Los sistemas automatizados de reconocimiento de emociones en el trabajo en el reglamento europeo de inteligencia artificial

Automated emotion recognition systems at work in the European Artificial Intelligence Act

Ana Belén Muñoz Ruiz

Profesora Titular de Derecho del Trabajo y de la Seguridad Social (catedrática acreditada)
Universidad Carlos III de Madrid

ORCID ID: 0000-0002-8863-9938

doi: 10.20318/labos.2024.9033

Resumen: Desde los comienzos de los años 90 se viene investigando en los sistemas de reconocimiento automatizados de las emociones. Los sistemas automatizados de reconocimiento de emociones basados en los datos biométricos (rostro, voz, movimiento corporal, entre otros) pueden ser una fuente potencial de información para las empresas sobre sus trabajadores. En el artículo se estudia la nueva regulación de estos sistemas en el Reglamento europeo de Inteligencia Artificial con un doble propósito: primero, identificar el alcance y contenido de las obligaciones y prohibiciones de las empresas que1 tratan esta tipología de datos de carácter personal con apoyo en la inteligencia artificial; y segundo, comprender que los sistemas de reconocimiento de emociones en el ámbito laboral suponen un salto cualitativo si los comparamos con los controles tradicionales (videovigilancia, email, Internet, ordenador y dispositivos semejantes). Como se verá, este tipo de control de nueva generación sitúa a las personas trabajadoras en una vulnerabilidad extrema y precisan de mayores garantías.

Palabras clave: Reglamento europeo, biometría, dato de carácter personal, sistema automatizado de reconocimiento de emociones, inteligencia artificial, derecho del trabajo, obligaciones empresariales, intimidad.

Abstract: Since the early 1990s, research has been carried out on automated emotion recognition systems. These automated emotion recognition systems, which are based on biometric data (face, voice, body movement, etc.), may be a potential source of worker information for companies. The paper analyses the new European regulation which includes those mentioned systems on twofold purpose. First, to identify the scope and content of the obligations and prohibitions of companies that process this type of personal data with the support of artificial intelligence; and second, to explain that emotion recognition systems in the labor setting are a qualitative leap as compared to traditional control measures (video surveillance, email, Internet, computer and similar devices). It is shown that these new generation control measures may place employees in situations of extreme vulnerability and therefore require additional guarantees.

Keywords: European Artificial Intelligence Act, biometrics, personal data protection, automated emotion recognition systems, artificial intelligence, labour law, duties of the employer, privacy.

1. Introducción

Desde los comienzos de los años 90 se viene investigando en los sistemas de reconocimiento automático de las emociones. Programadores y matemáticos sugieren tres modos de expresión emocional adecuados para la detección automatizada: (i) la emoción a partir de la expresión facial; (ii) la detección de la emoción a partir del habla y; (iii) la detección de la emoción multimodal, es decir, la combinación de la emoción facial y del habla. De los tres modos de expresión emocional, la expresión facial es una de las formas más poderosas que tienen las personas para entablar conversaciones y comunicar emociones y otras señales mentales, sociales y fisiológicas. Una de las formas más importantes en que las personas muestran sus emociones es a través de las expresiones faciales[1].

Cuando nos referimos a emociones comúnmente nos referimos a emociones concretas y a los estados de ánimo (por ejemplo, estados depresivos o ansiosos). ¿Y los sentimientos? Los sentimientos se definen como el poso que dejan las emociones[2].

Querer comprender el comportamiento humano ignorando las emociones es como querer comprender el funcionamiento de un coche ignorando su motor. Los sistemas de reconocimiento de emociones pueden ser una fuente potencial de información para las empresas sobre las personas trabajadoras[3]. En esta lógica, resulta igual de importante comprender las propias emociones como reconocer las emociones de los demás. Más aún cuando, con frecuencia, las motivaciones humanas residen en el estado emocional de los agentes[4].

Precisamente el nuevo Reglamento (UE) 2024/1689, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (en adelante, RIA), introduce el concepto de reconocimiento automatizado de emociones para establecer unos mecanismos de tutela en el ámbito laboral. Si bien se establece una entrada en vigor aplazada y distinta en función del articulado de esta norma[5], la novedad de su regulación es motivo suficiente para realizar un análisis pormenorizado del concepto legal de sistema automatizado de emociones, sus implicaciones para los trabajadores y los niveles de riesgo que introduce.

2. El concepto legal de sistema automatizado de reconocimiento de emociones y los motivos para su inclusión

Como se ha anticipado, el RIA da un paso adelante e introduce por primera vez en una norma jurídica de alcance comunitario el concepto de sistema automatizado de recono-

[1] K. Prasanthi Jasmine y K. Naga Prakash, *Reconocimiento de emociones humanas a partir de imágenes de rostros*, Ediciones Nuestro Conocimiento, 2021, pp. 5-6.

[2] V. Camps, *El gobierno de las emociones*, Herder, Barcelona, 2011, p. 40.

[3] Sobre este tema, me remito a mi libro A.B., Muñoz Ruiz, *Biometría y sistemas automatizados de reconocimiento de emociones: implicaciones jurídico-laborales*, Tirant Lo Blanch, Valencia, 2023.

[4] D. Pinea Oliva, *Sobre las emociones*, Ediciones Cátedra, 2019, p. 12.

[5] Con carácter general la entrada en vigor está prevista para el 2 de agosto de 2026. Vid. artículo 113 del RIA.

cimiento de emociones. Se define el sistema de reconocimiento de emociones como un sistema de inteligencia artificial (en adelante, IA) destinado a distinguir o inferir las emociones o las intenciones de las personas físicas a partir de sus datos biométricos (artículo 3 (39) RIA)[6]. Sustituir el texto tachado por este otro: En la condición de responsables del despliegue se enumeran en el RIA las obligaciones de las empresas y entidades públicas que usen estos sistemas de alto riesgo y que son las siguientes: a) Deber de transparencia y explicación individual. Transparencia informando al trabajador afectado y a la representación de los trabajadores de que están expuestos a este tipo de sistema. Dicha información debe proporcionarse con anterioridad a la puesta en servicio o utilización del sistema de IA en el lugar de trabajo (artículo 26.7 RIA). Por su parte, la explicación individual consiste en el derecho del trabajador (y obligación de la empresa) a recibir explicaciones claras y significativas acerca del papel que el sistema de IA ha tenido en el proceso de toma de decisiones y los principales elementos de la decisión adoptada cuando produzca efectos jurídicos o le afecte considerablemente del mismo modo, de manera que considere que tiene un efecto perjudicial para su salud, su seguridad o sus derechos fundamentales (artículo 86 RIA). b) Cuando proceda, los responsables del despliegue de sistemas de IA de alto riesgo utilizarán la información facilitada conforme al artículo 13 del RIA para cumplir la obligación de llevar a cabo una evaluación de impacto relativa a la protección de datos que les imponen el artículo 35 del RGPD o el artículo 27 de la Directiva (UE) 2016/680, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos. c) Supervisión humana. Los sistemas de IA de alto riesgo precisan de supervisión humana y los responsables del despliegue deben encomendar dicha supervisión a las personas físicas que tengan la competencia, la formación y la autoridad necesarias (artículo 26.2 RIA). d) Deber de vigilar el correcto funcionamiento de los sistemas sobre la base de las instrucciones de uso y, cuando proceda, informarán al proveedor o distribuidor y a la autoridad competente de acuerdo con el artículo 72 del Reglamento,

[6] En la tramitación del Reglamento comunitario se formularon algunas propuestas de cambio respecto del concepto originario que, finalmente, no prosperaron. La versión originaria definía el sistema automatizado de emociones del siguiente modo: "el sistema de reconocimiento de emociones es un sistema de inteligencia artificial (en adelante, IA) destinado a detectar o deducir las emociones o las intenciones de personas físicas a partir de sus datos biométricos". El 25.11.2022 el Consejo de la Unión Europea adoptó su posición (también denominada "orientación general") sobre la Ley de Inteligencia Artificial. En dicho documento se realizan algunas modificaciones en la definición de sistema de reconocimiento de emociones. Se propuso la siguiente definición: "un sistema de IA destinado a detectar o deducir los estados mentales, las emociones o las intenciones de las personas físicas a partir de sus datos biométricos". Como se observa, se incluyen los estados mentales para así dar cobertura a los estados de ánimo que podrían no considerarse emociones, por ejemplo, estar confuso, estar despistado, falta de concentración. Además de estados mentales, se podría añadir estados físicos (por ejemplo, estar cansado, tener una cojera, trabajador con una lesión). Con fecha de 14 de junio de 2023 el Parlamento Europeo incorporó en la definición los pensamientos: "el sistema de reconocimiento de emociones es un sistema de IA destinado a detectar o deducir las emociones, los pensamientos, los estados de ánimo o las intenciones de individuos o grupos a partir de sus datos biométricos y sus datos de base biométrica".

relativo al sistema de vigilancia poscomercialización. e) Los responsables del despliegue que sean autoridades públicas o instituciones, órganos y organismos de la Unión deben cumplir las obligaciones de registro previstas en el artículo 49 del Reglamento.

La definición de dato biométrico a que se refiere el RIA coincide con la establecida en el artículo 4 (14) del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (RGPD), según el cual los datos biométricos pertenecen a la categoría de datos especiales que se obtienen partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos. Otros ejemplos de datos biométricos son el reconocimiento de iris o retina, firma, voz, etc.[7]

En el RIA el concepto de emociones se refiere a emociones o intenciones como la felicidad, la tristeza, la indignación, la sorpresa, el asco, el apuro, el entusiasmo, la vergüenza, el desprecio, la satisfacción y la diversión. Pero no incluye los estados físicos, como el dolor o el cansancio, como, por ejemplo, los sistemas utilizados para detectar el cansancio de los pilotos o conductores profesionales con el fin de evitar accidentes. Tampoco incluye la mera detección de expresiones, gestos o movimientos que resulten obvios, salvo que se utilicen para distinguir o deducir emociones. Esas expresiones pueden ser expresiones faciales básicas, como un ceño fruncido o una sonrisa; gestos como el movimiento de las manos, los brazos o la cabeza, o características de la voz de una persona, como una voz alzada o un susurro (Considerando 18 del RIA).

En definitiva, la emoción de una persona expone su vulnerabilidad esencial. Se dice que un ser sin emociones porque se ha liberado de éstas no es un ser humano[8]. De hecho, los ordenadores emocionales que simulan intencionalidad, emociones, valores y sentido común son eso, únicamente simulaciones, no realidades. Hacen como si sintieran, pero para sentir se necesita un cuerpo[9].

¿Por qué abordar los sistemas de reconocimiento de emociones en el RIA? La razón se explica en el preámbulo de la norma cuando se dice que los datos biométricos pueden permitir las funciones tradicionales (autenticación, la identificación o la categorización de las personas físicas) pero también el reconocimiento de las emociones de las personas físicas (Considerando 14 RIA). Y esto es preocupante porque el tratamiento de datos biométricos por parte de las empresas puede derivar en el conocimiento de enfermedades de la persona trabajadora o predisposición a padecerlas sin cumplir la finalidad de prevención de riesgos laborales. A su vez, pueden producirse fallos del sistema de la IA debido a las singularidades culturales y derivar en discriminaciones[10].

[7] Vid. Artículo 3. 34) RIA.

[8] V. Camps, *El gobierno de las emociones*, Herder, Barcelona, 2011, p. 38.

[9] A. Cortina, Orts, "Ética de la inteligencia artificial", *Anales de la Real Academia de Ciencias Morales y Políticas*, 2019, nº 96, p. 385.

[10] AB, Muñoz Ruiz, *Biometría y los sistemas automatizados de reconocimiento de emociones: implicaciones jurídico-laborales*, Tirant Lo Blanch, Valencia, 2023.

A diferencia de los controles tradicionales (videovigilancia, ordenador, Internet, entre otros), los sistemas de reconocimiento de emociones emplean algoritmos e IA que, como se verá, incrementará la capacidad de análisis y explotación del resultado alcanzado. Lo que va a significar una mayor intromisión en los derechos de intimidad y protección de datos de carácter personal de las personas trabajadoras y producir lesiones indirectas de otros derechos fundamentales (discriminación, daños a la salud mental y su conexión con la seguridad y salud en el trabajo).

Los datos biométricos (voz, rostro, movimiento corporal, entre otros) han experimentado una profunda transformación. Por lo que se refiere a la voz, los científicos han conseguido con varias técnicas de procesado capturar esta capa de información, oculta a primera vista, amplificando y extrayendo características tonales y acústicas del habla humana. Las emociones que este sistema puede detectar pueden ser calma, felicidad, tristeza, ira, temor, asco, sorpresa… o simplemente "neutral"[11]. En este sentido, se describen supuestos donde las máquinas juzgan a los humanos y, por ejemplo, pueden tener como resultado que una persona no consiga un trabajo por el tono de voz[12].

La fiabilidad de este sistema se ve comprometida por la presencia de "Ruido" en la grabación, a pesar de ser un procedimiento computacionalmente más sencillo que el reconocimiento del habla ("Ok Google", Alexa, Siri….)[13]. De todas formas, los valores de fiabilidad esperados en los resultados superan el 80% y pueden llegar a alcanzar el 95%, dependiendo de la calidad de la fuente de grabación y de su procesamiento.

El análisis de la voz por sí mismo no es capaz de analizar ningún aspecto de la salud ("física") de la persona propietaria de la voz, incluso la salud "psíquica" es de muy difícil interpretación usando sólo la voz, pero eso no quiere decir que no pueda ser de ayuda en la salud de las personas. Por ejemplo, en Taiwán se realizó un ensayo en el cual se utiliza este sistema para el análisis de las reacciones de los pacientes durante las consultas. Esto sirve para "entrenar" a los doctores a tener una mayor empatía. El objetivo es conseguir este análisis y un asesoramiento al facultativo a tiempo real[14].

Existe consenso científico en relación a que al afirmar que es el rostro donde están ubicados muchos de los rasgos en que los humanos nos fijamos para reconocer a otro, juega además un papel clave en la comunicación e interacción con los demás, en la transmisión de la identidad y de la emoción[15]. Precisamente, el reconocimiento facial de emociones (FER/Facial Emotion Recognition) es la tecnología que analiza las expresio-

[11] https://www.projectpro.io/article/speech-emotion-recognition-project-using-machine-learning/573

[12] Vid. F. Pasquale, *Las nuevas leyes de la robótica. Defender la experiencia humana en la era de la IA*, Galaxia Gutenberg, S.L. 2024, p. 173.

[13] Instituto Nacional de Ciberseguridad (INCIBE), Tecnologías biométricas aplicadas a la ciberseguridad. Una guía de aproximación para el empresario, 2016, p. 12; accesible en: https://www.incibe.es/sites/default/files/contenidos/guias/doc/guia_tecnologias_biometricas_aplicadas_ciberseguridad_metad.pdf.

[14] https://www.mdpi.com/2076-3417/11/11/4782

[15] L, Escajedo San Epifanio, *Reconocimiento e identificación de las personas mediante Biometrías estáticas y dinámicas*, Tesis Doctoral, Universidad de Alicante, 2015, p. 86.

nes faciales, tanto de imágenes como de vídeos, con el objetivo de obtener información acerca del estado emocional del sujeto[16].

Debido a que la cara es uno de nuestros medios más importantes de comunicación no verbal, la cantidad de emociones detectables por este medio es alta: ira, asco, miedo, alegría, sorpresa… Los modelos pueden incluso detectar "emociones compuestas", por ejemplo "tristemente sorprendido", "sorprendentemente enfadado"….

Se sabe que la comunicación verbal supone solamente un 7% de la comunicación humana, mientras que la expresión facial supone entre un 38% y un 55%[17].

En cuanto al entorno laboral, esta técnica tendrá uso en dos vertientes: aspectos puramente laborales y detección de enfermedades de la persona trabajadora. En cuanto a los aspectos puramente laborales, esta tecnología ayudará a los reclutadores después de las entrevistas de trabajo a tomar decisiones (y en un futuro durante las entrevistas), detectará el interés del candidato durante la entrevista de trabajo y durante el trabajo, se podrá monitorizar las actitudes, atención y motivación de los empleados.

Los datos de salud que se pueden detectar con esta técnica son limitados, lógicamente se limitan a enfermedades mentales: detección de autismo, enfermedades degenerativas, trastornos psicóticos, tendencias suicidas, depresión…. También está en estudio la posible detección de enfermedades genéticas que tengan un reflejo en la cara de la persona.

Por lo que se refiere al estudio de emociones a través de los movimientos del trabajador, en el ámbito laboral es posible la detección de estados de ansiedad y de estrés[18]. El gran problema de desarrollo de esta técnica es la identificación de las características relacionadas con las emociones en los movimientos del cuerpo humano, es decir, relacionar ciertos movimientos o conjuntos de movimientos con emociones, o un conjunto de emociones. Las emociones que se pueden estudiar aquí son, de momento, felicidad, tristeza, miedo, ira y "neutral". La gran ventaja de este sistema es que el sujeto objeto de estudio puede ni siquiera saber que lo está siendo, ya que la imagen puede capturarse a gran distancia (no como el análisis de la voz, o de la expresión facial).

La precisión de estos sistemas llega a valores del 90%, incluso del 96% si la persona está sentada (y el algoritmo está preparado para ello). Si la persona puede estar haciendo diferentes acciones, la precisión del algoritmo puede bajar, teniendo un valor aproximado del 85%. De todas formas, el sistema tiene dificultades que todavía no han sido corregidas, por ejemplo, si la persona está andando[19].

---

[16] Vid. sobre el tema: https://edps.europa.eu/system/files/2021-05/21-05-26_techdispatch-facial-emotion-recognition_ref_en.pdf

[17] C. Blushan y otros, Facial Expression Recognition, https://www.wsj.com/articles/BL-DGB-42522

[18] https://www.researchgate.net/publication/338238356_Emotion_Recognition_From_Body_Movement

[19] Como ejemplo, se ha publicado noticia de un proyecto piloto que se desarrollará en las cárceles de Cataluña. La prueba piloto consiste en analizar a través de las imágenes registradas por las cámaras de vigilancia interna y de la IA las expresiones faciales y el lenguaje corporal de los reclusos. El objetivo es prevenir riesgos que puedan producirse, como una fuga o la introducción de droga en la prisión. La implantación de este sistema ha sido adjudicada a una empresa, que desarrolla proyectos detección biométrica. La empresa adjudicataria creará e instalará un sistema automatizado de identificación facial y control de movimientos de internos en zonas críticas del perímetro de seguridad del centro. El sistema servirá para realizar búsqueda de datos y su posterior

3. Los niveles de riesgo en el nuevo Reglamento europeo de Inteligencia Artificial

En el articulado del RIA se palpa el carácter inspirador de la propuesta legislativa alemana. La Comisión Ética de Datos constituida por el Gobierno alemán recomendó en 2019 adoptar un enfoque normativo basado en el riesgo distinguiendo cinco niveles de criticidad en función de las variables de probabilidad y severidad del daño como consecuencia del empleo de algoritmos. A diferencia de algunos informes que ponen el foco en las regulaciones nacionales, el documento de trabajo mencionado afirma que se precisa una nueva regulación europea sobre sistemas algorítmicos fijando unos requisitos generales horizontales que deberían ser desarrollados por normas sectoriales (entre ellas, el derecho del trabajo) [20].

Siguiendo este enfoque de regulación basado en el riesgo, el RIA menciona de forma expresa que el límite de los sistemas de IA son los derechos fundamentales: dignidad, intimidad, protección de datos de carácter personal, no discriminación y seguridad y salud de los ciudadanos y también de las personas trabajadoras.

En efecto, el RIA, a diferencia de otras normas previas con una lógica semejante como el RGPD, donde las referencias a la salud se concentran en el tratamiento de los datos de salud al ser considerado como datos especialmente protegidos, establece como uno de sus objetivos garantizar un nivel elevado de protección de la salud y seguridad: "Artículo 1º: El objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior y promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA (en lo sucesivo, «sistemas de IA») en la Unión así como prestar apoyo a la innovación".

Se afirma en el Considerando (47) del RIA que "los sistemas de IA pueden tener un efecto adverso para la salud y la seguridad de las personas, en particular cuando funcionan como componentes de seguridad de productos. (…) Por ejemplo, los robots cada vez más autónomos que se utilizan en las fábricas o con fines de asistencia y atención personal deben poder funcionar y desempeñar sus funciones de manera segura en entornos complejos. Del mismo modo, en el sector sanitario, donde puede haber repercusiones especialmente importantes en la vida y la salud, los sistemas de diagnóstico y de apoyo a las decisiones humanas, cuya sofisticación es cada vez mayor, deben ser fiables y precisos".

---

análisis para clasificar los perfiles que presenten riesgo de violencia en el interior de los recintos penitenciarios. A través de esta tecnología también se podrá evaluar si algún interno introduce droga u objetos prohibidos en la cárcel. El reconocimiento gestual también permitirá analizar expresiones, actitudes o comportamientos de los presos, incluso tras una comunicación íntima o con la familia (una vis a vis). En la actualidad, este control dependía, en gran parte, del conocimiento que tienen los funcionarios. Vid. noticia: https://www.elperiodico.com/es/sociedad/20230920/carceles-cataluna-inteligencia-artificial-control-presos-92321451

[20] Más extensamente sobre la propuesta alemana me remito a mi trabajo A.B., Muñoz Ruiz, ¿Se deben regular los algoritmos? Un breve análisis a la propuesta normativa alemana: la pirámide de criticidad basada en el riesgo, *Blog El Foro de Labos*, 11.12.2019.

Algunos estudios advierten que la tecnología de control emocional puede tener implicaciones para la salud y seguridad de las personas trabajadoras. Esta tecnología permite conocer los movimientos corporales, los signos vitales, los indicadores de estrés y fatiga, las micro expresiones faciales, el tono de voz y análisis de sentimiento. De hecho, pueden dar lugar a que los trabajadores pierdan el control de sus puestos de trabajo y aumenten la micro gestión, la presión por el rendimiento, la competitividad, la individualización y el aislamiento social. Al sentir que los trabajadores que su privacidad está siendo invadida, puede generar ansiedad y estrés. Es probable que los propios trabajadores declinen tomar descansos cuando los necesitan, lo que puede causar accidentes y problemas de salud, como trastornos musculoesqueléticos y enfermedades cardiovasculares. Los horarios de trabajo inestables, como los horarios a corto plazo establecidos automáticamente por algoritmos, pueden tener impacto negativo en los trabajadores, incluido un mayor conflicto entre el trabajo y la familia y el estrés laboral y la incertidumbre de ingresos [21].

El RIA supone un salto cualitativo respecto del RGPD por varias razones. En primer lugar, se da el paso de aprobar un marco jurídico de la IA que supone una mayor envergadura que el procesamiento de datos de carácter personal. En segundo lugar, se concede un especial protagonismo a los datos biométricos (rostro, voz, huella dactilar, movimientos corporales, entre otros) y sus nuevas formas de captación (remota, en tiempo real, en tiempo diferido, entre otros). Y tercera, se supera el plano de los datos de carácter personal y se incluyen las emociones de las personas trabajadoras.

Con la aprobación del RIA se incrementa la protección de los datos de salud de las personas trabajadoras. Es cierto que no se modifican de forma explícita las reglas y garantías de tratamiento de la salud de las personas empleadas previstas en el RGPD y en la Ley 31/1995, de 8 noviembre, de Prevención de Riesgos Laborales (LPRL). Sin embargo, se da un paso más en el RIA al incluir protección frente a los sistemas automatizados de reconocimiento de emociones de las personas.

A partir de estas premisas, el RIA establece cuatro niveles de riesgo donde podemos encontrar alguna expresión de los sistemas automatizados de reconocimiento de emociones de las personas empleadas.

En primer lugar, los sistemas de reconocimiento de emociones aparecen mencionados de forma explícita en el nivel de riesgo 1 (riesgo inaceptable), luego, están prohibidos. Sin embargo, se recogen dos excepciones. Señala el artículo 5.1 f) RIA: "Quedan prohibidas las siguientes prácticas de IA: (…) La introducción en el mercado, la puesta en servicio para este fin específico o el uso de sistemas de IA para inferir las emociones de una persona física en los lugares de trabajo y en los centros educativos, excepto cuando el sistema de IA esté destinado a ser instalado o introducido en el mercado por motivos médicos o de seguridad".

Se siguen de esta forma las directrices aportadas por el Comité Europeo de Protección de Datos (CEPD) y el Supervisor Europeo de Protección de Datos (SEPD) que

[21] OSHA-EU, Impact of artificial intelligence on occupational safety and health, 2021.

habían considerado que el uso de la IA para inferir emociones de una persona física es muy indeseable y deberá prohibirse, excepto en determinados casos de uso bien especificados, a saber, con fines de salud o investigación (por ejemplo, pacientes para quienes el reconocimiento emocional es importante), siempre con las salvaguardias adecuadas y, por supuesto, con sujeción a todas las demás condiciones y límites de protección de datos, incluida la limitación de la finalidad[22].

En segundo término, se debe añadir que los sistemas de reconocimiento de emociones también podrían estar en el nivel de riesgo 2: sistemas de IA de alto riesgo cuando no estén prohibidos, por ejemplo, aquéllos que estén justificados por motivos médicos o de seguridad. De hecho, en el Anexo III del RIA donde se enumeran los sistemas de IA de alto riesgo aparecen los sistemas de IA destinados a ser utilizados para el reconocimiento de emociones y se dice en el Considerando (54) del RIA que: "Además, deben clasificarse como de alto riesgo los sistemas de IA destinados a ser utilizados para la categorización biométrica conforme a atributos o características sensibles protegidos en virtud del artículo 9, apartado 1, del Reglamento (UE) 2016/679 sobre la base de datos biométricos, en la medida en que no estén prohibidos en virtud del presente Reglamento, así como los sistemas de reconocimiento de emociones que no estén prohibidos con arreglo al presente Reglamento".

En la condición de responsables del despliegue se enumeran en el RIA las obligaciones de las empresas y entidades públicas que usen estos sistemas de alto riesgo y que son las siguientes: a) Deber de transparencia y explicación individual. Transparencia informando al trabajador afectado y a la representación de los trabajadores de que están expuestos a este tipo de sistema. Dicha información debe proporcionarse con anterioridad a la puesta en servicio o utilización del sistema de IA en el lugar de trabajo (artículo 26.7 RIA). Por su parte, la explicación individual consiste en el derecho del trabajador (y obligación de la empresa) a recibir explicaciones claras y significativas acerca del papel que el sistema de IA ha tenido en el proceso de toma de decisiones y los principales elementos de la decisión adoptada cuando produzca efectos jurídicos o le afecte considerablemente del mismo modo, de manera que considere que tiene un efecto perjudicial para su salud, su seguridad o sus derechos fundamentales (artículo 86 RIA). b) Cuando proceda, los responsables del despliegue de sistemas de IA de alto riesgo utilizarán la información facilitada conforme al artículo 13 del RIA para cumplir la obligación de llevar a cabo una evaluación de impacto relativa a la protección de datos que les imponen el artículo 35 del RGPD o el artículo 27 de la Directiva (UE) 2016/680, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos. c) Supervisión humana. Los sistemas de IA de alto riesgo precisan de supervisión humana y los responsables del despliegue deben encomendar dicha supervisión a las personas físicas que tengan la competencia, la formación y la autoridad nece-

[22] CEPD-SEPD Dictamen conjunto 5/2021 sobre Ley de Inteligencia Artificial, 18.6.2021.

sarias (artículo 26.2 RIA). d) Deber de vigilar el correcto funcionamiento de los sistemas sobre la base de las instrucciones de uso y, cuando proceda, informarán al proveedor o distribuidor y a la autoridad competente de acuerdo con el artículo 72 del Reglamento, relativo al sistema de vigilancia poscomercialización. e) Los responsables del despliegue que sean autoridades públicas o instituciones, órganos y organismos de la Unión deben cumplir las obligaciones de registro previstas en el artículo 49 del Reglamento.

Si bien pensamos que los ejemplos que podrían encajar en estas excepciones son limitados (en la medida que el RIA ha excluido los sistemas utilizados para detectar el cansancio de los pilotos o conductores profesionales de la definición de sistema de reconocimiento de emociones, podemos mencionar alguno. Sería el caso antes mencionado de las cámaras de IA que algunas cárceles españolas han implantado para identificar expresiones faciales y lenguaje corporal de los reclusos[23].

En tercer lugar, también podríamos entender que habrá sistemas de este tipo que encajen en el nivel de riesgo 3 (nivel de riesgo limitado) cuando no se basen en los datos biométricos de las personas empleadas. El artículo 50 del RIA se refiere a la IA que se destina a interactuar con personas físicas tales como los robots de software (por ejemplo, *chatbots*). Desde nuestra perspectiva, si el robot de software se apoya en el lenguaje escrito para inferir emociones o intenciones estaríamos en el nivel de riesgo 3 y se aplicarán las obligaciones del artículo 50 RIA (deber de transparencia). Según el artículo 50 RIA los proveedores garantizarán que los sistemas de IA destinados a interactuar directamente con personas físicas se diseñen y desarrollen de forma que las personas físicas de que se trate estén informadas de que están interactuando con un sistema de IA, excepto cuando resulte evidente desde el punto de vista de una persona física razonablemente informada, atenta y perspicaz, teniendo en cuenta las circunstancias y el contexto de utilización.

Y el artículo 50.3 RIA indica que los responsables del despliegue de un sistema de reconocimiento de emociones o de un sistema de categorización biométrica informarán del funcionamiento del sistema a las personas físicas expuestas a él y tratarán sus datos personales de conformidad con los Reglamentos (UE) 2016/679 y (UE) 2018/1725 y con la Directiva (UE) 2016/680, según corresponda.

Ahora bien, si en la conversación entre el chatbot y la persona trabajadora procesa la voz de la persona física (dato biométrico) podríamos estar en el nivel de riesgo 1 o 2 según las circunstancias.

En efecto, existe un tipo de robótica denominada robots de software que trabaja en las sombras y que puede afectar a los derechos de privacidad de las personas empleadas. Nos referimos a los robots RPA (Automatización robótica de procesos) y a los chatbots que habitan en nuestros ordenadores y smartphones y sobre todo tienen la capacidad de hablar, escuchar, reconocernos y contestarnos.

Se han definido como los trabajadores virtuales de cuello azul frente a la robótica industrial que se asimila más a los trabajadores de cuello blanco. A diferencia de los ro-

[23] https://www.elperiodico.com/es/sociedad/20230920/carceles-cataluna-inteligencia-artificial-control-presos-92321451

bots industriales y los coches robóticos, los robots software no son directamente visibles y no tienen una realidad física (en el sentido que no pesan, no ocupan espacio), sino que son pura lógica, puro software. Si se prefiere decirlo de otra manera, son solo programas. Por eso ni los vemos ni les podemos tocar. Se dice que son habitantes de las sombras y que actúan sobre las aplicaciones o los ficheros[24].

Podemos distinguir dos clases de robots software. Por un lado, los RPA o lo que es lo mismo la automatización robótica de procesos. Los robots RPA se dedican a trabajar con otros activos digitales. Fundamentalmente, interactúan con las pantallas de otras aplicaciones y con documentos ofimáticos como hojas de cálculo o ficheros PDF. Lo que hacen, dicho de forma simplificada, es leer datos de pantallas y de ficheros, realizar cálculos o tomar decisiones basadas en esa información y volver a escribir en pantallas o ficheros los resultados o conclusiones obtenidas[25].

De otra parte, existe una categoría de robots que normalmente se denominan chatbots. Un chatbot es un módulo software cuya misión es interaccionar con personas de forma abierta y natural mediante conversaciones. Se trata de una variedad de robots especializados en dialogar con personas mediante conversaciones naturales. Es decir, en esta categoría de robots incluimos aquellos chatbots que interaccionan con nosotros a través de sistemas de mensajería como Facebook Messenger, Slack o WhatsApp, pero también aquellos otros que hablan con nosotros, emiten voz y escuchan y entienden, a su vez, la voz humana. Es decir, incluimos también a los a veces denominados voicebots y los altavoces inteligentes. Nos estamos refiriendo con ello a software del tipo de Alexa, Siri, Cortana y todo tipo de robots más especializados construidos con capacidad de mantener conversaciones por voz [26].

Tanto los RPA como los chatbots incluyen elementos de inteligencia artificial para reconocimiento de voz, para procesamiento de lenguaje natural, para visión artificial y para reconocimiento óptico de caracteres. Por tanto, estos robots, aparte de la programación, incluyen dosis mayores o menores de adaptación a través de lo que denominamos como aprendizaje[27].

En todo caso, los robots de software pueden leer el correo electrónico y los documentos que se procesan en el puesto de trabajo pudiendo llegar a pulverizar el derecho fundamental de intimidad de la persona trabajadora.

Por último, el nivel de riesgo 4 (riesgo mínimo) incluye entre otros videojuegos con IA o filtros de *spam*. En este nivel de riesgo podríamos mencionar la aventura

[24] La doctrina especializada ha identificado seis características de los robots que se aplican a los robots industriales, biológicos y robots software. Se refieren a la artificialidad, capacidad de adaptación, interacción con el entorno, autonomía, sustitución de personas y similitud con la forma de trabajar las personas. Vid. I. G.R., Gavilán, *Robots en la sombra. RPA, robots conversacionales y otras formas de automatización cognitiv*a, Anaya, Madrid, 2021.

[25] Más extensamente G.R., Gavilán, *Robots en la sombra. RPA, robots conversacionales y otras formas de automatización cognitiv*a, Madrid, Anaya, 2021.

[26] I.G.R., Gavilán, *Robots en la sombra. RPA, robots conversacionales y otras formas de automatización cognitiva*, Anaya, Madrid, 2021, pp. 55 y 88.

[27] I.G.R., Gavilán, *Robots en la sombra. RPA, robots conversacionales y otras formas de automatización cognitiva*, Anaya, Madrid, 2021, pp. 55-56.

gráfica y el uso de la gamificación con finalidad laboral. Este tipo de tecnología puede provocar en las personas empleadas distintas emociones durante el proceso como alegría, frustración, decepción, triunfo…

4. El impacto del nuevo Reglamento europeo de Inteligencia Artificial en España

El impacto del nuevo RIA en España es notable por dos razones principales. En primer lugar, si bien no se suscitado hasta el momento casuística en nuestro país sobre los sistemas automatizados de emociones, algunas empresas del sector del telemarketing están desarrollando programas piloto de control laboral y calidad con apoyo en la IA. Es decir, primero se graban las llamadas entre los clientes y las personas trabajadoras. A continuación, la IA evalúa estas llamadas conforme a los parámetros fijados por la empresa y, finalmente, un trabajador se encarga de verificar si la IA ha cometido errores.

En Europa se han identificado supuestos de reconocimiento automatizado de emociones cuyo interés adquiere más relevancia tras la aprobación del RIA. En el momento de los hechos no estaba aprobado el RIA y se aplicó el RGPD. En efecto, uno de los primeros supuestos aborda el dato biométrico de la voz y fue resuelto por la resolución de la Agencia Húngara de Protección de Datos de 8 febrero 2022. Los hechos describen a un banco que utilizó un software de procesamiento de señales de voz basado en inteligencia artificial. El período de datos del procesamiento fue de 45 días con respecto a la grabación de sonido que se puede escuchar dentro del software y un año con respecto a las estadísticas y listas de llamadas clasificadas generadas a través de la operación de software. El software analizaba y evaluaba los estados emocionales de los clientes y de las personas empleadas. Utilizando los resultados del análisis, la empresa establecía qué cliente insatisfecho precisaba que se le devolviera la llamada y, en relación con esto, analizaba automáticamente, entre otros aspectos, el estado emocional del interesado que llama, así como del empleado del servicio de atención al cliente, junto con otras características de la conversación. La finalidad de esta tecnología fue gestionar las quejas, controlar la calidad de las llamadas y del trabajo y aumentar la eficiencia de las personas trabajadoras. El asunto llegó a la Agencia Húngara de Protección de Datos por la queja de uno de los clientes. El cliente-denunciante formuló al banco algunas preguntas sobre la información publicada en la web de la empresa donde se hacía referencia al análisis de las grabaciones de sonido, sin embargo, no recibió una respuesta satisfactoria e instó el procedimiento ante la Autoridad nacional de Protección de Datos.

En la resolución de la Agencia Húngara de Protección de Datos se analiza el cumplimiento del principio de proporcionalidad. Según el banco, el software no incluía IA; no tomaba decisiones automatizadas y los resultados de su análisis podían ser utilizados exclusivamente con intervención e interpretación humana. Sin embargo, el análisis realizado por la Agencia de Protección de Datos concluye que el software utilizaba IA para tratar de forma automatizada datos personales, resultando, por un lado, en una lista de llamadas en el orden en que deben ser devuelto, y, por otro lado, las emociones recono-

cidas y las características de grabación de voz asociadas con cada llamada (por ejemplo, la duración de las pausas).

Es especialmente interesante la conclusión alcanzada por la Agencia de Protección de Datos cuando analiza la posibilidad de aplicar la protección reforzada de los datos biométricos recogida en el RGPD. A juicio de la Agencia, no se cumplen los requisitos en el caso concreto para ser datos biométricos. Según los hechos explorados del caso, el análisis de voz genera datos, pero estos datos no permiten identificar al titular de los datos, por lo que está ausente la condición de datos biométricos. Con base en esta información, los empleados de la entidad bancaria podían decidir a quién devolver la llamada del servicio de atención al cliente para abordar la insatisfacción. Si bien el software no está diseñado para manejar quejas individuales, las quejas reportadas por teléfono son atendidas de alguna manera por el personal de atención al cliente, independientemente del funcionamiento del software.

Tampoco se aplica el artículo 21 del RGPD sobre las decisiones automatizadas. Esto se debe a que quedan excluidos del artículo 21 RGPD las decisiones automatizadas negativas, es decir, cuando los interesados no son seleccionados para ser llamados o no se reporta ningún error administrativo, por lo que en estos casos se toma una decisión negativa sin intervención humana. Y sobre las decisiones positivas, es decir, el grupo seleccionado de clientes sí había intervención humana. En efecto, se requiere la intervención humana para tomar medidas adicionales en el caso de personas seleccionadas por el software para ser llamados o de empleados para ser revisados, por lo que para estas personas se logra un impacto significativo, pero la condición para una decisión basada en procesos totalmente automatizados no se cumple en el caso concreto.

En definitiva, la Agencia de Protección de Datos aplica el régimen general de garantías en materia de protección de datos. Y a partir de este indica que las actividades de análisis de voz realizadas por el banco utilizando IA, en particular la evaluación de las emociones de los interesados, plantean en sí mismos problemas de protección de datos. Cuando se utilizan herramientas para examinar las características psicolingüísticas y los tonos emocionales del habla no es suficiente la existencia formal del consentimiento de la persona. La tecnología de priorización basada sobre el tratamiento del habla supone una invasión de la privacidad y conlleva el riesgo de que el interesado no sea capaz de reconocer en el momento de dar el consentimiento y evaluar la incidencia en sus derechos. La Agencia indica que la tecnología aplicada permite a la entidad financiera obtener datos de los que el cliente ni siquiera es consciente, por lo que el uso de dichas herramientas reduce la posición del interesado de ser sujeto del procedimiento a ser objeto de este.

Además, se plantean incumplimientos de los principios de proporcionalidad y transparencia. No se proporcionó información a los afectados en relación con el análisis de voz por parte de IA o el propósito de dicho procesamiento y, por lo tanto, no hubo derecho a oponerse al tratamiento realizado. El banco no consideró adecuadamente los intereses en juego. El problema debería haber sido establecer la adecuación y proporcionalidad para el propósito dado del procesamiento; en cambio, la evaluación de la empresa se basó únicamente en sus propios intereses. En realidad, no consideró la pro-

porcionalidad y la posición del afectado, menospreciando los riesgos significativos para los derechos fundamentales.

La Agencia considera que el argumento de la empresa de que podía realizar las actividades empleando menos personal no es en sí mismo una justificación proporcionada y adecuada para la supresión de los derechos fundamentales de los interesados y por el uso de una forma de procesamiento de datos que considera indeseable y que implica un alto riesgo, incluso si se garantizan los derechos adecuados de los interesados. La innovación sólo beneficia a las personas si es apropiada, eficaz y va acompañado de garantías fuertes.

Se insiste en la resolución en la finalidad para la que fueron recabados los datos. Si bien la grabación de la voz es un elemento inevitable en la actividad de atención al cliente, incluso obligatorio en caso de reclamaciones, si el banco desea realizar más operaciones de procesamiento con la voz para analizarlas de forma automatizada utilizando nuevas y no del todo conocidas tecnologías, también debe cumplir con el artículo 6 (4) de RGPD, ya que pretende procesar datos personales para una finalidad distinta para la finalidad para la que se recogieron los datos.

Resulta clave el apartado de la resolución referido a los trabajadores y su posición de debilidad contractual. Indica la Agencia que no se proporciona un sistema adecuado de garantías para los empleados que se encuentran en relación de subordinación y, por tanto, son más vulnerables que los terceros. Los análisis de las emociones, cuya eficacia sigue sin probarse y profunda y severamente limita su derecho a la libre determinación, no puede sustentarse de manera razonable en el caso de los empleados. Dado que los empleados también están sujetos explícitamente a las normas relacionadas con el desempeño en el lugar de trabajo sobre la elaboración de perfiles según el artículo 4 (4) del RGPD, un análisis exhaustivo de las reglas y garantías aplicables a este también es necesario con carácter previo al procesamiento de datos con una nueva tecnología y la empresa no lo tuvo presente. Si un controlador utiliza métodos innovadores y tecnologías menos conocidos, las expectativas son más altas que para las tecnologías clásicas, por lo que mayores garantías y la planificación cuidadosa también deben aplicarse en la supervisión de los empleados. La elaboración de perfiles, en particular el análisis de las emociones de los empleados plantea una serie de problemas legales y éticos. Problemas que no han sido identificados y abordados por el banco en el curso del tratamiento de datos. Expresa la Agencia que son dudosas las posibilidades reales de las personas trabajadoras para oponerse a este tratamiento si tenemos en cuenta la subordinación.

Sobre la base de los argumentos expuestos, la Agencia de Protección de Datos concluye que las prácticas de procesamiento de datos del banco en relación con el análisis de las grabaciones de voz por parte del servicio de atención al cliente suponen una vulneración de los artículos del RGPD 5 (1)(a), 6 (1) y 6 (4). De conformidad con el artículo 12 (1) del RGPD, el banco debe proporcionar a los interesados la información mínima información necesaria para comprender el tratamiento de forma concisa y comprensible, de forma que los interesados sean al menos conscientes de la naturaleza básica del tratamiento. Esta información no se proporcionó por el banco y los clientes no podían sospechar que su voz se analizara automáticamente, y tampoco podía razonablemente

esperar que le devolvieran la llamada sin solicitarlo, entre otras cosas, por el tono de su voz. Por todo ello, se impuso al banco la multa de 670.000 € y le obligó a suspender el análisis de emociones con fundamento en los artículos 12, 24 y 25 RGPD.

Si se produjera un supuesto como el descrito, tras la aprobación del RIA, se aplicaría el primer nivel de riesgo y, por tanto, se trataría de un sistema de IA prohibido en el ámbito laboral al no concurrir las excepciones explicadas sobre los motivos médicos o de seguridad. Por lo tanto, este tipo de prácticas en las empresas están prohibidas.

El segundo de los impactos se refiere a una posible vertiente preventiva del RIA. La reciente aprobación del RIA ha supuesto un hito en el ámbito laboral pero también en el terreno de la seguridad y salud en el trabajo. Las referencias reiteradas a la salud y seguridad en el nuevo reglamento comunitario plantean algunos interrogantes que invitan a la reflexión: ¿Es el RIA una norma de prevención de riesgos laborales?; ¿qué impacto tiene el RIA sobre la Ley española de Prevención de Riesgos Laborales?

El sistema normativo de la prevención de riesgos laborales se compone de normas que tienen orígenes distintos. La cláusula de apertura del sistema que se recoge en el artículo 1º de la Ley de Prevención de Riesgos Laborales (Ley 31/1995, de 8 noviembre, de Prevención de Riesgos Laborales), pone de relieve la procedencia múltiple de las normas de prevención. El tenor literal del artículo 1º de la LPRL dice así: «La normativa sobre prevención de riesgos laborales está constituida por la presente Ley, sus disposiciones de desarrollo o complementarias y cuantas otras normas, legales o convencionales, contengan prescripciones relativas a la adopción de medidas preventivas en el ámbito laboral susceptibles de producirlas en dicho ámbito». Esto es, el sistema se nutre de normas específicamente preventivas, pero también de normas y reglas externas[28].

Como se ha dicho, el nuevo RIA realiza numerosas referencias a la salud y seguridad haciendo alusión expresa al ámbito laboral. Lo que nos permite plantearnos si en realidad esta norma se ha sumado al bloque normativo de la prevención de riesgos laborales. A diferencia de las Directivas comunitarias que han sido abundantes en materia de prevención de riesgos laborales y que precisan de una norma nacional de transposición, el RIA es una norma comunitaria de directa aplicación en España.

Como se ha anticipado, el RIA introduce prohibiciones respecto de determinados sistemas de IA y prevé obligaciones de cierta exigencia para las empresas y entidades públicas que empleen sistemas de IA de alto riesgo. Con la aprobación del RIA se incrementa la protección de los datos de salud de las personas trabajadoras. Si bien no se alteran de forma explícita las reglas y garantías de tratamiento de la salud de las personas empleadas previstas en el RGPD y en la LPRL, se producen avances en el RIA al incluir, con carácter general, la prohibición de los sistemas automatizados de reconocimiento de emociones de las personas trabajadoras (nivel 1 de riesgo) y su admisión condicionada en los niveles 2, 3 y 4 de riesgo.

En relación con las obligaciones de transparencia para las empresas que utilicen los sistemas automatizados de emociones en los niveles de riesgo permitidos, si adopta-

[28] Vid. mi libro A.B. Muñoz Ruiz, *Sistema normativo de la prevención de riesgos laborales*, Lex Nova, Valladolid, 2009.

mos un enfoque preventivo la empresa tendría además la obligación de incorporar los sistemas de IA de alto riesgo en la evaluación de riesgos laborales cuando su uso pueda producir al trabajador un efecto perjudicial considerable en su salud.

En definitiva, sería deseable un mayor desarrollo de la normativa de IA que clarifique cómo afecta la nueva normativa comunitaria a las empresas en el ámbito de la prevención de riesgos laborales. Este desarrollo podría tomar la forma de reforma de algunos preceptos de la Ley de Prevención de Riesgos Laborales y también podría ser útil la aprobación de una guía de actuación para empresas y sindicatos por parte del Instituto Nacional de Seguridad y Salud del Trabajo.