueden avisar al supervisor en caso de conflicto. Los algoritmos de control en el ámbito laboral también pueden resultar *opacos* de forma intencional debido al temas de secreto comercial y propiedad intelectual, por la alfabetización técnica y digital requerida o debido a la opacidad del propio aprendizaje automático (Burell 2016).

Kellogg et al. (2020) concluye que el control algorítmico en el entorno laboral se lleva a cabo mediante seis mecanismos principales: “los empresarios pueden usar el algoritmo para manejar a los empleados mediante *restricciones* y *recomendaciones*, evaluarlos mediante la *grabación* y la *calificación* y disciplinarlos mediante el *reemplazo* y la *recompensa*”. Cuando aparecen los sistemas de control algorítmico, los trabajadores (de forma individual o colectiva) desarrollan nuevas formas de relacionarse con esas nuevas tecnologías digitales y actuar ante ellas. Kellogg llama a estas nuevas tácticas resistencia emergente como *algoactivismo*. El siguiente cuadro recoge estas categorías y sus impactos:

Control laboral algorítmico	Cómo funciona	Acciones ejemplo	Experiencia potencial del trabajador
Dirigir	Recomendación algorítmica	- Toma de decisiones - Recomendación de pasos a seguir específicos	- Frustración - Prejuicios - Menor capacidad de opinión e intervención
	Restricción algorítmica	- Restringir el comportamiento y/o el acceso a información	- Precariedad
Evaluar	Grabación algorítmica	- Grabación del comportamiento - Rastreo de movimiento	- Pérdida de privacidad - Precisión de los datos
	Calificación algorítmica	- Información a tiempo real	- Discriminación

		- Clasificación y calificación online	
Disciplinar	Reemplazo algorítmico	- Despedir a trabajadores de bajo rendimiento - Reemplazar a trabajadores por aquellos que siguen las directrices de los gerentes	- Precariedad - Frustración - Estrés
	Recompensa algorítmica	- Proporcionar recompensas a tiempo real	

Tabla 2: Control algorítmico del entorno laboral. Adaptado de Kellogg et. al. (2020).

- DIRECCIÓN ALGORÍTMICA: RECOMENDACIÓN Y RESTRICCIÓN

La dirección algorítmica tiene lugar cuando los algoritmos guían a los trabajadores en cuanto a qué necesita realizarse, en qué orden y en qué periodo de tiempo. Esto puede funcionar tanto por recomendación como por restricción algorítmica. La *recomendación algorítmica* funciona analizando los patrones de productividad anteriores y la prescripción de acciones basada en la mejora de los mismos. Así, el algoritmo impulsa al trabajador a tomar una determinada ruta o ritmo de acción preferido por el sistema. Por ejemplo, los sistemas usados en centros de atención al cliente como Cogito¹⁰ monitorizan las llamadas, detectan los niveles de satisfacción de los clientes, sugieren acciones, advierten si se habla muy rápido, mandan avisos sobre requerimientos legales faltantes o sugieren cuando hay que llamar a un supervisor/a (Sánchez-Monedero and Dencik 2019, 22). Los supervisores tienen acceso a paneles a tiempo real donde se toman mediciones de la experiencia del cliente y el comportamiento de los trabajadores y trabajadoras. También pueden escuchar de forma proactiva las llamadas y monitorizar el trabajo a tiempo real.

La *restricción algorítmica* funciona proporcionando información selectiva, por ejemplo, permitiendo el acceso del trabajador o trabajadora a ciertos tipos de información y restringiendo el mismo para otros. Las prácticas de dirección algorítmica pueden resultar en frustración por parte de la plantilla, si existe conocimiento sobre esas restricciones. Los avisos y advertencias instantáneas pueden incrementar los niveles de estrés y dañar la estabilidad emocional de los trabajadores. No tener el control total sobre la toma de decisiones en el trabajo puede debilitar la autonomía, libertad de elección y racionalidad del trabajador (Schmidt and Engelen 2020). Las restricciones pueden también ocultar la visión global del trabajo y dividirlo en micro-tareas que pueden distanciar a los trabajadores, debilitando su sensación de control e incrementando la sensación de precariedad.

- EVALUACIÓN ALGORÍTMICA: GRABACIÓN Y CALIFICACIÓN

La evaluación algorítmica incluye la revisión de las actividades de los trabajadores, evaluando su rendimiento e identificando aquellos que no están desempeñando su labor de forma adecuada. Hay una amplia gama de mecanismos mediante los cuales se puede

¹⁰ <https://cogitocorp.com/>

realizar la evaluación algorítmica. Los dos principales son la grabación y calificación algorítmica. La *grabación algorítmica* incluye el uso de "procedimientos computacionales para monitorizar, agrupar e informar, a menudo a tiempo real, una gran variedad de datos precisos de fuentes tanto internas como externas" (Kellogg, Valentine and Christin 2020, 376). La monitorización se puede realizar mediante numerosos dispositivos del entorno laboral como ordenadores, teléfonos móviles, videovigilancia, dispositivos portátiles y tarjetas de acceso. Algunas de estas tecnologías no resultan novedosas, pero la creciente integración de sistemas de datos y su análisis instantáneo no tiene precedentes. Por ejemplo, mientras que la tecnología de monitorización por videovigilancia no es nueva, ahora posee nuevas capacidades como el reconocimiento facial y la detección de emociones. Mediante la "clasificación automatizada y semántica de escenas", los sistemas de grabación de video ahora pueden informar de cuánto tiempo pasa una persona haciendo ciertas actividades y, aparentemente, de su estado "emocional" (Deng et al. 2015), que es utilizado para tomar decisiones sobre la productividad o adherencia a normas.

Los sistemas algorítmicos integran también herramientas de control presencial, autenticación biométrica y análisis de datos para rastrear actividades de los empleados. Por ejemplo, el uso del aseo por parte de los trabajadores se puede monitorizar mediante un acceso con requerimiento de tarjeta de identificación. En casos extremos, el tiempo que se pasa en el aseo se resta del total de horas trabajadas (Sánchez-Monedero and Dencik 2019, 18). Además de fichar a la entrada y salida del edificio, las tarjetas de identificación pueden ser biométricas. Humanyze¹¹ desarrolla tarjetas de identificación que tienen una amplia gama de sensores de rastreo que captan desde el "movimiento e interacciones en la oficina hasta la duración de conversaciones y tono de voz" (Derousseau 2017). OccupEye¹² proporciona servicios de monitorización del calor y movimiento por el espacio de la oficina.

Fuera de las oficinas, los movimientos y la movilidad de los trabajadores se puede seguir mediante el teléfono móvil, el GPS o dispositivos de Identificación de Radio Frecuencia. Los trabajadores de las industrias del transporte, logística o sanitaria a menudo deben llevar dispositivos de rastreo de movilidad. La empresa Amazon, por ejemplo, usa una "aplicación de seguridad del conductor digital" llamada Mentor, que rastrea y monitoriza el comportamiento de los transportistas de mercancías en la carretera. La aplicación proporciona una puntuación de seguridad al volante, conocida como puntuación FICO, basándose en su desempeño en la carretera con datos como número de frenazos, exceso de límite de velocidad, uso de cinturón de seguridad, conducción marcha atrás, llamadas telefónicas o envío de mensajes (Palmer 2021).

Los sistemas impulsados por IA también permiten a los empresarios algunos métodos de vigilancia como la revisión de los emails, chats y conversaciones telefónicas, rastrear sus actividades durante reuniones online y monitorizar el uso del ordenador y de internet. La pandemia de COVID y la extensión del teletrabajo ha impulsado la adopción de este tipo de sistemas de vigilancia (Vitak and Zimmer 2021). De acuerdo a un estudio que encuestó a 239 grandes empresas, más del 50% de éstas declaró usar técnicas de

¹¹ <https://humanyze.com/>

¹² <https://occupeye.com/>

monitorización poco "tradicionales" (Wartzman 2019). Proveedores como Activtrak¹³ e Interguard¹⁴ ofrecen "software de almacenamiento de teclado", que almacenan todo lo que se tecldea o cualquier acción tomada en el dispositivo que se monitoriza (Martins 2021). Estas herramientas pueden trazar la productividad o falta de ella durante la jornada de cada empleado/a, registrar qué herramientas se usan durante las horas productivas, tomar capturas de pantalla y notificar a los supervisores en caso de uso de redes sociales o cualquier otro comportamiento o distracción que se salga de la rutina autorizada por la empresa.

Algunas organizaciones también dicen rastrear la atención de los trabajadores mediante datos biométricos como el movimiento de los ojos, cambios en el cuerpo y expresiones faciales (West 2021). Por ejemplo, la herramienta de teleconferencia Zoom detecta cuando la pestaña de la reunión de Zoom se minimiza o no es la pestaña activa del ordenador de una persona durante más de 30 segundos, y avisa al promotor de la sesión con un icono gris a la izquierda del nombre de la persona en la lista de participantes (Malani 2020).

La vigilancia de la plantilla también se puede realizar mediante dispositivos móviles o que se pueden implantar. Las iniciativas de bienestar corporativo impulsan un boom de dispositivos que se pueden llevar en la muñeca de salud y actividad, como los vendidos por Fitbit, Carmin y Apple. Estos dispositivos pueden enviar datos altamente sensibles sobre salud a la empresa, y vulnerar la privacidad de las personas contratadas (Rowland 2019). En algunos casos extremos, compañías como Three Square Market¹⁵ han implantado microchips de radiofrecuencia a sus trabajadores que les permiten abrir puertas, iniciar sesión en sus ordenadores y comprar artículos de las máquinas expendedoras de la compañía (Holley 2018), todo esto mientras dejan un rastro digital de sus actividades. La compañía sueca Biohax¹⁶, líder en la fabricación de implantes de microchips para humanos, promete "transformar el Internet de las Cosas en el internet de nosotros".

Además de recopilar datos directos del rendimiento del trabajador, Levy and Barocas (2018) hablan de una "vigilancia refractiva", un proceso que se da lugar cuando son los datos de los consumidores los que se recopilan para tomar decisiones sobre el rendimiento de los trabajadores y trabajadoras. Por ejemplo, RetailNext¹⁷ integra múltiples fuentes de datos usando varias tecnologías como sensores, cámaras y balizas que captan los comportamientos de los consumidores y usan los datos recopilados para optimizar los horarios de los trabajadores y así reducir costes laborales.

Los datos recopilados por estas aplicaciones y sistemas se usan para la toma de decisiones laborales, acciones disciplinarias, distribución de bonos y beneficios, a menudo basados en sistemas de puntuación algorítmicos opacos y que segregan a la fuerza laboral entre "buenos" y "malos" empleados. Las herramientas algorítmicas que evalúan el rendimiento de los trabajadores mediante rankings y puntuaciones usan distintas fuentes de

¹³ <https://www.activtrak.com/>

¹⁴ <https://www.interguardsoftware.com/>

¹⁵ <https://32market.com/public/>

¹⁶ <https://www.f6s.com/biohaxinternational>

¹⁷ <https://retailnext.net/>

datos. Por ejemplo, en los servicios de hostelería como restaurantes y hoteles, las plataformas abiertas como Yelp o TripAdvisor permiten a los clientes puntuar o valorar la calidad del servicio, las instalaciones y la relación calidad precio. Este flujo continuo de valoraciones por parte de clientes es usado por las empresas para conocer el rendimiento de los trabajadores en un caso de "vigilancia refractiva" como el que mencionamos más arriba, y a menudo ignora hechos como que los clientes suelen discriminar a ciertos perfiles de trabajadores en el sector servicios, o que estas críticas pueden estar manipuladas o tener intencionalidades no especificadas. Por ejemplo, Edelman et. al (2017) llevó a cabo un experimento en Airbnb que muestra que las aplicaciones de huéspedes con nombres claramente afroamericanos solían ser menos aceptadas por los anfitriones de Airbnb en comparación con huéspedes idénticos pero con nombres propios de personas blancas.

Las puntuaciones y rankings no solo evalúan el rendimiento de los trabajadores sino que también se usan para predecir rendimientos futuros. Así, datos pasados sobre logros, adaptabilidad, pensamiento analítico o comunicación de los empleados son los que determinan sus metas de rendimiento futuras (Kellogg, Valentine and Christin 2020, 379).

El control algorítmico conlleva importantes consecuencias para los trabajadores. La vigilancia constante puede hacer sentir a los trabajadores vulnerables, asustados y menos creativos. Puede reducir su satisfacción laboral global y disminuir su moral (Azer 2021). Se ha documentado que la vigilancia daña los niveles de autoestima y confianza, dando lugar a estrés, paranoia, desórdenes nerviosos y síndrome del túnel carpiano (European Parliamentary Research Service 2020, 35). Además, clasificar a los empleados puede dar lugar a una situación de competencia destructiva entre los mismos, dañando tanto el rendimiento individual como el grupal (Woike and Hafenbrädl 2020). Más aún, las puntuaciones de internet pueden no ser precisas y producir discriminación de género o etnia, con poca o ninguna opción por parte del empleado para recurrir una decisión algorítmica injusta, ya que de nuevo, los sistemas de soporte y asistencia tradicionales no han sido aún trasladados al entorno online y al contexto de los algoritmos. La vigilancia intensiva durante las horas laborales dentro y fuera de la oficina también implica una erosión de la privacidad y pone en jaque la protección de datos dentro y fuera del ámbito laboral (Collins 2020).

- DISCIPLINA ALGORÍTMICA: REEMPLAZO Y RECOMPENSA

Las empresas refuerzan la disciplina laboral mediante recompensas y castigos. La *recompensa algorítmica* incluye la compensación a trabajadores de alto rendimiento mediante la provisión de más oportunidades, mejores salarios, bonos y promociones. En este ámbito, la IA ofrece soluciones de gamificación. Por ejemplo Central¹⁸ ofrece soluciones avanzadas de gamificación para, asegura, aumentar el compromiso de los empleados mediante pequeñas dosis diarias de diversión y emoción.

La disciplina algorítmica funciona de un modo similar, mediante el reemplazo/despido o compensando a los empleados con la ayuda de indicadores algorítmicos. El *reemplazo algorítmico* sucede cuando un empleado de bajo rendimiento es

¹⁸ <https://central.com/>

despedido en base a una evaluación algorítmica. Aunque el reemplazo de empleados con bajo rendimiento no es nada nuevo, los algoritmos de reemplazo difieren de los métodos tradicionales de despido ya que los mismos se pueden ejecutar a una mayor escala y en una fracción de tiempo. Combinado con otras prácticas contractuales, el despido algorítmico también puede no tener en consideración los estándares y garantías laborales y elimina la posibilidad de trabajar mediante la “desconexión” remota de los empleados. Con el uso de herramientas de predicción, los trabajadores también pueden ser despedidos por sus potenciales rendimientos futuros, que aún no conocen o sobre los que no poseen conocimiento o control.

El reemplazo y recompensa algorítmicos pueden también afectar negativamente a los trabajadores. Pueden resultar en mayores niveles de precariedad, y a prácticas de despido ante los que los trabajadores y trabajadoras no tienen herramientas de defensa o amparo. Como Rosenblat (2018) argumenta, los algoritmos reescriben las normas laborales y refuerzan la asimetría de información entre empresarios y trabajadores, lo cual da lugar a una dependencia ciega de los trabajadores en los sistemas algorítmicos, una mayor conformidad y una menor autonomía.

Coordinación algorítmica: el caso de la economía de plataforma

Las nuevas herramientas digitales no solo transforman los entornos laborales tradicionales mediante los cambios en los sistemas de contratación, control, vigilancia y gestión laboral; también contribuyen al crecimiento de nuevos modos de organizar y coordinar los procesos de producción. Mientras que los algoritmos de coordinación se usan cada vez más en numerosos campos, se ha puesto mucho énfasis en las iniciativas económicas que se han *hecho posibles* debido a la posibilidad de la coordinación algorítmica, y específicamente la economía de plataformas, donde la aportación de los datos no se limita a su uso para la evaluación del rendimiento de trabajadoras y trabajadores y para establecer rankings y puntuaciones, sino también para organizar la producción e impulsar las relaciones laborales, convirtiendo al empresario prácticamente en invisible y sustituyendo el rol del supervisor por una app o plataforma. En esta sección discutiremos el uso de plataformas basadas en aplicaciones para organizar y distribuir la producción y la carga de trabajo, abordando los desafíos específicos de la economía de plataformas pero también proporcionando un mayor nivel de entendimiento sobre el uso de datos y algoritmos para organizar la mano de obra y la producción.

Las plataformas digitales pueden definirse como espacios de intercambio de información que vinculan oferentes y demandantes de bienes y servicios, así como también intermediación entre particulares para brindar tareas o soluciones específicas (Madariaga, Buenadicha, Molina y Ernst, 2019). Algunas plataformas facilitan actividades económicas, como las plataformas de micro-tareas para trabajadores tipo Amazon Mechanical Turk¹⁹,

¹⁹ <https://www.mturk.com/>

empleo por encargo como Uber²⁰, Cabify²¹, Rappi²² y PedidosYa²³, y plataformas para trabajo autónomo en la nube como Upwork²⁴, Freelancer²⁵ y Workana²⁶. La economía de plataformas no es un mercado homogéneo, y en general se clasifican estas iniciativas en base a sus factores distintivos: localización, habilidades y tipo de actividad (De Groen and Maselli 2016). El primer factor distintivo es la *localización* de los servicios que se ofrecen. Mientras que algunos productos/servicios son totalmente virtuales, como los servicios informáticos, otros se proporcionan desde una localización específica, como los envíos, servicios de transporte o servicios de reparación a domicilio. El segundo factor distintivo es el *nivel de habilidades* que se requiere para llevar a cabo dicho servicio. Además de la localización y habilidades, los investigadores han categorizado las plataformas basándose en el *tipo de actividades* que llevan a cabo y en si son de "capital intensivo" (venden o alquilan una propiedad o posesión) o de "mano de obra intensiva" (personas que realizan un trabajo pagado), el *tipo de pago* (por ejemplo, un precio cerrado frente a trabajadores pujando de forma competitiva) y el *nivel de control sobre el trabajo*, por ejemplo, un alto control frente a uno bajo por parte de los trabajadores (Forde et al. 2017).

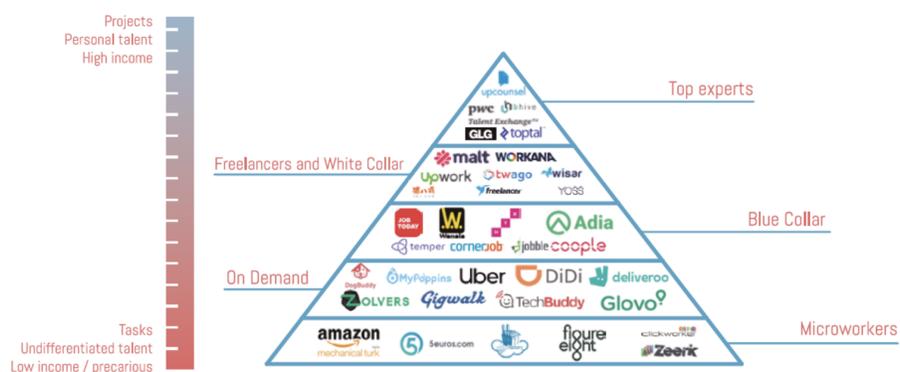


Figura 1: Pirámide de economía de plataforma a lo largo de distintos niveles de tareas y habilidades. Fuente: Cañigueral (2019)

Debido a las características del sector y a la falta de consenso en definir qué es exactamente lo que constituye la "economía de plataformas", aún no está claro cuántas personas están empleadas por estos servicios. Las estimaciones actuales varían ampliamente y según la OECD (2019), uno de los desafíos es precisamente el hecho que los estudios aún no han desarrollado una definición estándar, y que las definiciones actuales son algo ambiguas y pueden malinterpretarse por parte de los encuestados. Un reto adicional es que los

²⁰ <https://www.uber.com/es/en/>
²¹ <https://cabify.com/en>
²² <https://www.rappi.com/>
²³ <https://pedidosya.com/>
²⁴ <https://www.upwork.com/>
²⁵ <https://www.freelancer.com/>
²⁶ <https://www.workana.com/>

conjuntos de datos administrativos como la seguridad social y datos de impuestos no tienen un registro claro del trabajo en plataformas debido a las ambigüedades en su regulación.

A pesar de los obstáculos metodológicos sobre el peso del sector, hay consenso en su rápido crecimiento. Un estudio reciente (Kässi, Lehdonvirta, and Stephany 2021) estima que hay alrededor de 163 millones de autónomos registrados en plataformas online, lo cual supone un enorme crecimiento comparado con el anterior estudio del Banco Mundial (Kuek et al. 2015), el cual estimaba que este número estaba en los 50 millones en 2015.

La economía de plataformas también muestra una tendencia al crecimiento en América Latina y los países del Caribe. Según el informe redactado por Digital Future Society y el Banco Interamericano de Desarrollo (2021), la economía de plataformas en Latinoamérica (países analizados: Argentina, Brasil, Colombia, Ecuador, El Salvador, Guatemala y Uruguay) ha crecido de forma estable desde el estallido de la pandemia del COVID-19. El sector con mayor crecimiento ha sido el de servicios de envíos de supermercados (con una subida del 259%), seguido por el de envío de comida a domicilio (con una subida del 209%), y por los servicios de mensajería y paquetería (con una subida del 141%). Entre marzo y junio de 2020, el número de repartidores ha crecido un 38%. El perfil de los repartidores también ha pasado a ser más diverso incluyendo repartidores nativos, mujeres, estudiantes y personas desempleadas. En los casos de Argentina, Colombia y la República Dominicana, se estima que entre el 0.8% y el 1.0% del empleo está vinculado a trabajo mediante plataformas online (ECLAT/ILO 2021).

La distinción entre economía de plataformas y tradicional está bastante disputada. Mientras que algunos académicos estudian el trabajo en plataformas bajo la categoría de trabajadores autónomos e independientes (De Groen and Maselli 2016), otros enfatizan las características únicas de la economía de plataformas, que la convierten en "una distintiva forma de actividad económica" (Vallas and Schor 2020). El papel de la tecnología y la IA resulta crucial aquí, puesto que es lo que posibilita que en estos entornos todas las decisiones de gestión están automatizadas, incluyendo aquellas asociadas a la coordinación laboral, de forma que los empleadores "desaparecen" detrás de lo que parecen ser procesos "neutrales" donde todas las partes interactúan de forma igualitaria alrededor de una plataforma.

Precisamente uno de los principales aspectos distintivos de la economía de plataformas es el papel de las herramientas digitales y de la IA para diseñar, coordinar y optimizar actividades. Si en los escenarios de contratación y post-contratación descritos más arriba se podría decir que "tu jefe es un algoritmo", para los trabajadores de la economía de plataformas, "el algoritmo es la fábrica", en el sentido en el que es el factor que establece las condiciones de trabajo. Los algoritmos determinan qué rutas son más eficientes, cuántos trabajadores son necesarios en un momento determinado, cómo se fijan los precios y se distribuyen los beneficios, qué rankings y puntuaciones son aceptables o no aceptables y, por supuesto, quién debe ser despedido. Aquí, pues, la tecnología no solo "sigue" al trabajador, sino que establece las reglas del juego.

Marco regulador: leyes, sentencias y decisiones de las Autoridades de Protección de Datos (DPAs)

El uso de algoritmos en contextos laborales no ha sido abordado de forma decisiva en los marcos legales existentes. En el caso específico de la economía de plataformas, el sector más "algoritmizado", el debate legal se ha centrado en la dependencia laboral y contractual o no de los trabajadores y trabajadoras de estas plataformas, con mucha menor atención hasta el momento a los temas vinculados al impacto de los algoritmos.

Hasta hoy, en todo el mundo los temas algorítmicos se han abordado desde la normativa de protección de datos, un derecho vinculado a la protección de la privacidad que reconocen todas las constituciones modernas. En el caso de Argentina, el derecho a la privacidad e intimidad se funda en el artículo 19 de la Constitución Nacional, y son la Ley Nacional No. 25.326/2000 y el Decreto Ejecutivo Presidencial No. 1558/2001 los que regulan la Dirección Nacional de Protección de Datos Personales y la Agencia de Acceso a la Información Pública.

Categorías	Países, leyes y entidades administrativas
Países que tienen leyes nacionales y autoridades administrativas para hacer valer los derechos de protección de datos	<ul style="list-style-type: none"> - Argentina (Ley Nacional No. 25.326/2000, y Decreto Ejecutivo Presidencial No. 1558/2001; "Dirección Nacional de Protección de Datos Personales / Agencia de Acceso a la Información Pública"). - Uruguay (Ley No. 18.331/2008, y Decreto Ejecutivo no. 414/2009; "Unidad Reguladora y de Control de Datos Personales"). - México ("Ley federal de protección de datos personales en posesión de los particulares", 5 de julio de 2010; "Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos"). - Perú (Ley No. 29.733/2011: "Dirección General de Protección de Datos Personales"). - Costa Rica (Ley No. 8.969/2011: "Agencia para la Protección de Datos Personales"). - Colombia (Ley No. 1.581/2012, y Decreto Ejecutivo No. 1.377/2013: "Superintendencia de Industria y Comercio"); - Brasil (Ley Federal No. 13.709/2018: "Autoridade Nacional de Proteção de Dados"). - Panamá (Ley No 81/2019: "Autoridad Nacional de Transparencia y Acceso a la Información").
Países que tienen leyes nacionales, pero que no tienen la autoridad administrativa correspondiente para hacer valer los derechos	<ul style="list-style-type: none"> - Chile (Ley No. 19.628/1999). - República Dominicana (Ley No. 172/2013).

Tabla 1: Estado de las leyes de la protección de datos y autoridades administrativas para hacer valer los derechos en Latinoamérica.
Fuente: Revista Latinoamericana de Economía y Sociedad Digital.

Otro de los puntos evidentes de entrada a los temas algorítmicos lo constituyen las estrategias de IA. Según la Organización para la Cooperación y el Desarrollo Económico (OCDE), Colombia, Argentina y Uruguay son los países de la región que más esfuerzos reales realizan para desarrollar políticas de IA. No obstante, estas políticas no suelen incorporar temas o preocupaciones relacionados con la protección de datos, la privacidad o la ética. Como subraya Fabrício Polido (Polido, 2020), "la construcción de dispositivos de protección debe trascender a los discursos filosóficos y argumentativos sobre ética, seguridad, responsabilidad, justicia", algo que aún no ocurre. En el ámbito del empleo, no hemos encontrado ejemplos en América Latina que apunten a la realización de progresos para traducir la protección de trabajadores y trabajadoras a entornos digitales y de decisión algorítmica.

A nivel global, Europa es la región que más esfuerzos ha realizado para regular la IA y los algoritmos, también en el contexto del empleo. El Reglamento General de Protección de Datos (RGPD) de la UE posee un artículo (Artículo 88) sobre el "Procesamiento de datos en el contexto del empleo", y el Considerando 155 habla sobre el "Procesamiento de datos en el contexto laboral", aunque en términos muy generales. La primera propuesta de la próxima ley sobre la IA de la UE, que se prevé aprobar en 2023, sí incorpora el "empleo, gestión de trabajadores y acceso al auto-empleo" como un contexto de alto riesgo que merece precauciones y garantías adicionales, pero aún no está claro si esta será la versión final de esta legislación.

Un desarrollo reciente dirigido desde el Ministerio de Trabajo español ha sido el uso de la regulación de los servicios de economía de plataformas como una oportunidad para abrir las puertas a la regulación de la "transparencia algorítmica". En el decreto 9/2021, publicado en mayo de 2021, el Gobierno de España reconoce formalmente que todos los trabajadores tienen derecho a ser informados por la empresa sobre los parámetros, normas e instrucciones detrás de los algoritmos y sistemas de IA que intervienen en los procesos de toma de decisiones que afectan a sus condiciones laborales, a sus posibilidades de acceso y mantenimiento de un empleo, y de la configuración de éstos. Este decreto fue anulado tácticamente por la ley 12/2021, del 28 de septiembre de 2021, sin muchos cambios sobre el decreto 9/2021.²⁷ La ley 12/2021 modifica la "Ley del Estatuto de los Trabajadores" mediante una nueva disposición del artículo 64.4, capturando la urgente necesidad de promover la transparencia en el uso de algoritmos en el entorno laboral para proteger a trabajadores y trabajadoras.

Por lo tanto, existe en la legislación española un requerimiento específico para la revelación de información clave sobre los sistemas de IA que afectan a los trabajadores tanto ante ellos como ante sus representantes, un requerimiento que también se ha incluido en la propuesta Directiva de la UE para "mejorar las condiciones de las plataformas de trabajo" publicada el 9 de diciembre de 2021.

Además de la regulación sobre los datos, también es productivo tratar las cuestiones legales relacionadas con las tecnologías usadas en entornos laborales a través de la

²⁷ Ley 12/2021, de 28 de septiembre, por el que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores. Boletín Oficial del Estado, 233, de 29 de septiembre de 2021, 119341 a 119344. Recuperado de <https://www.boe.es/boe/dias/2021/09/29/pdfs/BOE-A-2021-15767.pdf>

legislación sobre protección de datos y derechos laborales ya consolidada, ya que los impactos de la incorporación de algoritmos en contextos de trabajo afectan a cuestiones de privacidad, seguridad y salud laboral, riesgos laborales y a la participación de los trabajadores en los procesos de toma de decisiones que les afectan. También prevemos que en un futuro próximo las autoridades de defensa de la competencia adquieran un rol más protagonista en los temas relacionados con el uso de los algoritmos en el ámbito laboral (y no laboral), sobre todo en lo que respecta a la fijación de precios y su impacto en salarios.

Este enfoque transversal puede observarse en algunas resoluciones judiciales. Por eso el resto de este capítulo resume cómo los tribunales y las Agencias de Protección de Datos (DPAs) han abordado y resuelto casos relacionados con el despliegue y uso de tecnología y algoritmos en el contexto de las relaciones laborales. Este resumen se centra en casos dirimidos en la UE y EEUU, puesto que no se han encontrado casos relevantes en la región latinoamericana.

Fallos judiciales y decisiones de DPAs

Existe un creciente número de litigios que abordan explícitamente el rol de los algoritmos en general y aquellos usados en el empleo y el entorno laboral en particular. Los casos contra los algoritmos en el ámbito laboral varían mucho dependiendo de la jurisdicción nacional donde los juicios tienen lugar, la forma en la que se formulan las cuestiones legales, qué disposición de la legislación se está vulnerando y el tipo de derechos fundamentales que se evocan. En los países con tradición legal inglesa como EEUU, Canadá o Reino Unido, las "debidas garantías procesales" parecen ser un marco legal adecuado para desafiar a los algoritmos (AI Now Institute 2018). Las denuncias relacionadas con las garantías procesales plantean temas de equidad y alegan falta de notificación, explicación o posibilidad de intervenir o impugnar el proceso de incorporación de tecnologías en el ámbito laboral. Aunque en términos estadísticos los algoritmos siempre prometen mejorar las probabilidades de una decisión justa, las garantías procesales adecuadas no se cumplen si el sistema no fomenta el estado de derecho ni respeta la dignidad de los individuos afectados por tales decisiones (Camilleri 2018).

Otra base común para los litigios contra los algoritmos es la detección de errores en el diseño y/o implementación del software. Esta base legal requiere que los expertos accedan a la información técnica del sistema (AI Now Institute 2018), lo cual puede ser complicado teniendo en cuenta la protección del secreto comercial o la opacidad de los sistemas de aprendizaje automático. Es por eso que conseguir transparencia y acceso público a los algoritmos se considera una meta importante para aquellos interesados en denunciar legalmente sistemas de decisiones algorítmicos.

En el caso de los países de la UE, el Reglamento General de Protección de Datos (RGPD) proporciona la vía principal para las demandas contra los sistemas de decisión automatizados. El RGPD es considerado como el reglamento de privacidad y seguridad más

estricto del mundo. Impone obligaciones para las organizaciones de cualquier lugar que recopilen datos asociados a ciudadanos europeos.²⁸ Artículo 5. 1-2 identifica siete principios de protección y responsabilidad, incluyendo: 1) Legalidad, equidad y transparencia en el procesamiento de datos; 2) Limitación de finalidades; 3) Minimización de los datos; 4) Precisión; 5) Limitación de almacenaje; 6) Integridad y confidencialidad; y 7) Responsabilidad. El Artículo 22(1) del RGPD establece que: *Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar.* Este artículo establece "el derecho a obtener una intervención humana" cuando la persona es sometida a una toma de decisiones puramente algorítmica y crea así un importante fundamento jurídico para el litigio.

En el cuadro que sigue proporcionamos un resumen de las causas legales más recientes que abordan temas tecnológicos y algorítmicos en el entorno laboral en diferentes contextos nacionales. Algunas de estas causas están aún pendientes, algunas han sido desestimadas por los tribunales, aunque dejando la puerta abierta a futuros litigios, y otras han sido ya resueltas y proporcionan precedentes para futuros casos. Aunque esta lista no pretende ser exhaustiva, cubre las principales cuestiones legales y los fallos más relevantes surgidos de casos contra los algoritmos en el contexto de las relaciones laborales y el empleo.

En cuanto a las DPAs, aunque en el ámbito de las relaciones laborales éstas tienden a centrarse en temas de protección de datos "estricta" (uso de datos de los empleados videovigilancia, etc.) existen ya decisiones que abordan específicamente temas vinculados a los algoritmos y, específicamente, a las herramientas utilizadas por empresas de la economía de plataformas.

País	Descripción del Caso	Tribunal / Agencia	Fecha de Decisión	Estado
EEUU	La Federación de Profesores de Houston (HFT) demandó al Distrito Escolar Independiente de Houston contra las evaluaciones algorítmicas opacas de los profesores públicos en base a la 14ava Enmienda, Proceso Debido y Clausulas de Protección de Igualdad	Tribunal del Distrito de los Estados Unidos, S.D. Tejas, División de Houston	2017	El tribunal encontró culpable al distrito de violar los derechos del debido proceso de los profesores ante su negativa a divulgar los datos internos del sistema algorítmico (el modelo de valor añadido)
EEUU	La Federal Trade Commission inició una investigación de Amazon Logistics por sospecha de que las propinas de los clientes	FTC	2019	Después de que la FTC emitiera una "consent order" (con rango de ley) en 2019, Amazon aceptó pagar más de 60 millones

²⁸ <https://gdpreu/what-is-gdpr/>

	estaban siendo destinadas a sufragar salarios desde 2016.			de dólares a los trabajadores afectados.
Italia	La inspección de la DPA a la empresa Foodinho, filial de Glovo, descubrió que Foodinho no había informado de forma adecuada a sus "riders" sobre el funcionamiento del sistema de IA, no había implementado protecciones para asegurar su equidad y fiabilidad y no había puesto en marcha procesos para constatar la intervención humana y desafiar la toma de decisiones de los algoritmos	DPA	2019	La DPA impuso una multa de 2,6 millones de EUR a Foodinho y ordenó medidas correctivas a tomar en 60 días.
EEUU	Discriminación racial del sistema de puntuación por estrellas de Uber	Tribunal del Distrito de California	2020	Desestimado con permiso a presentar una nueva demanda
EEUU	Se presenta una demanda colectiva contra la Agencia de Seguros de Desempleo de Michigan (UIA) por el presunto uso de un sistema de automatización deficiente (MiDAS) para detectar fraude (fallando en un 93% de los casos). Los demandantes declaran que se violaron los derechos del debido proceso.	Tribunal Supremo de Michigan	2017	El Tribunal de Reclamaciones deniega la petición del acusado y falla que la queja se presentó durante los seis meses requeridos. El Tribunal de Apelaciones revierte la decisión del anterior tribunal. El caso es apelado al Tribunal Supremo de Michigan.
EEUU	El uso por parte de IBM de IA para ocultar los despidos discriminatorios de trabajadores más mayores	Tribunal del Distrito de Tejas y otros litigios pendientes en distintos estados	2021	En proceso
Reino Unido	Acusación del Servicio de Correos Británico a +700 empleados por fraude basándose en la información defectuosa proveniente de un sistema de IA	Tribunal de Apelación Criminal	2021	39 de las condenas se han revertido, lo cual despeja el camino para nuevas apelaciones de otros empleados
Países Bajos	Los conductores de Uber impugnan el "robo-despido", consistente en la desactivación automática de los trabajadores sobre los que el algoritmo	Juzgado del Distrito de Ámsterdam	2021	Las decisiones de Uber de desactivar las cuentas de los conductores no están totalmente automatizadas según el RGPD

	identifica posible actividad fraudulenta, bajo el RGPD			
Países Bajos	Los conductores impugnan a Uber y Ola debido a los derechos de acceso a datos y gestión algorítmica bajo el RGPD	Juzgado del Distrito de Ámsterdam	2021	El juzgado rechaza estas reclamaciones específicas, mientras pide a ambas compañías que proporcionen más datos de los que han mostrado hasta ahora.
Italia	Los conductores de Deliveroo impugnan el algoritmo de rankings de reputación por motivos de discriminación	Corte de Bolonia	2021	La Corte juzga al sistema de rankings como discriminatorio, ya que hace caso omiso de las causas de protección legales de negación de trabajo.
Italia	La Autoridad de Protección de Datos Italiana (DPA) sanciona el algoritmo de rankings de reputación de Mevaluate por vulnerar la privacidad. Mevaluate apela al Tribunal de Roma, que anula la orden de la DPA. La DPA apela el caso ante la Corte de Casación.	Corte de Casación (La Corte Suprema Italiana)	2021	La Corte de Casación anula la decisión del Tribunal de Roma y falla que los algoritmos han de ser transparentes para que su consentimiento sea válido, y el sujeto de los datos debe ser informado de forma apropiada sobre la finalidad del procesamiento de datos.

Tabla 2: Resumen de casos.

Finalmente, es importante destacar que algunos temas no han llegado a tribunales, pero aún así han logrado cambios en las políticas de empresas que utilizan algoritmos de forma intensiva. En EEUU, por ejemplo, Uber llegó a un acuerdo en 2021 con los abogados generales de Washington y Pennsylvania para mejorar la transparencia de su sistema de fijación de precios, que impacta directamente sobre los salarios que reciben los conductores de ese servicio.²⁹ Otros colectivos de EEUU, como los trabajadores de plataformas de distribución de comida, han organizado movilizaciones y huelgas para exigir transparencia sobre las propinas que abonan los usuarios, que alegan no siempre recibir de forma completa. En España, hace tiempo que los taxistas denuncian prácticas de acuerdo de precios entre los proveedores de servicios de transporte, lo que está siendo investigado por las autoridades de competencia.

²⁹ [Uber Fats pressured by attorneys general on pricing transparency | Nation's Restaurant News \(nrn.com\)](https://www.nrn.com/uber-fats-pressured-by-attorneys-general-on-pricing-transparency)

Buenas prácticas y avances en transparencia algorítmica

A medida que proliferan las tecnologías y herramientas algorítmicas, existe una creciente demanda de transparencia. Desde los motores de búsqueda y recomendación a los sistemas de navegación o reconocimiento facial, la IA se ha convertido en parte de nuestro día a día, y aunque es cierto que la transparencia en sí no garantiza la equidad, la seguridad y la legalidad de la IA, la transparencia permite a expertos y expertas externos, usuarios y al regulador acceder y desafiar las suposiciones y decisiones que se toman en base a datos personales. Además, asegurar que las herramientas de toma de decisiones automatizadas están abiertas y son transparentes para la evaluación pública promueve la responsabilidad, lo que es un componente crucial para generar confianza en la IA.

Teniendo esto en consideración, algunas administraciones han tomado la iniciativa de desarrollar prácticas para proporcionar información sobre cómo las organizaciones públicas usan la IA para la toma de decisiones, así como entender cómo los principios de equidad y responsabilidad se traducen en ellas.³⁰ En 2018, la ciudad de Nueva York fue la primera en desarrollar una división encargada de realizar propuestas para supervisar el uso en la ciudad de sistemas de decisión automatizados.³¹ No obstante, el Grupo de Trabajo de Sistemas de Decisión Automatizados, o Grupo de Trabajo ADS, dirigido por la oficina del alcalde, no duró demasiado debido a una falta de consenso sobre sus objetivos, procedimientos y competencias.

Desde entonces, han surgido otras iniciativas específicas, como los registros, etiquetas y calificaciones. Revisaremos estas referencias como contexto de la propuesta de una herramienta de transparencia algorítmica concreta: el prospecto algorítmico.

Registros algorítmicos

Los registros de IA son bases de datos públicas donde la información clave sobre los sistemas de IA se guarda y comparte. Son una iniciativa prometedora para recopilar información sobre la existencia de estos sistemas, sobre las decisiones y suposiciones de las herramientas algorítmicas usadas en instituciones públicas y empresas privadas, y un modo efectivo de hacer que la transparencia algorítmica sea real y concreta. Ámsterdam y Helsinki han sido las primeras ciudades en adoptar los registros de IA como un modo de incrementar la confianza de los ciudadanos en los sistemas usados por las organizaciones gubernamentales.³² Siguiendo la iniciativa de estas ciudades, otros países también han mencionado su intención de abrir los algoritmos, pero sin una acción concreta al día de hoy.

³⁰ <https://research.ngi.eu/two-days-to-change-the-internet-the-ngi-policy-summit-2020/>

³¹ <https://www1.nyc.gov/site/adstaskforce/index.page>

³² Los registros de IA en Ámsterdam y Helsinki fueron desarrollados en colaboración con Ssaidot, una plataforma finlandesa de gestión de IA

La idea tras los registros de IA es capturar las tres dimensiones principales de un sistema de IA: su desarrollo y gestión técnicos, su evaluación y su uso. En la práctica, los registros de IA actúan como una forma estandarizada e indexada de documentar y consultar las decisiones y suposiciones realizadas durante el ciclo de vida de un algoritmo, mientras que también permiten a los usuarios proporcionar su opinión para poder mejorar dicho sistema. La información se proporciona de una forma accesible con distintos niveles de abstracción: desde una visión general a una foto más detallada, dependiendo de la naturaleza e intereses de los actores que hagan la búsqueda.

Los registros de IA deben proporcionar información entendible sobre 1) Qué, dónde y cómo se están usando las aplicaciones de IA, 2) Qué conjuntos de datos fueron usados para entrenar el modelo, 3) Qué algoritmos han sido evaluados para identificar sesgos o riesgos. Saidot, la organización que ha ayudado a Ámsterdam y Helsinki a desarrollar sus propios registros, mantiene que cualquier registro debe contener los siguientes elementos:

- **Resumen:** Quizás la parte más importante, pero también la más difícil. La documentación debe escribirse de una forma simple, clara y entendible para que el contenido resulte asequible a todo el mundo independientemente de sus conocimientos técnicos. El enfoque debería centrarse en definir el objetivo o el beneficio social del sistema, cómo y cuándo se está usando, quién sufre el mayor impacto del mismo y cuáles son los impactos esperados, así como un resumen a más alto nivel de cómo funciona el sistema.
- **Responsabilidad:** Esta sección aborda la trazabilidad de responsabilidades. Esto implica que aquellos responsables del desarrollo y mantenimiento del sistema lo hagan durante todo su ciclo de vida. Esto se traduce en mencionar qué entidades son responsables del sistema, cómo se puede contactar con ellas y cuáles son los actores externos involucrados. En este sentido se recomienda encarecidamente tener auditores externos que revisen cada sistema periódicamente.
- **Conjuntos de datos:** Esta parte está pensada para facilitar el acceso a los conjuntos de datos usados para entrenar a la tecnología, así como identificar las posibles fuentes de errores o sesgos. Se debería incluir el tipo y origen de los datos, así como su calidad. Además, cualquier uso de datos externos o comercialización debe ser reportado para permitir la trazabilidad de los éstos.
- **Procesamiento de Datos:** Eso es, información sobre la arquitectura del sistema. Esta sección será algo más técnica e incluirá información sobre los modelos matemáticos y computacionales usados. También se deberían incluir métricas sobre la interpretación y medición del rendimiento para facilitar la identificación de posibles desajustes.
- **No Discriminación:** Esta sección debe cubrir los mecanismos incluidos en el sistema para identificar y mitigar los sesgos y los impactos injustos sobre grupos protegidos.
- **Supervisión Humana:** Esta parte debe incluir las formas en las que el sistema está diseñado para asegurar que una persona o grupo preservan el control, la autoridad y

la autonomía sobre la decisión tomada por el sistema. Para poder hacerlo, se deben listar tanto los lugares y las formas mediante las cuales los humanos pueden ejercer el control sobre el sistema como la formación necesaria para tener las competencias que permitan hacerlo.

- **Riesgos:** En este punto, se identifican riesgos y medidas para mitigarlos, así como la solución intermedia entre riesgos y beneficios o la evaluación de impacto en la privacidad.
- **"Explicabilidad":** Descripción de los mecanismos que permitan comprender las decisiones automáticas tomadas por los algoritmos.
- **Referencias:** Una lista de documentación ampliada sobre políticas de privacidad, modelos de gobierno, interacciones y transacciones con terceros, margo regulador y otras disposiciones.

Proporcionando información sobre estos aspectos, los registros de IA pretenden mejorar la confianza de los ciudadanos en dichas tecnologías. Incentivando las aportaciones de usuarios mediante canales de retroalimentación, la transparencia se convierte en un instrumento efectivo para generar confianza en la IA, a la vez que se abre un espacio para materializar las responsabilidades (Floridi, 2020: 545). Los registros deberían también brindar a los científicos, activistas, periodistas y todos los actores de la sociedad civil una oportunidad valiosa de participar y ejercer control sobre los algoritmos. Los servidores civiles también se pueden beneficiar de los registros de IA, puesto que también funcionan con herramientas para la gobernanza de la IA y documentan su uso y funcionamiento.³³

No obstante, los investigadores del Instituto de Internet de Oxford Corinne Cath y Fieke Jansen argumentan que los ejemplos existentes de registros de IA deben verse con cierto escepticismo, puesto que el registro de IA en Helsinki tiene cinco entradas, y el de Ámsterdam tres. Además, los algoritmos registrados son de naturaleza muy burocrática, como el uso de chatbots, el procesamiento automático de queja o un sistema que calcula la probabilidad de que se anuncien alquileres turísticos ilegales (Cath and Jansen, 2021:3), consolidando así la invisibilidad de sistemas de IA más controvertidos.

Además, esos registros solo recopilan información sobre los sistemas instalados por los gobiernos locales de esas ciudades, aunque Ámsterdam, por ejemplo, no sólo usa IA en servicios públicos, sino que también comercializa datos y compra datos de terceros (Munten, 2019). Estos aspectos, sumados al hecho que los registros de algoritmos solo permiten un control expost de los algoritmos, hacen que a menudo las iniciativas de transparencia algorítmica tengan poca credibilidad como ejercicios de transparencia real. Un enfoque vertical dificulta que se establezcan conversaciones de calado sobre si éstas tecnologías deberían siquiera instalarse. En este sentido, es imprescindible situar esas iniciativas en un marco más amplio sobre la gobernanza de los datos desde una perspectiva de protección de derechos y participación.

³³ https://uploads-ssl.webflow.com/5c8abedb10ed656ecfb65fd9/5f6f334b49d5444079726a79_AI%20Registers%20-%20White%20paper%2010.pdf

Etiquetas y calificaciones

En EEUU, una práctica de transparencia que ha generado bastante interés es la de usar un tipo de "etiquetas nutricionales" para los sistemas de IA, centrándose en soluciones visuales para afrontar el reto de la transparencia. Con la Ley de Alimentos y Medicamentos de 1906, EEUU fue el primer país del mundo en aprobar una ley de protección del consumidor que dio lugar a la creación de la Administración de Medicamentos y Alimentos (FDA), la cual reforzó el etiquetado obligatorio de los alimentos desde 1913 en adelante. El propósito de este etiquetado era elevar los estándares de la industria de los alimentos para evitar el fraude a los consumidores. Al inicio del siglo 20, la conveniencia de los alimentos envasados derrotó a la compra a granel, por lo que la necesidad de incluir la cantidad y composición del contenido se vio reflejada en la ley de 1906.³⁴ Aun así, no fue hasta que se aprobó la Ley de Educación y Etiquetado Nutricional en 1990, cuando este etiquetado se convirtió en obligatorio, dando lugar a las famosas etiquetas de Información Nutricional que empezaron a aparecer en 1994. Esta etiqueta gráfica debe incluirse en cualquier alimento empaquetado regulado por la FDA y tiene las mismas iniciales en todo el mundo. En el año 2005 se incorporó el mencionado Código a la Resolución del Grupo Mercado Común (GMC) N°26/03 1, que establece los requisitos para el rotulado general. Y, específicamente, desde el año 2006 en Argentina es obligatoria la declaración del rótulo nutricional a través de la incorporación de las Resoluciones GMC N°46/03 y 47/032 a la normativa nacional.³⁵ Incluir la información nutricional en una etiqueta es un gran paso para asegurar de que el consumidor tiene a mano la información relevante para decidir si desea o no consumir un producto.

Siguiendo esta lógica, un equipo de investigadores de Oxford desarrolló el sistema NutriScore (Imagen 1), unos códigos de color que representan la idoneidad del producto en una escala del verde al rojo (o de la A a la E). Esta puntuación se obtiene mediante un algoritmo que considera las contribuciones positivas y negativas del producto, alcanzando un valor global que se traduce a un color y su respectiva letra. A y verde quieren decir que es el más sano, mientras que E y rojo suponen la opción menos sana (Galan, et al., 2017).



Imagen 1: Modelo Nutriscore. Fuente: www.ocu.org

³⁴ <https://www.fooddive.com/news/the-origins-and-evolution-of-nutrition-facts-labeling/507016/>

³⁵ <http://www.alimentosargentinos.gob.ar/contenido/publicaciones/calidad/Guias/GuiaRotulo.pdf>

Este sistema ha sido validado por la Agencia de Normas Alimentarias (FSA) y la Organización Mundial de la Salud (OMS), pero no parece existir aún en la región latinoamericana. No obstante, en octubre de 2020 entró en vigor en México la NOM 051, conocida como "etiquetado frontal", por lo que los fabricantes de alimentos y bebidas no alcohólicas deben advertir del exceso de ingredientes críticos para los consumidores (Imagen 2). Un estudio sobre las diferentes alternativas de fuentes de etiquetado de envoltorios concluyó que NutriScore es, efectivamente, la forma más visualmente efectiva de comunicar estos procesos y estándares (Egnell, et al., 2018). A día de hoy, las puntuaciones visuales se usan también para mostrar fácilmente el impacto medioambiental y el consumo de otros bienes, como electrodomésticos, es un buen modo de alentar al consumidor para tener esto en cuenta y motiva a las compañías a crear productos más eficientes y sostenibles al estar éstos puntuados en una escala de la A a la G.



Imagen 2: Etiquetado frontal. Fuente: *Etiquetado frontal: el debate sobre la reformulación - The Food Tech*

Usando este ejemplo histórico de una solución parcial que ha funcionado para proteger los derechos del consumidor y rendir prácticas seguras, transparentes y responsables, el AI Ethics Impact Group del Bertelsmann Stiftung ha realizado una propuesta para un Etiquetado Ético de la IA que puntúa seis dimensiones:

- Transparencia
- Responsabilidad
- Privacidad
- Justicia
- Fiabilidad
- Sostenibilidad Medioambiental

Esta puntuación se obtiene mediante la combinación de dos herramientas diferentes. Por un lado, el modelo VCIO (Valores, Criterios, Indicadores y Observaciones) permite especificar y operativizar los valores relacionados con diferentes temas, consigue cerrar la brecha entre las discusiones teóricas sobre los valores y los problemas de la vida real, y proporciona un espacio para solventar conflictos entre valores. Además, usa una matriz de riesgos bi-dimensional para clasificar los distintos contextos de aplicación de los sistemas de IA,

proporcionando los requerimientos mínimos que el sistema de IA debe cumplir en un contexto de aplicación específico.³⁶

Esta iniciativa es quizás la más cercana a la metodología del etiquetado NutriScore en la industria de la alimentación o puntuación de los productos eléctricos. Además, la Etiqueta Ética (Imagen 3) no solo informa sobre las distintas categorías a considerar y valorar, sino que también proporciona una metodología para obtener esa puntuación, lo que es un buen punto de inicio para futuros trabajos.



Imagen 3: Etiqueta ética. Fuente: *Ethics of Algorithms | From principles to practice: How can we make AI ethics measurable? - Ethics of Algorithms*

Otra iniciativa relevante es la de Open Ethics, que propone un etiquetado voluntario que cubra los tres componentes básicos de la tecnología algorítmica:

- Datos de aprendizaje (propietarios, acceso limitado o abierto a los conjuntos de datos o análisis heurístico basado en reglas).
- Código Fuente (propietario o código abierto)
- Espacio de Decisión (restringido o no restringido)

Este sistema resulta en un sencillo resumen visual que el usuario final puede comprender y donde se presenta la información básica sobre estas tres dimensiones.³⁷

³⁶ <https://www.ai-ethics-impact.org/resource/blob/1961130/c6db9894ee73aefa489d6249f5ee2b9f/aieig---report---download-hb-data.pdf>

³⁷ <https://openethics.ai/label/>

“Model cards”

Debido a la confusión generalizada sobre qué son los algoritmos y lo que hacen, y su potencial poder de transformación, en 2019 Google propuso que todos los modelos de IA deberían ir acompañados de documentación que detallara sus características. Para esto, la compañía desarrolló un marco para facilitar, automatizar y potenciar este tipo información sobre los modelos a través de lo que ellos llaman “model cards” (Mitchell, et al., 2019), que podría traducirse como documento sobre el modelo.

Las “model cards” proporcionan una fotografía rápida del modelo utilizado, así como información sobre los procedimientos de evaluación de su rendimiento y otros datos importantes. De forma similar a como funcionan las etiquetas de nutrición, las model cards para la IA detallan cuáles son las técnicas que se han usado en el desarrollo de los modelos algorítmicos y como se asegura su rendimiento óptimo. Por supuesto, esto tiene un significado distinto para cada sistema y actor, pero facilita el entendimiento común de lo que pasa dentro del modelo.

Según Google, las model cards deben incluir:

- Detalles del modelo: Información básica sobre el modelo, como quién lo ha desarrollado, su versión y tipo, procedimientos de aprendizaje, igualdad de parámetros, licencia, etc.
- Finalidad: Finalidad, usuarios finales y usos fuera del ámbito de desarrollo identificados en la fase de diseño
- Factores: Grupos protegidos, factores relevantes y evaluación de factores
- Métricas: Métricas de rendimiento del modelo, límites y aproximación a variantes que reflejen casos potenciales de uso en la vida real e impactos
- Datos de Evaluación: Bases de datos, pre-procesado y la motivación de las bases de datos usadas para los análisis cuantitativos
- Datos de Entrenamiento: Distribución sobre varios factores en las bases de datos o replicando la sección de datos de evaluación
- Análisis Cuantitativo: Resultados unitarios e interseccionales
- Consideraciones Éticas
- Recomendaciones

Este es un ejercicio más narrativo, similar a una Evaluación de Impacto o al reciente Modelo de Transparencia Algorítmica desarrollado por el gobierno de Reino Unido,³⁸ que cubre los siguientes puntos:

- Propietarios del algoritmo: Incluyendo proveedores externos e internos, propietarios responsables ante la ley, cualquier otro propietario.

³⁸ <https://www.gov.uk/government/publications/algorithmic-transparency-template/algorithmic-transparency-template>

- Finalidad de la herramienta: Descripción del uso final de la herramienta, sus beneficios y alternativas. Esta parte también puede incluir especificaciones técnicas genéricas.
- Impacto en la toma de decisiones: Descripción del impacto del algoritmo. Este campo incluye explicaciones sobre cuánta información es compartida con el personal técnico, si hay intervención humana y si necesitan formación para intervenir. Los procesos de apelación se pueden incluir aquí.
- Datos: Esta es la descripción de los conjuntos de datos usados para el aprendizaje, despliegue del modelo, links a los datos o cualquier información relacionada con los datos.
- Evaluación del impacto: Este campo describe si se han realizado evaluaciones de impacto, en qué fechas y dónde se pueden encontrar los informes publicados.
- Riesgos: riesgos potenciales y medidas de mitigación relacionadas con el uso de la herramienta.

Este enfoque narrativo tiene ventajas y desventajas; puede tener más matices pero es menos visual, y la narrativa puede ser vaga o depender de la disponibilidad de redactores dentro de una organización. No obstante, los distintos planteamientos también pueden funcionar juntos, como hacen las etiquetas de NutriScore y de Información Nutricional: la puntuación de colores puede dar una visión general y rápida del modelo mientras que las model cards proporcionan datos más específicos y detallados que pueden resultar útiles para el control, supervisión y auditoría. La propuesta que detallamos a continuación se aleja de este enfoque narrativo.

Hacia un Prospecto Algorítmico

Basándonos en las experiencias expuestas más arriba y teniendo en cuenta la próxima Ley Europea de IA, se impone avanzar de forma decisiva hacia la consolidación de prácticas que puedan hacer tangibles la transparencia y responsabilidad de los sistemas de IA. El prospecto de IA propuesto aquí es fruto de las experiencias descritas más arriba y de las prácticas de control y garantías que existen en otro ámbito sensible como es el médico (y que también están informando el desarrollo de la futura Ley de IA europea), y su uso se prevé en todos los contextos de desarrollo e implementación de sistemas de IA de "alto riesgo" como son el trabajo, la salud, la seguridad, la educación, la justicia criminal o los servicios sociales.

El prospecto algorítmico es altamente técnico para asegurar que sean los desarrolladores los responsables de proporcionar una información que es necesaria e imprescindible para la rendición de cuentas de los sistemas que deciden sobre aspectos clave de la vida de las personas. Por lo tanto, se aleja de los enfoques narrativos puesto que, en nuestra experiencia, los requerimientos narrativos suelen traspasar el peso del

cumplimiento de los desarrolladores y equipos de producto a los abogados, una situación que daña la transparencia y capacidad de supervisión efectiva. Todas las soluciones existentes pueden, por supuesto, funcionar de forma combinada, pero pensamos que es importante tener un estándar que resulte técnico y entendible, así como que proporcione no sólo la información relevante de los sistemas sino también sobre las decisiones técnicas y las rutas que se pueden seguir para obtener más información y protección. Proponemos que el Prospecto Algorítmico se convierta en algo tan habitual e imprescindible como los prospectos médicos, y que contribuyan a una mejor gestión de los riesgos en los sistemas de IA.

Nuestra propuesta inicial para el prospecto algorítmico, que necesita ser pilotada y consensuada con los actores más relevantes, incluye los siguientes campos:

1. REGISTRO

1.1. Número de registro

Si se adoptara un registro algorítmico, el número de registro vendría dado y permitiría la trazabilidad del modelo y sus posteriores versiones

2. DETALLES DEL MODELO

2.1. Propietario

Detalles fiscales y de contacto del organismo propietario

2.2. Desarrollador

En caso de ser distinto al propietario, detalles fiscales y de contacto del organismo que desarrolló el modelo

2.3. Versión del Modelo

Si hubiera habido iteraciones o distintas versiones

2.4. Tipo de Algoritmo

¿Cuál es la arquitectura básica del modelo?

2.5. Código Abierto / Software Libre

De ser así, facilitar el link al repositorio

2.6. Licencia

Información de licencia sobre el modelo/algoritmo/sistema

2.7. Opinión / Feedback Sobre el Modelo

Datos de contacto para poder ofrecer feedback sobre el modelo

3. OBJETIVO Y USO INTENCIONADO

3.1. Uso Intencionado Principal del Algoritmo

¿Se diseñó el modelo para uso genérico o para alguna aplicación específica?

3.2. Usuarios Principales del Algoritmo

¿Qué tipo de público consumirá y utilizará el algoritmo? ¿Cómo se utiliza el servicio por parte de consumidores o usuarios?

3.3. Área de Impacto

¿Se utiliza el modelo en alguna de estas situaciones? De ser así, especificar cual.

- *Infraestructuras críticas* (e.g. transporte), que pudieran poner en riesgo la vida o la salud de los ciudadanos;
- *Fines educativos*, que pudieran condicionar el acceso a la educación o al futuro profesional de una persona (e.g., corrector de exámenes)
- *Componentes de seguridad de productos* (e.g. aplicación de IA en cirugía asistida por robot);
- *Empleo, relación con trabajadores y trabajadores autónomos* (e.g. software de selección de CV para candidatos en un puesto de trabajo);
- *Servicios esenciales públicos o privados* (e.g. credit scoring para aceptar o denegar el acceso a crédito);
- *Orden público* que pudiera interferir con los derechos fundamentales de los ciudadanos (e.g. evaluación de la fiabilidad de pruebas judiciales);
- *Migración, asilo y gestión de fronteras* (e.g. verificación de la autenticidad de los documentos de identificación);
- *Administración de justicia y procesos democráticos* (e.g. aplicación de la ley en un caso concreto, demarcación de los distritos electorales).

3.4. Nivel de Riesgo

(Esto debe ser determinado externamente)

3.5. Uso Fuera del Ámbito de Diseño

¿Existen algunas situaciones o contextos en los que el modelo no debiera ser utilizado?

4. FACTORES

4.1. Factores Relevantes

¿Cuáles son los factores para los cuales el desempeño del modelo podría variar?
¿Cómo han sido determinados? (e.g. entornos con poca luz para sistemas de reconocimiento facial)

4.2. Factores de Evaluación

¿Qué factores son monitoreados y por qué han sido seleccionados?

4.3. Poblaciones Afectadas

Definir las poblaciones afectadas o impactadas por el uso del sistema, de forma directa e indirecta

5. MÉTRICAS

5.1. Medidas de Desempeño del Modelo

¿Cómo se evalúa el desempeño del modelo?

¿Por qué se han seleccionado estas medidas?

¿Hay alguna forma de verificar los valores? (e.g., mediante una API)

¿Ha evaluado el desempeño del sistema alguna organización independiente?

5.2. Umbrales de Decisión

(En caso de que se hayan utilizado umbrales de decisión) ¿Cuáles son, y por qué se han seleccionado estos?

5.3. Variación

¿Cómo se han medido y estimado estas métricas? (e.g. desviación estándar, variancia, intervalos de confianza,...) — También se deberían incluir detalles de cómo se han aproximado los valores (e.g., medias,...)

6. DATOS DE EVALUACIÓN

6.1. Bases de Datos

¿Qué bases de datos se han utilizado para hacer la evaluación cuantitativa del desempeño del modelo?

¿Se han utilizado datos sintéticos?

¿Permite el servicio evaluar diferencias entre los datos de desarrollo y los de uso?

6.2. Justificación

¿Por qué se seleccionaron estas bases de datos?

6.3. Preprocesamiento

¿Cómo se han preprocesado los datos?

¿Fue necesaria alguna transformación de los datos?

7. DATOS DE ENTRENAMIENTO

7.1. Bases de Datos

¿Qué bases de datos se han utilizado para entrenar el modelo?

¿Se han utilizado datos sintéticos?

7.2. Justificación

¿Por qué se seleccionaron estas bases de datos?

7.3. Preprocesamiento

¿Cómo se han preprocesado los datos?

¿Fue necesaria alguna transformación de los datos?

8. ANÁLISIS CUANTITATIVOS

8.1. Métricas de Equidad

Define las métricas de equidad relevantes: grupos protegidos, edad, género, localización geográfica,...

8.2. Resultados Unitarios

¿Cuál es el desempeño del modelo en relación a cada uno de estos factores? Añadir variaciones (e.g. mediante barras de error)

8.3. Resultados Interseccionales

¿Cuál es el desempeño del modelo en relación a la intersección de los diversos factores? Añadir variaciones (e.g. mediante barras de error)

8.4.1. *(si se trata de un problema de clasificación)* False Positive, False Negative, False Discovery, and False Omission Rates

Para cada categoría relevante, especificar si se trata de valores reales o de laboratorio, e incorporar intervalos de confianza

8.4.2. *(si se trata de un problema score-based)* Distribution of Measured Metrics Across Groups

Medidas de tendencia central (moda, mediana y media), así como de dispersión o variación (rango, cuartiles, desviación absoluta, variancia y desviación estándar)

9. CONSIDERACIONES ÉTICAS

9.1. Datos

¿Utiliza el modelo datos sensibles sobre los grupos protegidos?

9.2. Impacto Severo

¿Ha sido el modelo diseñado para informar o tomar decisiones sobre cuestiones centrales de la vida humana? (Seguridad, salud,...)

9.3. Sesgo y Problemas Éticos

¿Se han identificado instancias de sesgo, o problemas de seguridad o éticos derivados del uso del servicio?

9.4. Riesgos

¿Qué riesgos pueden derivar de tales sesgos, y cuál sería la población principal afectada?

9.4. Mitigación

¿Qué medidas se han adoptado para mitigar el sesgo?

9.5. Human-in-the-loop

Descripción del rol específico de los humanos en el desarrollo del sistema

9.6. Testeo con Grupos Interesados

¿Se ha involucrado algún grupo interesado durante las fases de desarrollo y testeo previas a la implementación del sistema?

Esta primera propuesta, para estar completa, deberá acompañarse de una guía de uso con las definiciones clave, ejemplos de implementación y otras instrucciones que faciliten su cumplimentación. El Prospecto Algorítmico puede complementar otros enfoques más visuales y/o menos esquemáticos, que se pueden desarrollar en paralelo. En las fase de pilotaje, se valorará:

- Validez y practicidad del enfoque
- Aplicabilidad para abordar temas vinculados a:
 - Transparencia en fijación de precios/salarios
 - Discriminación en distribución de carga de trabajo
 - Discriminación en selección de perfiles
 - Imposición de ritmos
 - Problemas vinculados con la seguridad y salud laboral

Después de la validación del enfoque, se elaborará el documento de apoyo a la cumplimentación del Prospecto Algorítmico.

Bibliografía

- AI Now Institute. 2018. "Litigating Algorithms: Challenging Government Use of Algorithm Decision Systems." AI Now Institute.
http://www.law.nyu.edu/sites/default/files/litigatingalgorithms_0.pdf.
- Ajunwa, Ifeoma, and Daniel Greene. 2019. "Platforms at Work: Automated Hiring Platforms and Other New Intermediaries in the Organization of Work." In *Work and Labor in the Digital Age*, edited by Daniel Greene, Steve P. Vallas, and Anne Kovalainen, 33:61–91. Bingley: Emerald Publishing. <https://doi.org/10.1108/S0277-283320190000033005>.
- Almalis, Nikolaos D., George A. Tsihrintzis, and Nikolaos Karagiannis. 2014. "A Content Based Approach for Recommending Personnel for Job Positions." In *IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications*, 45–49. <https://doi.org/10.1109/IISA.2014.6878720>.
- Aneesh, A. 2009. "Global Labor: Algocratic Modes of Organization." *Sociological Theory* 27 (4): 347–70. <https://doi.org/10.1111/j.1467-9558.2009.01352.x>.
- Askehave, Inger, and Karen Korning Zethsen. 2014. "Gendered Constructions of Leadership in Danish Job Advertisements." *Gender, Work & Organization* 21 (6): 531–45. <https://doi.org/10.1111/gwao.12053>.
- Azer, Evronia. 2021. "Remote Working Has Led to Managers Spying More on Staff – Here Are Three Ways to Curb It." *The Conversation*. 2021.
<http://theconversation.com/remote-working-has-led-to-managers-spying-more-on-staff-here-are-three-ways-to-curb-it-159604>.
- Ball, Kirstie. 2010. "Workplace Surveillance: An Overview." *Labor History* 51 (1): 87–106. <https://doi.org/10.1080/00236561003654776>.
- Barocas, Solon, and Andrew D. Selbst. 2016. "Big Data's Disparate Impact." SSRN Scholarly Paper ID 2477899. Rochester, NY: Social Science Research Network. <https://doi.org/10.2139/ssrn.2477899>.
- Bernhardt, Annette, Lisa Kresge, and Reem Suleiman. 2021. "Data and Algorithms at Work: The Case for Worker Technology Rights." San Francisco: UC Berkeley Labor Center. <https://laborcenter.berkeley.edu/data-algorithms-at-work/>.
- Bogen, Miranda, and Aaron Rieke. 2018. "HELP WANTED: An Examination of Hiring Algorithms, Equity, and Bias." Washington, D.C.: Upturn. <https://www.upturn.org/reports/2018/hiring-algorithms/>.
- Broughel, James, and Adam Thierer. 2019. "Technological Innovation and Economic Growth: A Brief Report on the Evidence." Mercatus Center at George Mason University, Arlington, VA: Mercatus Research. <https://www.mercatus.org/system/files/broughel-technological-innovation-mercatus-research-v1.pdf>.

- Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 81: 1–15.
- Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (1): 2053951715622512. <https://doi.org/10.1177/2053951715622512>.
- Calo, Ryan, and Alex Rosenblat. 2017. "The Taking Economy: Uber, Information, and Power." *Columbia Law Review* 117–1690 (6): 1623–. <https://doi.org/10.2139/ssrn.2929643>.
- Camilleri, Michael. 2018. "Algorithmic Decision Making: To What Extent Can Due Process Principles Assist in Protecting Individuals from the Threat of Algocracy." Scotland: The UK: The University of Edinburgh. https://www.academia.edu/40419191/Algorithmic_decision_making_To_what_extent_can_due_process_principles_assist_in_protecting_individuals_from_the_threat_of_algocracy.
- Cañigueral, Albert. 2019. "How Can Tech Meet the Needs of Platform Workers?" *The RSA* (blog). 2019. <https://www.thersa.org/blog/2019/06/tech-platform-workers>.
- Caprettini, Bruno, and Hans-Joachim Voth. 2017. "Rage against the Machines: New Technology and Violent Unrest in Industrialising Britain." *VoxEU.Org* (blog). May 9, 2017. <https://voxeu.org/article/rage-against-machines-new-technology-and-violent-unrest>.
- Collier, Danielle, and Charlotte Zhang. 2016. "Can We Reduce Bias in the Recruiting Process and Diversify Pools of Candidates by Using Different Types of Words in Job Descriptions?," October. <https://ecommons.cornell.edu/handle/1813/74363>.
- Collins, Philippa. 2020. "The Right to Privacy, Surveillance-by-Software and the 'Home-Workplace.'" UK Human Rights Blog. September 8, 2020. <https://ukhumanrightsblog.com/2020/09/08/the-right-to-privacy-surveillance-by-software-and-the-home-workplace-philippa-collins/>.
- Datta, Amit, Michael Carl Tschantz, and Anupam Datta. 2015. "Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination." *Proceedings on Privacy Enhancing Technologies* 2015 (1): 92–112. <https://doi.org/10.1515/popets-2015-0007>.
- Daub, Adrian. 2018. "The Return of the Face." *Longreads* (blog). October 3, 2018. <https://longreads.com/2018/10/03/the-return-of-the-face/>.
- De Groen, Willem Pieter, and Ilaria Maselli. 2016. "The Impact of the Collaborative Economy on the Labour Market." Brussels: Centre for European Policy Studies. https://www.ceps.eu/system/files/SR138CollaborativeEconomy_0.pdf.
- Dencik, Lina, and Sanne Stevens. 2021. "Regimes of Justification in the Datafied Workplace: The Case of Hiring." *New Media & Society*, October, 14614448211052892. <https://doi.org/10.1177/14614448211052893>.
- Deng, Zhiwei, Mengyao Zhai, Lei Chen, Yuhao Liu, Srikanth Muralidharan, Mehrgan Javan

- Roshtkhari, and Greg Mori. 2015. "Deep Structured Models For Group Activity Recognition." In *Proceedings of the British Machine Vision Conference 2015*, 179.1-179.12. Swansea: British Machine Vision Association. <https://doi.org/10.5244/C.29.179>.
- Derousseau, Ryan. 2017. "The Tech That Tracks Your Movements at Work." 2017. <https://www.bbc.com/worklife/article/20170613-the-tech-that-tracks-your-movements-at-work>.
- Drahokoupil, Jan, and Agnieszka Piasna. 2017. "Work in the Platform Economy: Beyond Lower Transaction Costs." *Intereconomics* 52 (6): 335-40. <https://doi.org/10.1007/s10272-017-0700-9>.
- Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment." *American Economic Journal: Applied Economics* 9 (2): 1-22. <https://doi.org/10.1257/app.20160213>.
- Eurofound. 2018. "Automation, Digitisation and Platforms: Implications for Work And." Luxembourg: Publications Office of the European Union. <https://www.eurofound.europa.eu/publications/report/2018/automation-digitisation-and-platforms-implications-for-work-and-employment>.
- European Parliamentary Research Service. 2020. "Data Subjects, Digital Surveillance, AI and the Future of Work." Brussels: European Parliament. <https://data.europa.eu/doi/10.2861/879078>.
- Feng, Siyuan, Olya Kudina, Bence Mark Halpern, and Odette Scharenborg. 2021. "Quantifying Bias in Automatic Speech Recognition." *ArXiv:2103.15122 [Cs, Eess]*, April. <http://arxiv.org/abs/2103.15122>.
- Forde, Chris, Mark Stuart, Simon Joyce, Liz Oliver, Danat Valizade, Gabriella Alberti, Kate Hardy, Vera Trappmann, Charles Umney, and Calum Carson. 2017. "The Social Protection of Workers in the Platform Economy." PE 614.184. Brussels: European Parliament. [https://www.europarl.europa.eu/RegData/etudes/STUD/2017/614184/IPOL_STU\(2017\)614184_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2017/614184/IPOL_STU(2017)614184_EN.pdf).
- Freeman, Chris, and Francisco Louçã. 2002. *As Time Goes By: From the Industrial Revolutions to the Information Revolution*. Oxford: Oxford University Press. <https://doi.org/10.1093/0199251053.001.0001>.
- Fuller, Joseph B, Manjari Raman, Eva Sage-Gavin, and Kristen Hines. 2021. "Hidden Workers: Untapped Talent." Harvard Business School Project on Managing the Future of Work and Accenture. <https://www.hbs.edu/managing-the-future-of-work/Documents/research/hiddenworkers09032021.pdf>.
- Gaucher, Danielle, Justin Friesen, and Aaron C. Kay. 2011. "Evidence That Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality." *Journal of Personality and Social Psychology* 101 (1): 109-28. <https://doi.org/10.1037/a0022530>.

- Glassdoor. 2019. "How To Calculate Cost-Per-Hire." *US | Glassdoor for Employers* (blog). 2019. <https://www.glassdoor.com/employers/blog/calculate-cost-per-hire/>.
- Gómez, Manuel V. 2020. "Spanish Supreme Court Rules Food-Delivery Riders Are Employees, Not Self-Employed." *EL PAÍS English Edition*. September 24, 2020. https://english.elpais.com/economy_and_business/2020-09-24/spanish-supreme-court-rules-food-delivery-riders-are-employees.html.
- Gordon, Robert J. 2016. *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*. Princeton University Press.
- Hauben (ed.), Eftheia, Karolien Lenaerts, and Willem Waeyaert. 2020. "The Platform Economy and Precarious Work." Publication for the Committee on Employment and Social Affairs, Policy Department for Economic, Scientific and Quality of Life Policies. Luxembourg: European Parliament. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652734/IPOL_STU\(2020\)652734_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652734/IPOL_STU(2020)652734_EN.pdf).
- Heaven, Will Douglas. 2020. "Predictive Policing Algorithms Are Racist. They Need to Be Dismantled." *MIT Technology Review* (blog). 2020. <https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/>.
- Hind, Michael, Stephanie Houde, Jacquelyn Martino, Aleksandra Mojsilovic, David Piorkowski, John Richards, and Kush R. Varshney. 2019. "Experiences with Improving the Transparency of AI Models and Services." *ArXiv:1911.08293 [CS]*, November. <http://arxiv.org/abs/1911.08293>.
- Holley, Peter. 2018. "This Firm Already Microchips Employees. Could Your Ailing Relative Be Next?" *Washington Post*, 2018. <https://www.washingtonpost.com/technology/2018/08/23/this-firm-already-microchips-employees-could-your-ailing-relative-be-next/>.
- Human Resources - Northwestern University. 2019. "The Cost of a Bad Hire." 2019. <https://www.northwestern.edu/hr/about/news/february-2019/the-cost-of-a-bad-hire.html>.
- Huws, Ursula. 2020. *Reinventing the Welfare State: Digital Platforms and Public Policies*. London: Pluto Press.
- Imana, Basileal, Aleksandra Korolova, and John Heidemann. 2021. "Auditing for Discrimination in Algorithms Delivering Job Ads." *Proceedings of the Web Conference 2021*, April, 3767-78. <https://doi.org/10.1145/3442381.3450077>.
- Jin, Li, Scott Duke Kominers, and Lila Shroff. 2021. "A Labor Movement for the Platform Economy." *Harvard Business Review*, September 24, 2021. <https://hbr.org/2021/09/a-labor-movement-for-the-platform-economy>.
- Kässi, Otto, Vili Lehdonvirta, and Fabian Stephany. 2021. "How Many Online Workers Are There in the World? A Data-Driven Assessment." *Open Research Europe* 1 (lversion 4; peer

- review: 4 approved): 1–53. <https://doi.org/10.12688/openreseurope.13639.4>.
- Kellogg, Katherine C., Melissa A. Valentine, and Angèle Christin. 2020. "Algorithms at Work: The New Contested Terrain of Control." *Academy of Management Annals* 14 (1): 366–410. <https://doi.org/10.5465/annals.2018.0174>.
- Koenecke, Allison, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, and Sharad Goel. 2020. "Racial Disparities in Automated Speech Recognition." *Proceedings of the National Academy of Sciences* 117 (14): 7684–89.
- Kübler, Dorothea, Julia Schmid, and Robert Stüber. 2018. "Gender Discrimination in Hiring across Occupations: A Nationally-Representative Vignette Study." *Labour Economics* 55 (December): 215–29. <https://doi.org/10.1016/j.labeco.2018.10.002>.
- Kuek, Siou Chew, Cecilia Paradi-Guilford, Toks Fayomi, Saori Imaizumi, Panos Ipeirotis, Patricia Pina, and Manpreet Singh. 2015. "The Global Opportunity in Online Outsourcing." Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/22284>.
- Lakshmi, Soanpet .Sree, and T.Adi Lakshmi. 2014. "Recommendation Systems:Issues and Challenges." In , 5(4):5771–72.
- Lambrecht, Anja, and Catherine Tucker. 2019. "Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads." *Management Science* 65 (7): 2966–81. <https://doi.org/10.1287/mnsc.2018.3093>.
- Leckcivilize, Attakrit, and Alexander Straub. 2018. "Headscarf and Job Recruitment—Lifting the Veil of Labour Market Discrimination." *IZA Journal of Labor Economics* 7 (1): 11. <https://doi.org/10.1186/s40172-018-0071-x>.
- Levy, Karen, and Solon Barocas. 2018. "Refractive Surveillance: Monitoring Customers to Manage Workers." *International Journal of Communication* 12: 1166–88. <https://ijoc.org/index.php/ijoc/article/viewFile/7041/2302>.
- Lomas, Natasha. 2020a. "Ola Is Facing a Drivers' Legal Challenge over Data Access Rights and Algorithmic Management." *TechCrunch* (blog). 2020. <https://social.techcrunch.com/2020/09/10/ola-is-facing-a-drivers-legal-challenge-over-data-access-rights-and-algorithmic-management/>.
- . 2020b. "Uber's 'Robo-Firing' of Drivers Targeted in Latest European Lawsuit." *TechCrunch* (blog). 2020. <https://social.techcrunch.com/2020/10/27/ubers-robo-firing-of-drivers-targeted-in-latest-european-lawsuit/>.
- . 2020c. "UK Uber Drivers Are Taking the Algorithm to Court." *TechCrunch* (blog). 2020. <https://social.techcrunch.com/2020/07/20/uk-uber-drivers-are-taking-its-algorithm-to-court/>.
- . 2021a. "Dutch Court Rejects Uber Drivers' 'Robo-Firing' Charge but Tells Ola to Explain

- Algo-Deductions." *TechCrunch* (blog). 2021.
<https://social.techcrunch.com/2021/03/12/dutch-court-rejects-uber-drivers-robo-firing-charge-but-tells-ola-to-explain-algo-deductions/>.
- . 2021b. "Italy's DPA Fines Glovo-Owned Foodinho \$3M, Orders Changes to Algorithmic Management of Riders." *TechCrunch* (blog). 2021.
<https://social.techcrunch.com/2021/07/06/italys-dpa-fines-glovo-owned-foodinho-3m-orders-changes-to-algorithmic-management-of-riders/>.
- Madariaga, Javier; Buenadicha, Cesar.; Molina, Erika and Ernst, Christoph. 2019 "Economía de plataformas y empleo ¿Cómo es trabajar para una app en Argentina?" Buenos Aires: OIT.
- Malani, Shivam. 2020. "How Does Zoom Attention Tracking Works." *All Things How* (blog). April 3, 2020. <https://allthings.how/how-does-zoom-attention-tracking-works/>.
- Martins, Andrew. 2021. "Should You Consider Keystroke Logging Software?" *Business.Com* (blog). January 2021. <https://www.business.com/articles/what-is-keystroke-logging/>.
- McColl, Rod, and Marco Michelotti. 2019. "Sorry, Could You Repeat the Question? Exploring Video-Interview Recruitment Practice in HRM." *Human Resource Management Journal* 29 (4): 637–56. <https://doi.org/10.1111/1748-8583.12249>.
- McKinsey Global Institute. 2017. "Technology, Jobs, and the Future of Work." <https://www.mckinsey.com/featured-insights/employment-and-growth/technology-jobs-and-the-future-of-work>.
- Meijerink, Jeroen, and Tanya Bondarouk. 2021. "The Duality of Algorithmic Management: Toward a Research Agenda on HRM Algorithms, Autonomy and Value Creation." *Human Resource Management Review*, November, 100876.
<https://doi.org/10.1016/j.hrmr.2021.100876>.
- Milkround. 2021. "My First Virtual Job."
https://www.milkround.com/file/general/MK_CAMPAGN-My-First-Virtual-Job_Report.pdf.
- Mohamed, Marwa, Mohamed Khafagy, and Mohamed Ibrahim. 2019. "Recommender Systems Challenges and Solutions Survey." In . <https://doi.org/10.1109/ITCE.2019.8646645>.
- Murray, Iain. 2018. "The Platform Economy Can Change the World." *Competitive Enterprise Institute* (blog). 2018.
<https://cei.org/blog/the-platform-economy-can-change-the-world/>.
- OECD. 2019. "Measuring the Digital Transformation: A Roadmap for the Future." OECD.
<https://www.oecd.org/going-digital/mdt-roadmap-platform-workers.pdf>.
- Paige, Mark A., and Audrey Amrein-Beardsley. 2020. "'Houston, We Have a Lawsuit': A Cautionary Tale for the Implementation of Value-Added Models for High-Stakes Employment Decisions." *Educational Researcher* 49 (5): 350–59.
<https://doi.org/10.3102/0013189X20923046>.

- Palmer, Annie. 2021. "Amazon Uses an App Called Mentor to Track and Discipline Delivery Drivers." *CNBC* (blog). February 12, 2021. <https://www.cnbc.com/2021/02/12/amazon-mentor-app-tracks-and-disciplines-delivery-drivers.html>.
- Papscun, Dan, and Erin Mulvaney. 2021. "Uber Driver's Ratings Bias Class Action Dismissed by Judge (1)," 2021. <https://news.bloomberglaw.com/daily-labor-report/uber-drivers-ratings-race-bias-class-action-dismissed-by-judge>.
- Partnership for Finance in a Digital Africa. 2019. "Micro-Entrepreneurs in a Platform Era." Farnham, Surrey: UK: Caribou Digital Publishing. <https://mse.financedigitalafrica.org/>.
- Peachey, Kevin D. 2021. "Post Office Scandal: What the Horizon Saga Is All About." *BBC News*, July 22, 2021, sec. Business. <https://www.bbc.com/news/business-56718036>.
- Perez, Carlota. 2003. *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*. Cheltenham: Edward Elgar Pub.
- Pérez, Gorka R. 2021. "Spain Approves Landmark Law Recognizing Food-Delivery Riders as Employees." *EL PAÍS English Edition*. May 12, 2021. https://english.elpais.com/economy_and_business/2021-05-12/spain-approves-landmark-law-recognizing-food-delivery-riders-as-employees.html.
- Polido, Fabricio. 2020. "Inteligência Artificial entre Estratégias Nacionais e a Corrida Regulatória Global: Rotas Analíticas para uma Releitura Internacionalista e Comparada," *Revista da Faculdade de Direito da UFMG*, no. 76 (2020): pp. 229-256, <https://doi.org/10.12818/p.0304-2340.2020v76p229>, 233.
- Raghavan, Manish, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 469–81. FAT* '20. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372828>.
- Reynolds, Tania, Luke Zhu, Karl Aquino, and Brendan Strejcek. 2021. "Dual Pathways to Bias: Evaluators' Ideology and Ressentiment Independently Predict Racial Discrimination in Hiring Contexts." *The Journal of Applied Psychology* 106 (4): 624–41. <https://doi.org/10.1037/apl0000804>.
- Rosenblat, Alex. 2018. *Uberland: How Algorithms Are Rewriting the Rules of Work*. First edition. Oakland, California: University of California Press.
- Rowland, Christopher. 2019. "With Fitness Trackers in the Workplace, Bosses Can Monitor Your Every Step — and Possibly More." *Washington Post*, 2019. https://www.washingtonpost.com/business/economy/with-fitness-trackers-in-the-workplace-bosses-can-monitor-your-every-step--and-possibly-more/2019/02/15/75e0848-2a45-11e9-b011-d8500644dc98_story.html.
- Sanchez-Monedero, Javier, and Lina Dencik. 2019. "The Datafication of the Workplace." Cardiff:

- Cardiff University.
<https://datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-The-datatification-of-the-workplace.pdf>.
- Schmidt, Andreas T., and Bart Engelen. 2020. "The Ethics of Nudging: An Overview." *Philosophy Compass* 15 (4): e12658. <https://doi.org/10.1111/phc3.12658>.
- Schor, Juliet B., William Attwood-Charles, Mehmet Cansoy, Isak Ladegaard, and Robert Wengronowitz. 2020. "Dependence and Precarity in the Platform Economy." *Theory and Society* 49 (5): 833–61. <https://doi.org/10.1007/s11186-020-09408-y>.
- Schwab, Klaus. 2016. "The Fourth Industrial Revolution: What It Means and How to Respond." <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>.
- Sloane, Mona, Emanuel Moss, and Rumman Chowdhury. 2021. "A Silicon Valley Love Triangle: Hiring Algorithms, Pseudo-Science, and the Quest for Auditability." *ArXiv:2106.12403 [Cs]*, June. <http://arxiv.org/abs/2106.12403>.
- Tang, Shiliang, Xinyi Zhang, Jenna Cryan, Miriam J. Metzger, Haitao Zheng, and Ben Y. Zhao. 2017. "Gender Bias in the Job Market: A Longitudinal Analysis." *Proceedings of the ACM on Human-Computer Interaction* 1 (CSCW): 99:1-99:19. <https://doi.org/10.1145/3134734>.
- Thornhill, John. 2021. "Post Office Scandal Exposes the Risk of Automated Injustice." *Financial Times*, April 29, 2021. <https://www.ft.com/content/08f485bf-6cea-46d6-g62c-46263aaec5f3>.
- Turner Lee, Nicol, Paul Resnick, and Genie Barton. 2019. "Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms." Brookings. <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>.
- Urzi Brancati, Maria Cesira, Annarosa Pesole, and Enrique Fernández-Macías. 2020. *New Evidence on Platform Workers in Europe Results from the Second COLLEEM Survey*. Luxembourg: Publications Office of the European Union. <https://doi.org/10.2760/459278>.
- Vallas, Steven, and Juliet Schor. 2020. "What Do Platforms Do? Understanding the Gig Economy." *Annual Review of Sociology* 46 (July). <https://doi.org/10.1146/annurev-soc-121919-054857>.
- Vincent, James. 2018. "Facebook, Instagram, and Twitter Limit Data Access for AI-Powered Babysitter Background Checks." *The Verge* (blog). November 29, 2018. <https://www.theverge.com/2018/11/29/18117660/ai-background-checks-babysitter-predictim-data-limited-facebook-twitter-instagram>.
- . 2021. "Automated Hiring Software Is Mistakenly Rejecting Millions of Viable Job Candidates." *The Verge*. September 6, 2021. <https://www.theverge.com/2021/9/6/22659225/automated-hiring-software-rejecting-viable-candidates-harvard-business-school>.

- Vitak, Jessica, and Michael Zimmer. 2021. "Workers' Attitudes toward Increased Surveillance during and after the Covid-19 Pandemic." *Items: Insights from the Social Sciences* (blog). 2021.
<https://items.ssrc.org/covid-19-and-the-social-sciences/covid-19-fieldnotes/workers-attitudes-toward-increased-surveillance-during-and-after-the-covid-19-pandemic/>.
- Walker, Joseph. 2012. "Meet the New Boss: Big Data." *Wall Street Journal*, September 20, 2012, sec. Management.
<https://online.wsj.com/article/SB10000872396390443890304578006252019616768.html>.
- Wartzman, Rick. 2019. "Workplace Tracking Is Growing Fast. Most Workers Don't Seem Very Concerned." *Fast Company*. March 20, 2019.
<https://www.fastcompany.com/90318167/workplace-tracking-is-growing-fast-most-workers-dont-seem-very-concerned>.
- West, Darrell M. 2021. "How Employers Use Technology to Surveil Employees." *Brookings* (blog). January 5, 2021.
<https://www.brookings.edu/blog/techtank/2021/01/05/how-employers-use-technology-to-surveil-employees/>.
- Woike, Jan K., and Sebastian Hafenbrädl. 2020. "Rivals without a Cause? Relative Performance Feedback Creates Destructive Competition despite Aligned Incentives." *Journal of Behavioral Decision Making* 33 (4): 523–37. <https://doi.org/10.1002/bdm.2162>.

Guía de uso del cuestionario para la obtención de un prospecto algorítmico

El cuestionario a primera vista

Algoritmos sujetos al cuestionario

Un algoritmo puede definirse de forma general como un conjunto de instrucciones o reglas definidas y no-ambiguas, ordenadas y finitas que, a través de un estado inicial y una entrada, permite obtener un resultado final. En concreto, este prospecto ha sido diseñado para abordar:

- Aquellos algoritmos que mediante el uso intensivo de datos – ya sea previo a su aplicación o durante la misma–, permiten automatizar decisiones y procesos
- Aquellos algoritmos flexibles que sean capaces de percibir su entorno y transformar dichas observaciones en acciones que maximicen su probabilidad de éxito en alguna tarea

Sin embargo, y dada la ambigüedad y la falta de consenso en relación a la definición de "Inteligencia Artificial", cualquier algoritmo que pudiera incurrir y/o preservar la discriminación sistemática de algún grupo o clase protegida de forma accidental o intencionada, estará sujeto a la evaluación propuesta por el prospecto.

Agrupación de preguntas

Las preguntas de este cuestionario han sido agrupadas en tres bloques, con el fin de recopilar la información general sobre el modelo, los datos correspondientes con el análisis técnico, y la información sobre el impacto que el sistema en cuestión pueda tener sobre la sociedad.

Cabe señalar que, a pesar de que los bloques primero y tercero son de carácter socio-técnico (y no técnico exclusivamente), todas las preguntas deberán ser contestadas por un equipo compuesto por los ingenieros y científicos encargados de desarrollar, implementar, y/o supervisar el sistema.

CUESTIONARIO PARA EL PROSPECTO ALGORÍTMICO								
1. INFORMACIÓN GENERAL			2. ANÁLISIS TÉCNICO				3. IMPACTO	
Registro	Detalles del Modelo	Finalidad y Uso Previsto	Factores	Métricas	Datos de Evaluación	Datos de entrenamiento	Análisis Cuantitativos	Consideraciones Éticas

Figura 1. Esquema de la estructura del cuestionario

Glosario de términos

Algoritmo: Implementación computacional de un conjunto finito de instrucciones y ecuaciones matemáticas que permiten la obtención de una respuesta (output) adecuado ante distintas señales de entrada (input)³⁹

Datos de Entrenamiento: Datos utilizados por un algoritmo de machine learning para aprender a identificar patrones. Estos datos pueden ser estructurados (fácilmente comprensibles por algoritmos, como por ejemplo vectores de datos alfanuméricos) o no estructurados (como por ejemplo imágenes)

Encargado del Tratamiento de Datos: (Según GDPR: ART. 4.8.) "persona física o jurídica, autoridad pública, servicio u otro organismo que trate datos personales por cuenta del responsable del tratamiento"

Equidad: (Eng. "Fairness") Entendido como la no discriminación negativa de un individuo o colectivo por el simple hecho de pertenecer a un grupo o clase protegida, entre las cuales se incluyen la distinción por sexo (hombre/mujer) raza, origen, creencia religiosa, color de piel, origen nacional, edad, o discapacidad de cualquier naturaleza

Etiquetas o Variables Objetivo: Corresponden a los resultados que el modelo de machine learning pretende predecir. Pueden ser de tipo numérico o categórico. Si el modelo es de aprendizaje supervisado, el algoritmo aprende la relación entre los datos y las etiquetas que los categorizan

Implementador: Persona física o jurídica encargada de adaptar e integrar el algoritmo en los productos o servicios comercializados o distribuidos por el propietario, incluyendo el entrenamiento del modelo, si se diera el caso

Modelo: Producto resultante de la exposición de un algoritmo a un conjunto de datos de entrenamiento. El comportamiento del modelo – o los modos en los que el modelo genera

³⁹ En el presente documento, algoritmo se utiliza como hiperónimo de algoritmo, modelo, y sistema si no se especifica lo contrario

respuestas adecuadas (output) a distintos estímulos (input) – depende de los datos de entrenamiento

Número de Registro: Identificador único para un modelo o sistema de inteligencia artificial

Propiedades: Corresponden a las variables utilizadas por los algoritmos para predecir los resultados. Cualquier dimensión que sea medible y almacenable en una base de datos puede constituir una propiedad

Parámetros del modelo: Los parámetros del modelo son las características que aprenden los modelos de machine learning para realizar predicciones – también son conocidos como pesos o “weights”

Propietario: Persona física o jurídica que tiene derecho de propiedad sobre el algoritmo o sobre el producto del cual éste forma parte

Responsable del Tratamiento de Datos: *(Según GDPR: ART. 4.7.)* “persona física o jurídica, autoridad pública, servicio u otro organismo que, solo o junto con otros, determine los fines y medios del tratamiento; si el Derecho de la Unión o de los Estados miembros determina los fines y medios del tratamiento, el responsable del tratamiento o los criterios específicos para su nombramiento podrá establecerlos el Derecho de la Unión o de los Estados miembros”

Sistema: Conjunto compuesto por un modelo y ciertos mecanismos de relación e interacción entre el modelo y el mundo real que permiten el procesamiento de datos reales por parte del modelo

Ámbito de aplicación

A pesar de que este cuestionario ha sido desarrollado con la intención de facilitar el registro y seguimiento de sistemas algorítmicos por parte de los organismos reguladores correspondientes, también está pensado como una herramienta proactiva para incrementar la transparencia en el uso de inteligencia artificial.

Así pues, los autores recomiendan la utilización del cuestionario como una guía para hacer frente a:

- Demandas legales por una supuesta falta de transparencia en relación a la existencia de sesgos sobre poblaciones protegidas, fijación de precios, y otros ámbitos de las leyes de la competencia y derechos fundamentales

- Necesidades comerciales o estratégicas de evaluar y comunicar tanto el rendimiento como la aceptabilidad de los sistemas algorítmicos
- Voluntades técnicas (o de carácter sociotécnico) en relación a una mayor comparabilidad entre algoritmos

Cuestionario anotado

1. Registro

1.1. Número de Registro

¿Cuál es el número de registro del modelo/algoritmo/sistema?

- *Tipo de Input:* Alfanumérico
- *Descripción:* El número identificador facilitado por el Registro Algorítmico, si existe. De no estar registrado o no existir un registro algorítmico, puede especificarse como "N/A"

2. Detalles del Modelo

2.1. Implementador/Propietario

¿Quién ha implementado y es propietario del modelo/algoritmo/sistema?

- *Tipo de Input:* Alfanumérico.
- *Descripción:* Detalles fiscales completos y contacto para el Delegado de Protección de Datos del implementador y/o propietario del algoritmo

2.2. Desarrollador

(Si es distinto del propietario y/o implementador)

¿Quién ha desarrollado el modelo/algoritmo/sistema?

- *Tipo de Input:* Alfanumérico
- *Descripción:* Detalles fiscales completos y contacto para el Delegado de Protección de Datos del desarrollador del modelo

2.3. Versión del Modelo

¿Cuál es la versión del modelo/algoritmo/sistema que se está evaluando?

- *Tipo de Input:* Alfanumérico

- *Descripción:* Versión del modelo, de modo que sea posible trazar e identificar versiones anteriores que pudieran estar registradas con anterioridad

2.4. Tipo de Algoritmo

¿Cuál es la arquitectura básica del modelo?

- *Tipo de Input:* Texto
- *Descripción:* Aportar información sobre la arquitectura básica del modelo. La recomendación es detallar de general a más específico (e.g. Aprendizaje Supervisado → Clasificación → K-NN)

2.5. Código Abierto/ Software Libre

¿Se trata de un sistema de código abierto o de código cerrado?

- *Tipo de Input:* Texto / link
- *Descripción:* Especificar si se trata de un modelo / programa de acceso abierto o si, de lo contrario, se trata de código cerrado. En el caso de que sea acceso libre, adjuntar el link al repositorio o fuente original

2.6. Licencia

¿Cuál es la referencia de la licencia del modelo?

- *Tipo de Input:* Texto
- *Descripción:* Adjuntar la licencia del modelo/algoritmo/sistema en cuestión

2.7. Feedback sobre el Modelo

¿Cuál es la dirección de contacto para solicitar información o hacer sugerencias sobre el modelo/algoritmo/sistema?

- *Tipo de Input:* Alfanumérico
- *Descripción:* Aportar una dirección de correo para que los usuarios puedan solicitar información o hacer sugerencias sobre el uso del modelo/ algoritmo/sistema

3. Finalidad y Uso Previsto

3.1. Principal Uso Previsto del Algoritmo

¿Cuál es el principal uso intencionado para este modelo/sistema?

- *Tipo de Input:* Texto
- *Descripción:* Especificar el uso intencionado de diseño del modelo, y si dicho uso iba dirigido a la ejecución de tareas específicas o genéricas

3.2. Principales Usuarios Previstos del Algoritmo

¿A qué tipo de usuarios va dirigida la implementado este modelo/sistema?

- *Tipo de Input:* Texto
- *Descripción:* Especificar el público objetivo del algoritmo – tanto usuario como intermediario –, así como el modo de uso intencionado para clientes y usuarios

3.3. Área de Impacto

¿El modelo/sistema será utilizado en alguna de las siguientes áreas?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si el modelo es o será utilizado en alguna de las siguientes áreas y, de ser así, concretar cuál:
 - *Infraestructuras Críticas* (e.g. transporte), cuyo uso podría poner en riesgo la vida y salud de ciudadanos en riesgo;
 - *Educación o formación profesional*, cuyo uso podría determinar el acceso a la educación y carrera profesional de alguna persona (e.g. calificación de exámenes);
 - *Componentes de seguridad de productos* (e.g. aplicación de IA en cirugía robótica o cirugía asistida por robot);
 - *Empleo, gestión de trabajadores, y acceso al trabajo autónomo* (e.g. software para el triaje de currículums en procesos de selección);
 - *Servicios esenciales públicos y privados* (e.g. credit scoring, o denegar a ciudadanos la oportunidad de acceder a crédito);
 - *Aplicación de la Ley* que podría interferir con los derechos fundamentales de las personas (e.g. evaluación de la fiabilidad de pruebas judiciales);
 - *Migración, asilo, y control de fronteras* (e.g. verificación de la autenticidad de los documentos de identidad o de viaje);
 - *Administración de Justicia y procesos democráticos* (e.g. aplicar la ley a un conjunto de hechos concretos).

3.4. Nivel de Riesgo⁴⁰

Estimación del nivel de riesgo que conlleva el sistema (Dicha estimación será confirmada externamente)

- *Input:* Texto
- *Descripción:* Especificar en cuál de las siguientes categorías debe incluirse el algoritmo en cuestión, en términos del nivel de riesgo que conlleva para los usuarios y ciudadanos

⁴⁰ Fuente: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682

- *Riesgo Inaceptable*: Entre los cuales se incluyen aquellos sistemas de inteligencia artificial que tengan como objetivo la manipulación del comportamiento y libre albedrío de los usuarios, así como sistemas de “*social scoring*” por parte de organismos públicos o privados
- *Riesgo Alto*: Aquellos sistemas diseñados para su uso en infraestructuras críticas, educación o formación profesional, seguridad, empleo, gestión de trabajadores o acceso al trabajo, servicios esenciales tanto públicos como privados, aplicación de la Ley, migración, asilo y control de fronteras, o administración de Justicia y procesos democráticos.
- *Riesgo Limitado*: Aquellos sistemas para los cuales la revelación de su naturaleza esté requerida por la ley (por ejemplo, en el caso de chatbots)
- *Riesgo Mínimo*: Sistemas de inteligencia artificial diseñados para uso en videojuegos, filtros de correo, u otras aplicaciones que no conllevan mayor riesgo para los usuarios o ciudadanos

3.5. Uso Fuera de Ámbito

¿Hay alguna aplicación o contexto en el que el modelo no debería ser utilizado?

- *Tipo de Input*: Texto
- *Descripción*: Especificar las aplicaciones y/o contextos en los que el modelo no debería ser utilizado (e.g. datos insuficientes, reconocimiento facial en entornos poco iluminados,...)

4. Factores

4.1. Factores Relevantes

¿Cuáles son los factores previstos para los cuales el desempeño del modelo podría variar?

- *Tipo de Input*: Texto
- *Descripción*: Especificar las condiciones de uso, contextos de aplicación, y otros factores que podrían afectar al desempeño del modelo (e.g. un modelo de NLP que no puede identificar sarcasmo, un clasificador de imágenes que no funciona con ciertas poblaciones,...)

¿Cómo han sido determinados?

- *Tipo de Input*: Texto
- *Descripción*: Especificar los mecanismos utilizados para detectar dichos factores

4.2. Factores de Evaluación

¿Qué factores son monitorizados y por qué han sido seleccionados?

- *Tipo de Input:* Texto
- *Descripción:* Especificar los factores de evaluación que han sido seleccionados y monitorizados y los motivos para escoger dichos factores y no otros.

4.3. Poblaciones Afectadas

¿Cuáles son las poblaciones afectadas, directa o indirectamente, por este modelo/ sistema?

- *Tipo de Input:* Texto
- *Descripción:* Especificar los grupos de población que podían verse más afectados por el uso o implementación del modelo

5. Métricas

5.1. Métricas de Rendimiento del Modelo

¿Cómo se mide el rendimiento del modelo?

- *Tipo de Input:* Texto
- *Descripción:* Especificar los mecanismos utilizados para medir el rendimiento del modelo

¿Por qué se seleccionaron éstas métricas?

- *Tipo de Input:* Texto
- *Descripción:* Exponer la motivación para seleccionar las métricas anteriores, así como posibles alternativas y por qué no han sido seleccionadas

¿Hay algún modo de verificar tales métricas?

- *Tipo de Input:* Texto
- *Descripción:* Indicar si es posible verificar las métricas (a través de un API, por ejemplo) de forma externa

¿Además de por el proveedor, ha sido este modelo/ algoritmo/ servicio evaluado por algún tercero?

- *Tipo de Input:* Texto
- *Descripción:* Indicar si el modelo ha sido auditado o supervisado por un tercero y, de ser así, indicar cuál y con qué finalidad

5.2. Umbrales de Decisión

¿Se utilizan umbrales de decisión? De ser así, ¿cuáles y por qué fueron escogidos estos?

- *Tipo de Input:* Texto
- *Descripción:* Indicar si se han utilizado umbrales de decisión y, de ser así, cuáles se han escogido y por qué razones

5.3. Medidas de Dispersión

¿Cómo se calculan, miden y estiman la variación en las métricas de rendimiento?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si se calcula la desviación estándar, variancia, intervalos de confianza, divergencia de Kullback-Liebler,... y explicar los resultados

¿Cómo se aproximan estos valores?

- *Tipo de Input:* Texto
- *Descripción:* Especificar cómo se calculan (e.g. A través de la media de X iteraciones, mediante validación cruzada...)

6. Datos de Evaluación

6.1. Bases de Datos

¿Qué bases de datos han sido utilizados para evaluar cuantitativamente el rendimiento del modelo?

- *Tipo de Input:* Texto
- *Descripción:* Aportar un link o referencia a la base de datos utilizada

¿Se han utilizado datos sintéticos?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si se han utilizado datos sintéticos para evaluar el rendimiento del modelo

¿Permite el modelo/ algoritmo/ sistema comprobar diferencias entre los datos de entrenamiento y uso?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si hay mecanismos para comprobar diferencias en cuanto al rendimiento ante datos de entrenamiento y datos de uso

6.2. Motivación

¿Por qué se escogieron estas bases de datos?

- *Tipo de Input:* Texto
- *Descripción:* Justificar por qué se han seleccionado las bases de datos utilizadas

6.3. Pretratamiento

¿Cómo se han pretratado los datos para la evaluación?

- *Tipo de Input:* Texto
- *Descripción:* Especificar qué mecanismos de pretratamiento de datos se han implementado

¿Ha requerido el servicio algún tipo de transformación de los datos?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si se han transformado los datos y, de ser así, especificar por qué y cómo

6.4. Responsable del Tratamiento de Datos

(Según GDPR) "«responsable del tratamiento» o «responsable»: la persona física o jurídica, autoridad pública, servicio u otro organismo que, solo o junto con otros, determine los fines y medios del tratamiento; si el Derecho de la Unión o de los Estados miembros determina los fines y medios del tratamiento, el responsable del tratamiento o los criterios específicos para su nombramiento podrá establecerlos el Derecho de la Unión o de los Estados miembros" (GDPR: ART. 4.7)

- *Tipo de Input:* Texto
- *Descripción:* Especificar el responsable del tratamiento de datos

6.5. Encargado del Tratamiento de Datos

(Según la GDPR) "«encargado del tratamiento» o «encargado»: la persona física o jurídica, autoridad pública, servicio u otro organismo que trate datos personales por cuenta del responsable del tratamiento" (GDPR: ART. 4.8)

- *Tipo de Input:* Texto
- *Descripción:* Especificar el encargado del tratamiento de datos

7. Datos de Entrenamiento

(Esta sección reformula las preguntas de la sección anterior, aplicadas al contexto de entrenamiento. El modelo/algoritmo/sistema ha sido entrenado externamente, es posible que sea imposible rellenar esta sección)

7.1. Bases de Datos

¿Qué bases de datos han sido utilizados para entrenar el modelo?

- *Tipo de Input:* Texto
- *Descripción:* Aportar un link o referencia a la base de datos utilizada

¿Se han utilizado datos sintéticos?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si se han utilizado datos sintéticos para entrenar el modelo

7.2. Motivación

¿Por qué se escogieron estas bases de datos?

- *Tipo de Input:* Texto
- *Descripción:* Justificar por qué se han seleccionado las bases de datos utilizadas

7.3. Pretratamiento

¿Cómo se han pretratado los datos para el entrenamiento?

- *Tipo de Input:* Texto
- *Descripción:* Especificar qué mecanismos de pretratamiento de datos se han implementado

¿Ha requerido el entrenamiento algún tipo de transformación de los datos?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si se han transformado los datos y, de ser así, especificar por qué y cómo

7.4. Responsable del Tratamiento de Datos

(Según la GDPR) «responsable del tratamiento» o «responsable»: la persona física o jurídica, autoridad pública, servicio u otro organismo que, solo o junto con otros, determine los fines y medios del tratamiento; si el Derecho de la Unión o de los Estados miembros determina los fines y medios del tratamiento, el responsable del tratamiento o los criterios específicos para su nombramiento podrá establecerlos el Derecho de la Unión o de los Estados miembros” (GDPR: ART. 4.7)

- *Tipo de Input:* Texto
- *Descripción:* Especificar el responsable del tratamiento de datos

8. Análisis Cuantitativos

Ver sección “recursos” para scripts/APIs para ejecutar el análisis. Si no hay opensource disponible, desarrollar internamente

8.1. Métricas de Equidad (Fairness)

¿Qué métricas se han definido en cuanto a factores de equidad?

- *Tipo de Input:* Texto / Gráfico
- *Descripción:* Especificar si se han identificado métricas para monitorizar posibles discriminaciones hacia grupos protegidos (e.g. edad, género, color de piel, localización,...)

8.2. Resultados Unitarios

¿Cuál es el rendimiento del modelo respecto a cada uno de estos factores?

- *Tipo de Input:* Texto / Gráfico
- *Descripción:* Especificar el rendimiento del modelo respecto a cada uno de estos factores

Adjuntar la variación de métricas

- *Tipo de Input:* Texto / Gráfico
- *Descripción:* Especificar la variación de métricas (e.g. mediante barras de error) en relación a cada uno de estos factores

8.3. Resultados interseccionales

¿Cuál es el rendimiento del modelo respecto a la intersección de los factores evaluados?

- *Tipo de Input:* Texto / Gráfico
- *Descripción:* Especificar el rendimiento del modelo respecto a la intersección de los diversos factores de riesgo

Adjuntar la variación de métricas

- *Tipo de Input:* Texto
- *Descripción:* Especificar la variación de métricas (e.g. mediante barras de error) en relación a la intersección de estos factores

8.4.1. (si el modelo es un clasificador) Sensibilidad, Especificidad, y Matriz de Confusión

Para cada una de las categorías relevantes, especificar si los valores se han obtenido en el laboratorio o en un entorno operacional, y adjuntar intervalos de confianza

- *Tipo de Input:* Texto / Gráfico

8.4.2. (si el modelo es score-based) Distribución de las Métricas Medidas Entre Grupos

Medidas de tendencia central (moda, media y mediana), y dispersión o variación (rango, cuartiles, desviación absoluta, varianza y desviación estándar)

- *Tipo de Input:* Texto / Gráfico

9. Consideraciones Éticas

9.1. Datos

¿Usa el modelo datos sensibles sobre clases protegidas?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si los datos están aglomerados o pretratados, de modo que las clases protegidas sean atributos (metadata)

9.2. Impacto Severo

¿Está diseñado el modelo para informar o tomar decisiones sobre cuestiones centrales de la vida humana?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si el campo de decisión y acción del sistema afecta dimensiones como la salud, la seguridad, el acceso al trabajo,...

9.3. Sesgo y Problemas Éticos

¿Tienen constancia de posibles problemas de sesgo o seguridad que pudieran derivar del uso del modelo/ algoritmo/ sistema?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si se han detectado posibles situaciones o contextos en los que el uso del modelo pudiera desencadenar en problemas graves de seguridad o sesgo para ciertos colectivos o para la población en general

9.4. Riesgos

¿Qué riesgos podrían derivar de tal sesgo, y cuál sería la población potencialmente afectada?

- *Tipo de Input:* Texto
- *Descripción:* Especificar la estimación de los riesgos que podrían derivar de los posibles sesgos del modelo

9.4. Mitigación

¿Qué medidas se han tomado o se tomarían para mitigar tal sesgo?

- *Tipo de Input:* Texto
- *Descripción:* Especificar los protocolos que se han establecido para mitigar la existencia y los efectos de posibles sesgos en el modelo

9.5. Human-in-the-loop

¿Cuál ha sido el rol específico de los humanos en el ciclo de vida del modelo, en cuanto a input y supervisión?

- *Tipo de Input:* Texto
- *Descripción:* Especificar las dinámicas de interacción entre el modelo y los humanos, en caso de existir, así como el ámbito en el que han ocurrido (modelaje, simulación,...) y la finalidad para la cual se ha introducido un actor humano (mitigación de sesgo, aprendizaje cooperativo,...)

9.6. Inclusión de Interesados en el Testeo y Pilotaje

¿Cuáles son las principales partes interesadas?

- *Tipo de Input:* Texto
- *Descripción:* Especificar las partes interesadas en el uso e implementación del modelo (*stakeholders*)

¿Han sido dichas partes incluidas en el proceso de prueba y/o pilotaje del modelo/algoritmo/sistema, con anterioridad al lanzamiento del sistema?

- *Tipo de Input:* Texto
- *Descripción:* Especificar si se han incluido dichas partes en el diseño, desarrollo, e implementación del sistema, así como los mecanismos utilizados y los resultados obtenidos de dicha participación

^[1] Fuente: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682

CIPPEC[®]
eticas

