Digital Future Society

Datathon sobre Derechos Digitales y Privacidad en Reconocimiento Facial

Un programa de

GOBIERNO DE ESPAÑA
VICEPRESIDENCIA PRIMERA DE GOBIERNO
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es

MOBILE WORLD CAPITAL BARCELONA

Índice

1
Introducción y Contexto

El reconocimiento facial automatizado se realiza a través de software que identifica o confirma la identidad de una persona a partir del rostro. Funciona mediante la identificación y medición de los rasgos faciales en una imagen. Los valores de la identificación y mediciones son datos biométricos, es decir, datos personales relativos a características únicas del ser humano, sean físicas, fisiológicas o asociadas al comportamiento.

Los sistemas de reconocimiento facial facilitan y garantizan la **identificación de una persona**, determinan si el rostro que aparece en dos imágenes pertenece a la misma persona, y permiten encontrar un rostro concreto entre una gran colección de imágenes existentes.

En los últimos años, ha habido un notable avance en el desarrollo y uso de los sistemas de reconocimiento facial. Estos sistemas han encontrado aplicaciones en diversos sectores de gran relevancia, como la seguridad, el comercio, las finanzas y la salud. Sus empleos abarcan una amplia gama de áreas, como **la identificación de sospechosos de delitos, el rastreo de personas desaparecidas y la mejora de la seguridad en espacios públicos**. Además, en la industria minorista, han contribuido a enriquecer la experiencia del cliente al permitir la personalización de la publicidad y la identificación de clientes valiosos, entre otros ejemplos destacados.

Si, por un lado, existen claros beneficios, por otro lado, genera **preocupación su impacto en la privacidad y en los derechos fundamentales de las personas**, ya que puede permitir una recopilación masiva de datos biométricos sin el consentimiento explícito y consciente de las personas, o una utilización para la discriminación y el acoso, o combinado con otras funcionalidades de sistemas de Inteligencia Artificial, puede multiplicar el impacto en la generación de prejuicios o discriminación sistemática.

En este contexto, Digital Future Society, una iniciativa conjunta de la **Secretaría de Estado de Digitalización e Inteligencia Artificial, Red.es** y **Mobile World Capital Barcelona**, ha impulsado la realización de un **ejercicio de** Datathon **sobre identidad digital, privacidad, protección de datos y derechos digitales** con el objetivo de abordar dos aspectos clave relacionados con el reconocimiento facial: cómo se realiza la recuperación y clasificación de datos abiertos necesarios para el reconocimiento facial y el análisis de cómo los sistemas existentes de reconocimiento facial que utilizan estos datos pueden cumplir con el nuevo marco de legislación europea relacionada con el tema.

En la Unión Europea, el uso de tecnologías de reconocimiento facial está regulado por el **Reglamento General de Protección de Datos (RGPD)**, que establece las condiciones en las que se puede procesar y emplear los datos biométricos, como los obtenidos a través del reconocimiento facial. El RGPD establece que los datos biométricos solo pueden ser procesados con el consentimiento explícito de la persona y solo si es necesario para el cumplimiento de una tarea específica.

Por otro lado, la Comisión Europea ha publicado una serie de directrices en las que se establecen los requisitos para la utilización de la tecnología de reconocimiento facial. En estas directrices se indica que el empleo de la tecnología de reconocimiento facial debe ser **transparente, proporcionado y necesario** para la finalidad específica para la que se recopilan los datos.

La Datathon en cuestión es parte de una serie de acciones llevadas a cabo por Digital Future Society, centradas en la **importancia del origen y uso de los datos en el diseño, construcción y regulación de la inteligencia artificial (IA)**. Además de abordar aspectos como el Reglamento General de Protección de Datos (RGPD) y las directrices existentes para el uso de la tecnología de reconocimiento facial, esta iniciativa también se enfoca en el primer marco legal sobre IA propuesto por la Comisión Europea, conocido como el "EU AI Act". Se espera que este marco legal se implemente en los países durante el segundo semestre de 2023 y clasifica los riesgos asociados con la utilización específica de la IA en cuatro niveles: riesgo inaceptable, riesgo alto, riesgo limitado y riesgo mínimo. Cabe destacar que este marco legal está estrechamente vinculado con el Plan Coordinado de la Comisión Europea sobre IA.

Este marco establece que un tribunal puede ordenar la divulgación de evidencias relevantes sobre sistemas de IA de alto riesgo específicos de los que se sospecha que han causado daños. Y clarifica y ajusta muchos de **los requisitos para los sistemas de IA de alto riesgo, como el de reconocimiento facial,** de tal manera que son técnicamente más factibles y menos difíciles de cumplir para las partes interesadas, en aspectos como, por ejemplo, en lo que respecta a la calidad de los datos o en relación con la documentación técnica que deben redactar las PyMEs para demostrar que sus sistemas de IA de alto riesgo cumplen los requisitos.

En los recientes cambios que se han aplicado en cuanto a este entorno normativo destacar que se han dado pasos importantes hacia la adopción de **una ley que regula las herramientas de inteligencia artificial (IA).** Aspectos tan relevantes como la limitación del uso de la IA sin dejar de incentivar la innovación en el sector, la prohibición de ciertos usos como el uso de sistemas de reconocimiento biométrico y facial que pueden calificar socialmente a las personas en función de su comportamiento social, estatus socioeconómico o personal están completamente prohibidos. Además, no será posible recurrir a soluciones de reconocimiento de emociones por parte de agencias de seguridad, controles fronterizos, centros educativos y lugares de trabajo. El proyecto de ley se presentará al pleno parlamentario en junio para su votación. Una vez aprobado, se iniciarán las negociaciones con los estados miembros de la UE para acordar una ley definitiva.

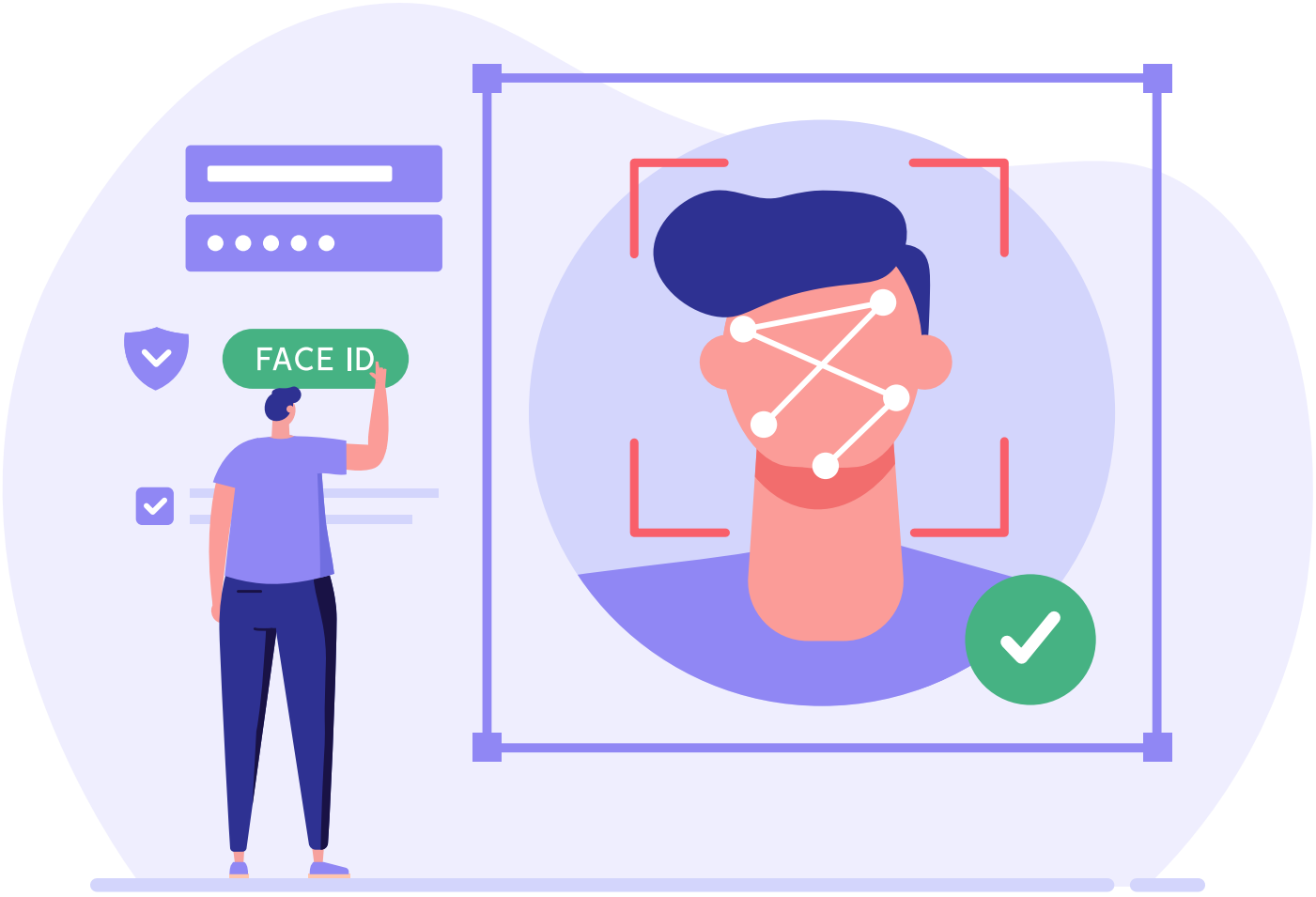
Dado que los sistemas de IA se desarrollan y distribuyen a través de cadenas de valor complejas, el texto de este nuevo reglamento incluye cambios que aclaran **la asignación de responsabilidades y roles de los diversos actores en esas cadenas, en particular los proveedores y usuarios de los sistemas de IA.** Y aclara la relación entre las responsabilidades en virtud este reglamento y las responsabilidades que ya existen en virtud de otra legislación, como la legislación sectorial o el RGPD, incluso en lo que respecta al sector de los servicios financieros.

En este contexto, este ejercicio de Datathon es la tercera acción en la línea de actuación que viene realizando Digital Future Society desde el pasado 1 de diciembre del 2022, sobre ética e innovación alrededor de la procedencia y el uso de los datos y su relación con la evolución exponencial del diseño, construcción y regulación de la Inteligencia Artificial (IA).

En esta tercera fase de nuestra iniciativa, hemos trabajado arduamente para recopilar datos abiertos que desempeñen un papel fundamental en el establecimiento de un marco de normalización. Nuestro enfoque se ha centrado específicamente **en considerar los datos biométricos como una fuente confiable y auténtica**. Para ello, nos hemos basado en la definición de datos abiertos, que se refiere a aquellos datos que están disponibles para que cualquier persona los utilice, reutilice y redistribuya, siempre y cuando se atribuya adecuadamente su fuente o se reconozca su autoría, en caso de ser requerido.

En este sentido, el principal resultado que hemos querido alcanzar con la datathon es la **creación de un repositorio de conjuntos de datos abiertos sobre reconocimiento facial que cumplan con altos estándares de seguridad.** Este repositorio será valioso para el desarrollo de nuevas aplicaciones, especialmente aquellas que hacen uso de tecnologías emergentes como la inteligencia artificial. Además, también nos hemos enfocado en identificar y alertar sobre aquellos conjuntos de datos que, aunque puedan parecer abiertos, presenten riesgos de seguridad.

Dado que es crucial considerar tanto los beneficios como los riesgos asociados con la tecnología de reconocimiento facial como parte de los sistemas de inteligencia artificial, el diseño de la datathon va más allá de la búsqueda, recuperación y clasificación de datos de reconocimiento facial. También nos esforzamos por **fomentar la reflexión y el debate en torno al uso responsable y ético de esta tecnología**. A través de un enfoque colaborativo, buscamos alertar y gestionar los riesgos potenciales, al tiempo que maximizamos los beneficios que esta tecnología puede ofrecer.


FACE ID

¿Por qué focalizar en datos sobre reconocimiento facial?

En el caso específico del reconocimiento facial, es necesario encontrar conjuntos de datos de alta **calidad** y **relevantes** que sean **representativos** de la **diversidad** de **características** faciales que existen en la población, para evitar posibles sesgos en los resultados.

Además, se deben considerar **aspectos éticos y legales en la recopilación y empleo de estos datos.** La búsqueda y selección cuidadosa de conjuntos de datos es un paso crucial en la preparación para la Datathon, que garantizará la calidad y fiabilidad de los resultados obtenidos.

> En la búsqueda de datos se han considerado aquellos que sean procedentes de **fuentes de datos confiables.** Entendiendo que proporcionan información precisa, verificable y actualizada sobre un tema específico. Una fuente de datos confiable también debe ser imparcial y objetiva en su presentación de la información, evitando cualquier forma de sesgo o influencia. Para determinar si una fuente de datos es confiable, es importante considerar quién la creó y cómo se recopiló la información. Las fuentes de datos confiables suelen ser aquellas que provienen de organizaciones o instituciones reconocidas y establecidas en el área temática en cuestión. Además, estas fuentes deben utilizar métodos de investigación rigurosos y bien documentados para recopilar y analizar los datos.

Las fuentes de datos pueden variar en su nivel de confiabilidad según el tema o el contexto en el que se empleen. Por lo tanto, es necesario **evaluar cuidadosamente la confiabilidad de las fuentes de datos en cada situación específica.**

¿Cuáles son los atributos de los datos de reconocimiento facial?

En el contexto del análisis de datos, un atributo se refiere a una característica o propiedad de un objeto o entidad que se está estudiando. En el caso del reconocimiento facial, un atributo se puede entender como una característica o propiedad de la cara de una persona que se emplea para su identificación o verificación.

De los diversos atributos resulta importante **priorizar** aquellos conjuntos de datos que más impacto pueden tener en la protección de las personas desde el punto de vista de la normativa. Con este objetivo a continuación se describen los atributos que se propone buscar, categorizar y testear de forma prioritaria:

Forma de la cara

Este atributo se refiere a la forma general de la cara de una persona, que puede variar en términos de su ancho, longitud y proporciones entre diferentes partes de la cara.

Características faciales

Estos atributos se refieren a características específicas de la cara, como los ojos, la nariz, los labios, las orejas y las cejas, que pueden usarse para identificar a una persona.

Textura de la piel

La textura de la piel, incluyendo arrugas, manchas y cicatrices, también se puede usar como un atributo para la identificación facial.

Patrones de cabello y barba

Estos atributos pueden ser útiles en la identificación de personas con diferentes estilos de cabello y barba.

Expresiones faciales

Los atributos relacionados con las expresiones faciales, como la sonrisa, la mueca o el ceño fruncido, también pueden ser útiles en la identificación facial.

Género y edad

Estos atributos se refieren al género y la edad de la persona, que pueden ser relevantes para ciertos propósitos de identificación.

Digital Future Society

¿Cuál es la relación entre conjuntos de datos y repositorios?

Cuando hablamos de conjunto de datos se hace referencia a un conjunto de observaciones o medidas recopiladas para su análisis y empleo posterior. En el contexto de una "Datathon" orientada al reconocimiento facial, un conjunto de datos puede incluir imágenes faciales etiquetadas con información como la edad, el género, la expresión facial, la iluminación, la resolución, entre otros.

Por otro lado, cuando hablamos de **repositorio** nos referimos a un lugar centralizado donde se almacenan, organizan y gestionan los datos y otros recursos digitales. En general, un repositorio puede ser una fuente valiosa de datos y recursos para una datathon, ya que puede **proporcionar un acceso centralizado y organizado a una amplia variedad de datos relacionados con el tema.** Además, un repositorio puede fomentar la colaboración y el intercambio de conocimientos entre los participantes de la datathon.

2
Detalles sobre la Datathon

Objetivos

El **objetivo** del datathon era **encontrar conjuntos de datos abiertos sobre reconocimiento facial** para clasificarlos bajo una serie de parámetros propuestos y posteriormente poder comprobar si cumplen con la normativa de la Unión Europea en cuanto a privacidad y protección de datos[1].

Se trataba de **elaborar un repositorio con estos datos abiertos de forma que se facilite su uso en soluciones que empleen la biometría como un elemento de autenticación unívoca**. Los conjuntos de datos resultantes deberán ser fiables, confiables y seguros para que puedan entrenar y poner en marcha soluciones/aplicaciones de tecnologías biométricas, que sean garantistas con la seguridad de la privacidad de las personas y cumplan la nueva 'Regulación del reconocimiento facial en la Unión Europea'.

Para ello se **convocó** a todos aquellas personas profesionales, **expertos en reconocimiento facial, daters y *data scientists*** a recopilar estos conjuntos de datos y crear colaborativamente el repositorio propuesto.

[1] La normativa a la que se hace referencia es el RGPD, paquete de medidas marco en Europa adoptado en mayo de 2016 y aplicación efectiva desde el 25 de mayo de 2018 (Reglamento (UE) 2016/679)

Punto de partida: buscar, clasificar y revisar

Las personas participantes han tenido que buscar conjuntos de datos abiertos que contengan datos biométricos y que puedan ser utilizados para el entrenamiento de aplicaciones de reconocimiento facial.

De los diversos atributos mencionados anteriormente para la clasificación, **definimos las siguientes categorías** que los agrupan e incorporan nuevos a tener en cuenta:

Imágenes faciales: Las imágenes faciales son la forma más común de datos usados en el reconocimiento facial. Estas imágenes se capturan empleando cámaras y se utilizan para crear una plantilla única de la cara de una persona.

- Imágenes en 2D
- Imágenes en 3D
- Imágenes en tiempo real
- Imágenes de alta resolución
- Imágenes con diferentes condiciones de iluminación
- ...

Características faciales: Las características faciales son los rasgos distintivos de la cara de una persona, como la forma de los ojos, la nariz, la boca y la barbilla. Estas características se analizan y comparan con una base de datos de características faciales para determinar la identidad de una persona.

- Características geométricas (por ejemplo, distancia entre ojos, ancho de la nariz)
- Características texturales (por ejemplo, patrones de la piel, líneas de expresión)
- Características basadas en puntos de referencia (por ejemplo, puntos específicos en la cara, como la punta de la nariz o la comisura de los labios)

Otros datos biométricos: Además de las imágenes faciales, algunos sistemas de reconocimiento facial también utilizan otros datos biométricos, como el reconocimiento de iris o de la retina, para aumentar la precisión del sistema:

- Reconocimiento de iris
- Reconocimiento de la retina
- Género
- Edad

Patrones de movimiento: Los patrones de movimiento también se utilizan en el reconocimiento facial, ya que cada persona tiene un patrón único de movimiento al hablar, sonreír o mover la cabeza.

- **Movimientos de los labios durante el habla**
- **Movimientos de los ojos durante el parpadeo**
- **Movimientos de la cabeza durante el movimiento y la orientación**
- **Gestos de la cara, como sonreír o fruncir el ceño**

Por ejemplo, en el conjunto de datos del proyecto D4Fly.eu, que ha recibido financiación del programa de investigación e innovación Horizonte 2020 de la Unión Europea en virtud del acuerdo de subvención n.º 833704, se incluyen seis tipos de datos: 3D face, thermal face, iris on-the-move, iris mobile, somatotype, smartphone sensors. Estos los podríamos clasificar por categorías de la siguiente manera:

- 3D face -> **imágenes faciales**
- thermal face -> **imágenes faciales**
- iris on-the-move -> **Otros datos biométricos**
- iris mobile -> **Otros datos biométricos**
- somatotype -> **Características faciales**
- smartphone sensors -> **otros**

Si realizamos una búsqueda en cualquier buscador podemos encontrar, sin problema, fuentes de datos sobre el tema, por ejemplo:

- Face Recognition Homepage - Databases (face-rec.org)
- Facial Expression | Hume AI
- Verificación de identidad online, segura y automática - Alice Biometrics

Esta primera clasificación se vio refrendada por otros equipos en una revisión de pares para asegurar que sea de consenso. Este proceso de verificación se realizará tanto en esta primera parte de búsqueda y clasificación como en la segunda de comprobación de que los datos no rompen ningún elemento de privacidad y seguridad establecido por la Unión Europea.


Data Challenge sobre derechos digitales
y privacidad en reconocimiento facial

Comprobar seguridad y privacidad

En la segunda parte de la Datathon, una vez efectuada la clasificación, se comprobó si estos **cumplían con los requisitos legislativos de la Unión Europea en materia de privacidad y protección de datos.**

Partimos de la base que para que un conjunto de datos de entrenamiento sea acorde con la legislación europea debe tener en cuenta una serie de elementos importantes:

- **Consentimiento informado**
- **Anonimización**
- **Minimización de datos**
- **Seguridad y protección**
- **Retención limitada**
- **Transparencia**
- **Responsabilidad**
- **Propósito legítimo**
- **Derechos de las personas afectadas**
- **Evaluación de impacto**

Es importante tener en cuenta que, aunque estos criterios son importantes para cumplir con las regulaciones de la UE, también es necesario **evaluar el impacto ético y social del uso de datos biométricos y tener en cuenta las preocupaciones sobre la discriminación y la privacidad.** Por lo tanto, se pueden tener en cuenta otra serie de elementos como:

- **Sesgo y discriminación**
- **Igualdad de género**
- **Proporcionalidad**

Se ha intentado verificar que cada uno de los conjuntos de datos encontrados cumplían con estos requisitos. En algunos casos no ha sido posible, como se explica más adelante.

Por otro lado, se recomienda conocer los principios éticos de la UE para identificar que el diseño de IA en los sistemas de reconocimiento facial no vulnera la privacidad y la protección de datos en cuanto al empleo de los conjuntos de datos encontrados y clasificados.

Debido a las consecuencias que puede tener generar falsos positivos o falsos negativos en el reconocimiento facial, el cumplimiento de las Directrices Éticas para una **IA fiable** requiere vigilancia, foco, efectividad y consciencia.

Una IA se considera fiable si se cumplen las siguientes tres características, necesarias pero no suficientes en sí mismas, de manera armónica y simultánea a lo largo de todo el ciclo de vida del sistema:

- debe ser **lícita**, es decir, cumplir todas las leyes y reglamentos aplicables;
- debe ser **ética**, de modo que se garantice el respeto de los principios y valores éticos;
- debe ser **robusta**, tanto desde el punto de vista técnico como social, ya que incluso si las intenciones son buenas, los sistemas de IA pueden provocar daños accidentales.

Contenidos complementarios

- Web de la datathon
- Empleo de datos biométricos: Evaluación desde la perspectiva de protección de datos
- Reconocimiento facial en servicios de seguridad privada
- Directrices éticas para una ia fiable
- Informe sobre la inteligencia artificial en la era digital
- The AI Act
- RGPD (Reglamento (UE) 2016/679)
- Regulating facial Recognition in Europe
- Datos abiertos del gobierno de España
- Datos abiertos y reutilización de la información del sector público
- Publicación en Nature sobre reconocimiento facial
- Publicación sobre los sesgos de los algoritmos

3
Resultados de la Datathon

En este repositorio se encuentran los conjuntos de datos analizados durante la datathon #DFSdatathon23 **sobre identidad digital, privacidad, protección de datos y derechos digitales,** organizada por Digital Future Society, una iniciativa conjunta de la **Secretaría de Estado de Digitalización e Inteligencia Artificial, Red.es y Mobile World Capital Barcelona.**

Durante la jornada se identificaron un total de **56 conjuntos de datos** que se han clasificado en "**COMPROBADOS**" y "NO COMPROBADOS".

La lista denominada «COMPROBADOS», la cual incluye aquellos conjuntos de datos en los que se pudo verificar al menos uno de los factores relacionados con ética y seguridad que se requerían.

Dentro del listado «COMPROBADOS», que consta de un total de **39 conjuntos de datos**, se identifican 13 que están completamente abiertos, mientras que los restantes 26 requieren algún tipo de registro para su descarga.

En cuanto a la versatilidad de los conjuntos de datos (porcentaje sobre el total de categorías en las que pueden ser clasificados basándose en los tipos de datos que contienen (imágenes faciales, características faciales, otros datos biométricos y patrones de movimiento), se observa que **la media de versatilidad es del 52,56%.** Es interesante destacar que aquellos conjuntos de datos que requieren registro presentan una mayor versatilidad, con un promedio del **57,69%.**

Si consideramos los aspectos éticos y legales (compuestos por 12 elementos; consentimiento informado, anonimización, minimización de datos, seguridad y protección, retención limitada, transparencia, propósito legítimo, derechos de las personas afectadas, evaluación de impacto, sesgo y discriminación, igualdad de género y proporcionalidad), se observa que **la media de cumplimiento es del 37.18%.** En este caso, los conjuntos de datos abiertos presentan un mayor porcentaje de cumplimiento, alcanzando el 42.95%, en comparación con el 34.29% de aquellos que requieren algún tipo de registro.

Acceso	Total	Versatilidad	Ética & legal
Registro	26	57,69%	34,29%
Abierto	13	42,31%	42,95%
Suma total	**39**	**52,56%**	**37,18%**

Por lo que hace a la tipología de datos, si son datos sobre reconocimiento facial o relacionados, de los datasets analizados hay **37 que son primarios y solo dos relacionados** (conjuntos de datos que no son directamente de datos faciales). Si miramos la versatilidad ambos están sobre el 50% y en cuanto a factores éticos y legales también se sitúan ambas categorías sobre el 37%.

Tipología	Total	Versatilidad	Ética & legal
Primario	37	52,70%	37,16%
Relacionado	2	50,00%	37,50%
Suma total	**39**	**52,56%**	**37,18%**

Si focalizamos aún más la atención en el tipo de datos encontrados en estos datasets y obtenemos la clasificación en las cuatro categorías tenidas en cuenta durante la datathon el resultado es que de los **39 comprobados, 26 contienen imágenes faciales, 20 características faciales, y 18 otros datos biométricos y patrones de movimiento**.

26

Imágenes faciales

20

Características faciales

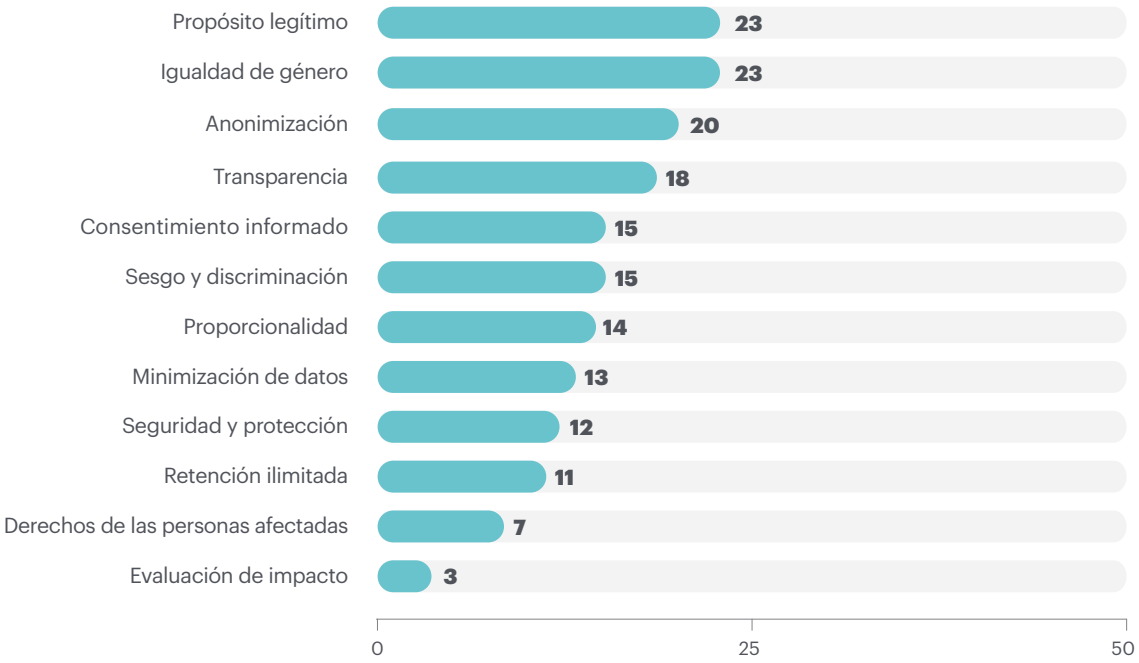
18

Patrones de movimiento

18

Otros datos biométricos

Y si miramos también más detenidamente el cumplimiento de estos conjuntos de datos, de los aspectos éticos y legales chequeados, **la igualdad de género y el propósito legítimo son los aspectos que más se cumplen con 23 datasets cada uno, 20 cumplen anonimización y 18 en transparencia**. El aspecto al que menos hacen referencia es la evaluación de impacto con solo 3 conjuntos de datos que lo mencionan explícitamente.



En cuanto al segundo listado, los «**NO COMPROBADOS**», una aclaración respecto de estos conjuntos de datos, que no se haya podido verificar ninguno de los aspectos evaluados, no quiere decir que no los cumplan, sencillamente, en la información referente al conjunto de datos no se ha encontrado ninguna referencia a ellos.

En estos conjuntos de datos **la versatilidad baja a un poco más del 35%** tanto en acceso como en tipología.

Acceso	Total	Versatilidad	Ética & legal
Registro	11	38,64%	0,00%
Abierto	6	29,17%	0,00%
Suma total	**17**	**35,29%**	**0,00%**

Tipología	Total	Versatilidad	Ética & legal
Primario	15	33,33%	0,00%
Relacionado	2	50,00%	0,00%
Suma total	**17**	**35,29%**	**0,00%**

Por último, en cuanto a los conjuntos de datos «NO COMPROBADOS», si miramos como se han clasificado, vemos que **10 contienen imágenes faciales, 6 características faciales, 5 patrones de movimiento y 3 otros datos biométricos.**

10

Imágenes faciales

6

Características faciales

5

Patrones de movimiento

3

Otros datos biométricos

4
Conclusiones y futuros pasos

El presente repositorio y el trabajo desarrollado durante la #DFSdatathon23 **sobre identidad digital, privacidad, protección de datos y derechos digitales**, organizada por Digital Future Society, una iniciativa conjunta de la **Secretaría de Estado de Digitalización e Inteligencia Artificial, Red.es** y **Mobile World Capital Barcelona** nos ofrece algunas pistas interesantes a tener en cuenta para el futuro.

Una de las áreas en las que se podría trabajar para mejorar la calidad y accesibilidad de los conjuntos de datos es **la versatilidad**. Un aumento en la tipología de los datos que contienen los datasets permitiría a los investigadores y desarrolladores trabajar con un mayor número de categorías y tipos de datos. Esto podría conducir a **una mayor innovación y avance en campos como el reconocimiento facial, la identificación biométrica y la privacidad**.

También es esencial **mejorar la accesibilidad de los conjuntos de datos**, ya que un mayor acceso a datos de calidad facilitaría el trabajo de los investigadores, las empresas y las organizaciones interesadas en desarrollar tecnologías y soluciones en el ámbito de la identidad digital, privacidad, protección de datos y derechos digitales. Esto incluye el **aumento de conjuntos de datos de acceso totalmente abierto, así como la reducción de barreras, como el registro o la petición de datos,** que pueden limitar el acceso a información valiosa.

Por lo tanto, la disponibilidad de conjuntos de datos abiertos debe mejorar para facilitar el acceso a datos de calidad. El mero hecho de tener que hacer un registro o una petición de datos, aunque estos sean gratuitos, ya supone una barrera al concepto de datos abiertos y, en consecuencia, a la accesibilidad de los mismos. De los datasets encontrados durante la jornada, **solo el 33% son de acceso totalmente abierto**, es decir, que no se pide ningún tipo de registro para poder obtener los datos.

Además, otra derivada que hemos obtenido con el trabajo realizado es que la accesibilidad a conjuntos de datos de calidad es difícil, sobre todo si queremos tener en cuenta aspectos éticos y legales. Parte del trabajo desarrollado durante la jornada se centró en **poder comprobar si los conjuntos de datos encontrados eran conformes a determinados aspectos propuestos**, como el consentimiento informado, el propósito legítimo, la igualdad de género o la proporcionalidad.

Para ello, el trabajo efectuado se centró en **revisar la información que acompañaba a los conjuntos de datos** para poder determinar si eran acordes a los factores sugeridos.

De los 39 conjuntos de datos comprobados analizados, la media de cumplimiento de estos aspectos es del 37.18%, lo que indica que aún hay un largo camino por recorrer en la mejora de la ética y legalidad en la creación y uso de datos.

Teniendo en cuenta estos resultados, es necesario que se fomente la creación de conjuntos de datos abiertos y accesibles que cumplan con los aspectos éticos y legales pertinentes. Esto permitirá un avance más rápido y responsable en el desarrollo de tecnologías relacionadas con la identidad digital, privacidad, protección de datos y derechos digitales.

Que en algunos casos no se haya podido verificar ni uno solo, conjuntos de datos «NO COMPROBADOS», de esos factores no quiere decir que no los cumplan, quiere decir que **no son lo suficientemente transparentes a la hora de compartir los datos**.

Por lo tanto, también se debe trabajar en la línea de incorporar en todos los conjuntos de datos información suficiente para dar la seguridad de que son de calidad y adecuados para poder desarrollar servicios y aplicaciones que sean totalmente respetuosas con los usuarios y acordes con la legislación vigente. Para lograr esto, se recomienda seguir ciertos pasos y acciones futuras:

1. **Fomentar la creación de guías y estándares** para la publicación y uso de conjuntos de datos abiertos, que incluyan aspectos éticos, legales y de accesibilidad, así como criterios de calidad y diversidad en la información proporcionada.

2. **Impulsar la colaboración entre entidades públicas, privadas y académicas** en la creación y mantenimiento de plataformas y repositorios de datos abiertos relacionados con la identidad digital, privacidad, protección de datos y derechos digitales.

3. **Desarrollar programas de formación y concienciación sobre la importancia de los datos** abiertos, la ética y la legalidad en la investigación y desarrollo de tecnologías relacionadas con la identidad digital, la privacidad, la protección de datos y los derechos digitales.

4. **Promover la transparencia y la responsabilidad en el uso de datos,** exigiendo a los proveedores de conjuntos de datos que compartan información clara y detallada sobre el cumplimiento de aspectos éticos, legales y de calidad.

5. **Establecer mecanismos de evaluación y certificación para conjuntos de datos,** que garanticen su adecuación en términos de ética, legalidad y calidad, facilitando así la identificación de aquellos conjuntos de datos que son adecuados para el desarrollo de tecnologías y soluciones respetuosas con los derechos de los usuarios y acorde con la legislación vigente.

6. **Fomentar la adopción de prácticas de "Privacy by Design" y "Privacy by Default"** en la creación y gestión de conjuntos de datos, asegurando que se adopten medidas para proteger la privacidad y seguridad de la información desde el inicio.

7. **Incentivar la investigación y el desarrollo de tecnologías y herramientas** que permitan garantizar la privacidad y protección de datos en la gestión y análisis de conjuntos de datos abiertos, como técnicas de anonimización, seudonimización y cifrado.

8 **Establecer políticas de gobernanza de datos** que promuevan la responsabilidad y la transparencia en el manejo de los conjuntos de datos, abordando cuestiones como la retención de datos, el acceso a datos y la eliminación de datos.

9 **Crear alianzas y colaboraciones internacionales** para compartir experiencias, conocimientos y mejores prácticas en la gestión de conjuntos de datos abiertos y éticos, impulsando así el desarrollo global de tecnologías y soluciones en el ámbito de la identidad digital, privacidad, protección de datos y derechos digitales.

10 **Promover la investigación y desarrollo de métodos y técnicas** que permitan medir el impacto social y económico de la adopción de conjuntos de datos abiertos y éticos en la creación y uso de tecnologías relacionadas con la identidad digital, privacidad, protección de datos y derechos digitales.

En conclusión, el acceso y empleo de conjuntos de datos abiertos, éticos y legales es fundamental para el progreso responsable en el ámbito de la identidad digital y derechos digitales.

Es esencial fomentar la creación de estos conjuntos de datos y generar colaboraciones entre diferentes entidades para asegurar su calidad y cumplimiento con los aspectos éticos y legales.

Además, es necesario promover la transparencia y la responsabilidad en la publicación y utilización de los datos, así como impulsar la formación y concienciación en estos temas. Solamente así podremos garantizar el desarrollo de tecnologías y soluciones que respeten los derechos de los usuarios y estén acorde a la legislación vigente, permitiendo así un avance sostenible y ético en el ámbito de la identidad digital, privacidad, protección de datos y derechos digitales.

5
Participantes de la Datathon

Todo este trabajo, el repositorio resultado de la datathon, ha sido posible gracias a la participación y colaboración de las siguientes personas:

Álvaro Lluís Rodríguez es analista de datos en Holaluz.

Ana Elizabeth Ledesma realiza visualización de datos en Médicos sin frontera

Ana Garzón es profesora en la Universidad Oberta de Catalunya.

Belen Arribas es abogada experta en Data Privacy.

Cristina Diez es Directora de contenidos en Qnari.

Didier Domínguez Herrera es Data Scientist en Fundació TIC Salut Social.

Ferran Deitx Roca es responsable de analítica de datos en la Agència de Ciberseguretat de Catalunya.

Gemma Romero Miguel es ingeniera Industrial.

Gerard Guarin es estudiante en la UAB.

Laura Martín González es analista de datos.

Pablo Cerralbo es analista de datos en Holaluz.

Paula Ruiz es ingeniera de datos.

Pol Colomer Campoy es ingeniero informático.

A todas ellas, en nombre de Digital Future Society, una iniciativa conjunta de la Secretaría de Estado de Digitalización e Inteligencia Artificial, Red.es y Mobile World Capital Barcelona, queremos expresar nuestro más sincero agradecimiento por haber colaborado y participado en la Datathon #DFSdatathon23 sobre identidad digital, privacidad, protección de datos y derechos digitales.

Su dedicación, entusiasmo y compromiso han sido fundamentales para el éxito de este evento. Su participación activa, ideas innovadoras y experiencia en estos temas tan relevantes han enriquecido enormemente las discusiones y los resultados obtenidos durante la Datathon.

También queremos reconocer y agradecer a todas las personas colaboradoras y expertas y que generosamente compartieron su tiempo, conocimientos y perspectivas para que el evento llegara a buen puerto. Su guía y apoyo han sido importantes para orientar a los equipos en las tareas encomendadas.

Esperamos que esta Datathon sea solo el comienzo de una larga trayectoria de colaboración y progreso en la construcción de un futuro digital más seguro, inclusivo y ético.

Nuevamente, nuestro más sincero agradecimiento a todas las personas involucradas en la Datathon #DFSdatathon23.

Digital Future Society

Digital Future Society

Un programa de

GOBIERNO DE ESPAÑA
VICEPRESIDENCIA PRIMERA DE GOBIERNO
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es

MOBILE WORLD CAPITAL BARCELONA