

Al Gobierno de España:

Mediante esta carta abierta PEDIMOS al gobierno una comisión de investigación que estudie la necesidad de establecer una MORATORIA en el uso y comercialización de sistemas de reconocimiento y análisis facial por parte de empresas públicas y privadas hasta que la Cortes Generales y las instituciones legislativas europeas debatan cuáles, de qué manera, en qué condiciones, con qué garantías y con qué objetivos debe permitirse, si cabe, el uso de estos sistemas.

Con regularidad se publican en medios generalistas y especializados noticias acerca de nuevos usos masivos para estos sistemas. A modo de ejemplo reciente podemos citar el caso de RENFE, que ha publicado (y luego retirado) una licitación para desarrollar un sistema de reconocimiento y análisis que, entre otras cosas, debía permitir identificar el género, la etnia o incluso el estado emocional de usuarias y usuarios¹. Asimismo, dicho sistema debía poder analizar las imágenes y alertar acerca de “peleas” o “actitudes antisociales”. Sistemas similares ya están disponibles para entornos educativos, laborales², sanitarios, de ocio y muchos otros, tanto públicos^{3 4} como privados⁵.

La preocupación que motiva esta carta tiene que ver con los potenciales efectos perniciosos que estos sistemas pueden tener sobre el bienestar, los intereses y las necesidades y derechos fundamentales⁶ de la población española en general. Por ello nos parece necesario que el Gobierno intervenga de manera rápida antes de que estos sistemas continúen expandiéndose y se conviertan en estándares de facto, a pesar de la intromisión que suponen en la esfera privada de las personas sin su consentimiento explícito. Están en juego cuestiones fundamentales de justicia social, dignidad humana, equidad, igualdad en el trato e inclusión.

Los sistemas de reconocimiento y análisis de imágenes de personas (de sus caras, gestos, peinados, posturas y movimientos corporales, vestimenta, texturas y/o colores de piel) y, por extensión, los algoritmos de aprendizaje automático que los sustentan computacionalmente tienen graves problemas que han sido ampliamente documentados y discutidos por la comunidad científica y entidades gubernamentales y civiles⁷. A continuación, presentamos una breve selección:

- 1) Asociar a una persona una determinada característica o tendencia (normalmente representada a través de una puntuación individual) con base en estadísticas poblacionales es altamente problemático, tanto desde un punto de vista epistémico como ético. La puntuación individual (un *score* o un porcentaje probabilístico) puede identificar a un estudiante como probable tramposo, a un posible cliente de un banco como potencial moroso debido a que otros “similares a él” han sido morosos, o a una candidata a un puesto de trabajo como poco concienzuda, debido también a que candidatas de perfil similar han resultado ser poco concienzudas. Es muy problemático cuando un sistema toma decisiones operativas que afectan a personas individuales con base en predicciones que pueden ser válidas únicamente a nivel de grupo. Que una persona sea clasificada en una categoría o perfil determinados no determina su comportamiento futuro, esto viola la idea fundamental de que los seres humanos tenemos *libre albedrío* y que nuestro comportamiento no está determinado de antemano, sino que disfrutamos de un espacio de libertad de decisión y propósito. Las eventuales predicciones que se puedan hacer sólo son válidas a nivel colectivo y nunca deberían transformarse en decisiones automatizadas que afecten a las personas a nivel individual.

- 2) En línea con lo anterior, hay suficientes evidencias que indican que asociar posturas, gestos, rasgos faciales, colores de piel, peinados o prendas de vestir a posibles comportamientos “problemáticos”, capacidades intelectuales o económicas puede resultar en clasificaciones racistas, clasistas y/o sexistas⁸. Estas clasificaciones, no cuentan, además, con base científica⁹ y están claramente asociadas a experimentos nefastos del pasado como la eugenesia, la frenología o la criminalística de Lombroso. Más allá de correlaciones estadísticas que no implican relaciones causa-efecto, no hay modelos científicos aceptados en psicología, antropología o sociología que indiquen que un tipo de nariz, un rictus particular o una manera de andar sean predictores adecuados de comportamientos individuales futuros. Por ejemplo, la probabilidad de delinquir, de aprender historia o ingeniería o de desempeñarse correctamente en un trabajo no dependen de ninguna de esas variables que los sistemas de reconocimiento facial recogen y analizan para clasificar a las personas y tomar decisiones sobre ellas.
- 3) En los sistemas de reconocimiento y análisis facial resulta especialmente complejo explicar por qué motivo y con base en qué criterios se ha clasificado a una persona en tal o cual categoría. Por esta dificultad para establecer una relación clara entre la entrada de datos y la predicción o recomendación de salida, los algoritmos de aprendizaje necesarios para estos sistemas se consideran *cajas negras* cuya opacidad dificulta saber cómo toman sus decisiones y en qué criterios las basan. Además, las propias categorías se integran en la caja negra aumentando, si cabe, su opacidad. Si bien es técnica y teóricamente posible desarrollar sistemas más transparentes, los sistemas actuales no están especialmente diseñados para permitir una rendición de cuentas que se ajuste a los requerimientos de una sociedad democrática.
- 4) Estos sistemas además son poco robustos ya que la calidad de sus resultados es altamente dependiente de cuestiones contextuales. Por ejemplo, que el sistema sea utilizado en condiciones de luz similares a las de las imágenes utilizadas para el “entrenamiento” del sistema. Cuando las condiciones difieren de las utilizadas durante el entrenamiento los resultados pueden no ser correctos dando lugar tanto a falsos positivos como a falsos negativos¹⁰. Incluso desde el ámbito privado hay resistencia al uso de esta tecnología por sus debilidades¹¹.
- 5) Es fácil que un sesgo muestral afecte de manera notable la calidad predictiva de un sistema cuando diferentes grupos no están representados de forma igualitaria en los datos de entrenamiento. Por ejemplo, aplicaciones de radiología que exhiben índices de éxito predictivo altamente prometedores en personas de piel clara muestran resultados mucho peores con personas de piel más oscura¹².

Debido a las graves deficiencias y riesgos que presentan estos sistemas, los posibles beneficios que quizás podrían ofrecer no compensan de ninguna manera sus potenciales efectos negativos, en especial para los grupos y colectivos que suelen sufrir injusticias y tratos discriminatorios: entre otros, mujeres, personas LGTBIQ+¹³, personas racializadas, migrantes, personas con discapacidad o personas en situación de pobreza y riesgo de exclusión social.

No hemos comentado aquí otra gama de cuestiones relevantes como el profundo condicionamiento coercitivo que trae aparejada una vigilancia ubicua. Tampoco hemos considerado cuestiones de privacidad, seguridad o el consentimiento para la recogida y tratamiento de los datos (algo que en investigación con humanos es absolutamente necesario, y

que no parece estar siendo aplicado en la utilización de estos sistemas)¹⁴ y hemos preferido concentrarnos en cuestiones de justicia social e igualdad. Sin embargo, vale la pena apuntar que se trata de problemas interrelacionados, dado que los datos recogidos mediante sistemas para reconocimiento y análisis facial potencian los riesgos y efectos negativos en los otros ámbitos.

Existe pues una variada gama de razones (tanto técnicas como éticas) que JUSTIFICAN la creación de una comisión para investigar la necesidad de una moratoria, que nosotros consideramos imprescindible y urgente. Sugerimos que esta comisión sea independiente y esté compuesta por científicos, juristas, expertos en ética e inteligencia artificial y por miembros de la sociedad civil, especialmente de aquellos colectivos que pueden verse *prima facie* afectados por estos sistemas. Es muy importante, además, aprovechar esta oportunidad para llevar a cabo un profundo debate social acerca del rol que estos sistemas pueden jugar en nuestras vidas y bajo qué condiciones.

Firmantes (en orden alfabético)

Ujué Agudo Díaz, *Responsable de Bicolabs*, Biko.

Patricia Amigot Leache, *Profesora Dpto. Sociología y Trabajo Social*, Universidad Pública de Navarra (UPNA).

Pablo Aragón, *Profesor Asociado*, Universitat Pompeu Fabra.

Ignacio Arganda Carreras, *Ikerbasque Research Associate*, Universidad del País Vasco/ Euskal Herriko Unibertsitatea.

César Astudillo González, *Asesor Independiente de Diseño Estratégico e Innovación Profesor Asociado / Program Leader - Masters in Design and Innovation*, Instituto Europeo de Design (Madrid).

Txetxu Ausín, *Científico Titular*, Instituto de Filosofía, CSIC.

José L. Aznarte, *Profesor titular del departamento de IA y Vicerrector adjunto de Digitalización e Innovación*, UNED.

Manuela Battaglini Manrique de Lara, *Abogada y consultora de Ética Digital, CEO*, Transparent Internet.

Alejandro Benito-Santos, *Investigador Postdoctoral*, Universidad de Salamanca.

Ana Bravo, *Profesora contratado doctor*, Universidad de Murcia.

Marcel Cano Soler, *Profesor asociado*, Universidad de Barcelona, Universidad de Vic.

David Casacuberta, *Profesor de filosofía*, Universidad Autónoma de Barcelona.

Diego Casado Mansilla, *Profesor e investigador (ayudante doctor)*, Facultad de ingeniería, Universidad de Deusto.

Mario Castro Ponce, *Catedrático de Universidad*, Universidad Pontificia Comillas.

Diego Cenzano Cestafe, *CEO*, Biko.

Laura Clèries García, *Directora de Elisava Research*, Facultad de Diseño e Ingeniería Elisava.

Jesús Conill Sancho, *Catedrático de Filosofía Moral*, Universidad de Valencia.

Adela Cortina, *Catedrática Emérita de Ética y Filosofía Política*, Universidad de Valencia.

Pilar Dellunde, *Catedrática Universitaria e Investigadora*, IIIA-CSIC.

Lorena Fernández Álvarez, *Directora de comunicación digital*, Universidad de Deusto.

Jorge Flores Fernández, *Director*, PantallasAmigas.

Luis Fuentes Melero, *Catedrático de Universidad*, Universidad de Murcia.

Albert Fuster Martí, *Director académico*, Facultad de Diseño e Ingeniería Elisava.

Gemma Galdon Clavell, *Presidenta*, Fundación Eticas www.eticasfoundation.org.

Daniel Gamper, *Profesor Titular de Filosofía Política*, Universitat Autònoma de Barcelona.

Pablo Garaizar Sagarminaga, *Profesor/investigador*, Universidad de Deusto.

Andrea García Rodríguez, *Investigadora*, CIDOB (Barcelona Centre for International Affairs).

Iñigo González Ricoy, *Profesor agregado de Filosofía*, Universitat de Barcelona.

Ariel Guersenzvaig, *Profesor titular*, Facultad de Diseño e Ingeniería Elisava.

Yusef Hassan Montero, *Director del Máster de Diseño de Experiencia de Usuario*, Universidad Internacional de la Rioja.

Angel Ferrández Izquierdo, *Catedrático de Universidad*, Universidad de Murcia.

Miren Itziar Laka Mugarza, *Catedrática del Departamento de Lingüística y Estudios Vascos*, Universidad del País Vasco/ Euskal Herriko Unibertsitatea.

Simona Levi, *XNet*, XNet.

Karlos G. Liberal, *Responsable de Bicolabs*, Biko.

Rosa Llop, *Profesora e investigador residente*, Hangar, Centro de Investigación y Producción Artística. / Facultad de Diseño e Ingeniería Elisava.

José Antonio Llorente Sesma, *Profesor asociado MU y Doctorando UOC*, MU Facultad de empresariales, Mondragón Unibertsitatea.

M^a Teresa López de la Vieja, *Catedrática*, Universidad de Salamanca.

Cristina López Tarrida, *Ingeniero, Analista de Contrainteligencia*, Profesional independiente.

Ramon López de Mántaras, *Premio Nacional de Investigación "Julio Rey Pastor" en Matemáticas y TIC 2018*, Institut d'Investigació en Intel·ligència Artificial CSIC.

Nerea Luis Minguenza, *Doctora, divulgadora y profesora asociada*, Universidad/institución UNIR y Universidad Nebrija.

David Luque Ruiz, *Investigador (Captación de Talento, Senior)*, Universidad Autónoma de Madrid.

Pablo Malo Ocejo, *Psiquiatra*, Red de Salud Mental de Bizkaia. Osakidetza.

Pablo Martín, *CTO*, Political Watch.

Manuel Martín-Loeches Garrido, *Catedrático de Psicobiología*, Universidad Complutense de Madrid.

Juan Ramón Martín San Román, *Profesor Encargado de Cátedra Facultad de Comunicación*, Universidad Pontificia de Salamanca.

Laureano Martinez, *Profesor Ayudante Doctor*, Universidad Pública de Navarra.

Emilio Martínez Navarro, *Catedrático de Filosofía Moral y Política, Director del Departamento de Filosofía*, Facultad de Filosofía de la Universidad de Murcia.

Pablo Martínez Garrido, *Investigador*, Universidad del País Vasco / Euskal Herriko Unibertsitatea.

Helena Matute Greño, *Catedrática de Psicología*, Universidad de Deusto, Bilbao.

Imanol Miramón Monasterio, *Editor*, Editorial Katakarak.

Núria Sara Miras Boronat, *Profesora agregada Facultad de Filosofía*, Universitat de Barcelona.

Luis Montero Montero, *Autor*, Pensar el diseño.

Roger Paez i Blanch, *Profesor titular e investigador*, Facultad de Diseño e Ingeniería Elisava.

M^a José Pedraja Linares, *Profesora Titular*, Departamento de Psicología Básica y Metodología, Universidad de Murcia.

Javier Peña Andrés, *Director Científico*, Facultad de Diseño e Ingeniería Elisava.

Juan Ignacio Pérez Iglesias, *Director de la Cátedra de Cultura Científica*, Universidad del País Vasco.

Angel Puyol, *Catedrático de Ética*, Universitat Autònoma de Barcelona.

Emmanuel Rodríguez López, *Editor*, Traficantes de Sueños.

David Rodríguez-Arias, *Profesor Titular*, Universidad de Granada.

Begoña Román Maestre, *Profesora agregada Facultad de Filosofía*, Universidad de Barcelona.

Charo Sádaba, *Profesora titular*, Facultad de Comunicación, Universidad de Navarra.

Javier Sánchez Monedero, *Investigador Distinguido "Beatriz Galindo"*, Universidad de Córdoba.

Ramon Sangüesa Solé, *Profesor Titular*, Dept. Ciencias de la Computación. Universitat Politècnica de Catalunya.

Rodrigo Santamaría, *Profesor Titular del Departamento de Informática y Automática*, Universidad de Salamanca.

Borja Sanz Urquijo, *Profesor e Investigador Facultad Ingeniería*, Universidad de Deusto.

Carles Sierra, *Director Institut d'Investigació en Intel·ligència Artificial*, IIIA-CSIC.

Ofelia Tejerina Rodríguez, *Presidenta*, Asociación de Internautas.

Roberto Therón Sánchez, *Profesor Titular, Ciencia de la Computación e Inteligencia Artificial*, Universidad de Salamanca.

Ana Valdivia, *Research Associate*, King's College London (UK).

Carissa Véliz, *Profesora Asociada*, Universidad de Oxford.

Francisco R. Villatoro Machuca, *Profesor Titular de Ciencia de la Computación e Inteligencia Artificial*, Universidad de Málaga.

Nota: las firmas son adhesiones a título personal y no representan a las instituciones mencionadas.

Nuevas adhesiones:

<http://bit.ly/CartaalGobierno>

Versión más actual de la carta:

<http://bit.ly/CartaGobiernoVersionActual>

Promotores y contacto:

cartarecfacial@gmail.com

Ujué Agudo

David Casacuberta

Ariel Guersenzvaig

Karlos G. Liberal

¹ Mcloughlin, M. (2021). *Origen étnico, sexo o vestimenta: El polémico sistema de RENFE para vigilar a sus viajeros*. https://www.elconfidencial.com/tecnologia/2021-02-17/renfe-videovigilancia-plego-condiciones-tecnicos_2953824/ Dado el carácter abierto de esta carta intentamos incluir fuentes generalistas. Es muy frecuente la publicación de *papers* presentando sistemas con supuestas capacidades similares en la literatura científico-técnica especializada.

² Sarabia, D. (2019). *Varias empresas multinacionales ya usan inteligencia artificial para analizar la personalidad en las entrevistas de trabajo*. https://www.eldiario.es/tecnologia/varias-multinacionales-inteligencia-psicologicamente-entrevistas_1_1159455.html

³ Aguiar, A. (2020). *Una estación de autobuses de Madrid lleva años utilizando tecnologías de reconocimiento facial sin que casi nadie se haya dado cuenta*. <https://www.businessinsider.es/usa-reconocimiento-facial-estacion-autobuses-madrid-651165>

⁴ De Vega, L. (2021). *Gran Hermano llega al Pequeño Caribe de Madrid*. <https://elpais.com/espana/madrid/2021-02-01/gran-hermano-llega-al-pequeno-caribe-de-madrid.html>

⁵ Herranz, A. (2021). *Iberia te permitirá facturar "por la cara": Así está desplegando un sistema de reconocimiento facial que te reconoce hasta con mascarilla*. <https://www.xataka.com/pro/iberia-te-permitira-facturar-cara-asi-esta-desplegando-sistema-reconocimiento-facial-que-te-reconoce-mascarilla>

⁶ FRA (2018). *Under watchful eyes – biometrics, EU IT-systems and fundamental rights*. <https://fra.europa.eu/en/publication/2018/under-watchful-eyes-biometrics-eu-it-systems-and-fundamental-rights>

⁷ Por ejemplo, European Union Agency for Fundamental Rights, European Digital Rights (EDRi), Ada Lovelace Institute, AlgorithmWatch, Algorithmic Justice League, y la American Civil Liberties Union entre muchas otras.

⁸ Ver por ejemplo Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim Code*. Cambridge: Polity; o Comisión Europea (2020). *Gendered innovations 2. How inclusive analysis contributes to research and innovation* Bruselas: Dirección General de Investigación e Innovación de la Comisión Europea.

⁹ Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., y Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1), 1-68.

¹⁰ Harlan, E., y Schnuck, O. (2021). *Objective or biased*. <https://web.br.de/interaktiv/ki-bewerbung/en/>

¹¹ Es el caso del proveedor corporativo más grande del mundo de cámaras para cuerpos de policía (Axon), que en el año 2019 anunció que no desplegaría tecnología de reconocimiento facial en ninguno de sus productos, "porque era demasiado poco fiable para el trabajo policial y podría exacerbar las desigualdades existentes en el trabajo policial, por ejemplo, penalizando a las comunidades negras o LGBTQ". Crawford, K. (2019). Regulate facial-recognition technology. *Nature*, 572, 565.

¹² Adamson, A. S., y Smith, A. (2018). Machine learning and health care disparities in dermatology. *JAMA Dermatology*, 154(11), 1247-1248.

¹³ Keyes, O. (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), Article 88.

¹⁴ Somos conscientes de que el Reglamento General de Protección de Datos de la Unión Europea, entre otras leyes, ya regula el tratamiento de datos especialmente protegidos como el origen racial o la orientación sexual, y dispone la prohibición del tratamiento de datos biométricos dirigidos a identificar a una persona específica, así como la necesidad de la obtención de consentimiento informado previo y de garantizar la seguridad de los datos tratados.