CONTESTA TRASLADO

Sra. Jueza,

La Fundación para la Difusión del Conocimiento y el Desarrollo Sustentable "Vía Libre" (en adelante "Fundación Vía Libre"), representado en este acto por María Beatriz Busaniche, en calidad Presidenta de la mencionada institución, con el patrocinio letrado de Margarita Trovato (T° 133 F° 719 CPACF), manteniendo el domicilio legal constituido y domicilio electrónico en xxxxxxxxxx, en la causa caratulada "Observatorio de Derecho Informático Argentino O.D.I.A. y Otros c/ G.C.B.A. s/ amparo – Otros", Expte. N° 182908/2020-0, a V.S. respetuosamente digo:

I. Objeto

Por la presente venimos a **contestar el traslado conferido** en nuestro carácter de amicus curiae de la resolución de V.S. de fecha 09/04/2024, notificada a esta parte el 10/04, donde se corre traslado de las presentaciones de la Defensoría del Pueblo de la Ciudad de Buenos Aires y de la Procuración General de la Ciudad de Buenos Aires (actuaciones 07009/2024 y 650654/2024 respectivamente. Ambas son concordantes en la idoneidad de la Facultad de Ciencias Exactas de la Universidad de Buenos Aires para la auditoría ordenada sobre el sistema.

En este punto, dada la *expertise* con la que la Fundación Vía Libre se presentó en este expediente como *amicus curiae*, y con el espíritu de aportar conocimientos que puedan asegurar el mejor proceso de auditoría posible en términos de derechos humanos, es que venimos en esta oportunidad a ofrecer algunos conceptos y lineamientos técnicos mínimos para llevarlo adelante, que desarrollaremos a continuación.

i. <u>Sobre la precisión de las tecnologías de reconocimiento facial. Factores que deben tenerse en cuenta.</u>

La **precisión** de las tecnologías de reconocimiento facial (TRF) difiere sustancialmente entre un sistema y otro. Estas diferencias surgen esencialmente de **tres factores: (a) la calidad de los conjuntos de datos empleados para el entrenamiento, (b) la capacidad del algoritmo en sí, y (c) el contexto de aplicación.**

La importancia del **primer factor** (conjunto de datos de construcción del modelo) se basa en que un número insuficiente de imágenes, o un conjunto

sesgado en cuanto a la composición fenotípica y etaria de las personas representadas respecto de las que se presentarán en la situación real, conduce a significativas disparidades demográficas, **penalizando de esa forma a grupos humanos que por otra parte resultan generalmente pertenecer a los sectores sociales más vulnerables**. El Instituto Nacional de Normas y Tecnología (NIST) de los Estados Unidos lleva a cabo una evaluación continuada de estas disparidades demográficas mediante la realización de pruebas comparativas bajo condiciones idénticas, en el marco del proyecto FRTE (ex-FRVT)[1][2]. A priori es posible extraer de los resultados de evaluación tres conclusiones de orden general: todos los algoritmos testeados (518 desde que se inició el proyecto) presentan sesgo, los sesgos son significativos, y existen enormes diferencias entre algoritmos. Aún el algoritmo con menor sesgo presenta una diferencia de 24 veces entre la clase más precisa (hombres de origen centroamericano entre 50 y 65 años)[3] y la menos precisa (mujeres de origen africano entre 65 y 99 años); el algoritmo con mayor sesgo penaliza con un error 7784 veces mayor a las mujeres africanas entre 65 y 99 años respecto de los hombres europeos entre 20 y 35. Un indicador práctico de esta distorsión es el coeficiente de Gini (que vale 0 para una distribución totalmente uniforme y 1 para una concentrada totalmente en un punto): para el mejor algoritmo arroja un valor de 0.38, mientras que para el peor resulta 0.87.

A la vez, **las diferencias demográficas pueden estar determinadas no solamente por la calidad de los datos, que es el factor de mayor importancia, sino también por la precisión del algoritmo en sí**, es decir, por cuánto y cómo "aprenda" a partir de los datos con que fue entrenado y los que le son suministrados en la operación real. Este es entonces el segundo factor que incide en la precisión de la tecnología. No existen evaluaciones públicas e independientes respecto de estos criterios cualitativos, pero es una buena guía saber que en condiciones ideales (comparar dos fotografías frontales de muy buena calidad en un proceso 1:1, también llamado de verificación) existen diferencias abismales entre algoritmos: tasas de error de falsos negativos, para falsos positivos ajustados a 0.00001, entre 0.0006 (o 0.06 %) y 0.9998 (o 99.98 %) entre el mejor y el peor de 550 analizados respectivamente.

---

[1] P. Grother, M. Ngan y K. Hanaoka, *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects*, Interagency Report NISTIR 8280 Draft Supplement, diciembre 2019. Gaithersburg, MD: National Institute of Standards and Technology.

[2] P. Grother, *Face Recognition Vendor Test (FRVT) Part 8: Summarizing Demographic Differentials*, Interagency Report NISTIR 8429, julio 2022. Gaithersburg, MD: National Institute of Standards and Technology.

[3] Este comportamiento del algoritmo *idemia_09*, de origen francés, es ligeramente anómalo pero no único (otros algoritmos de origen chino tienen como grupo más preciso a los hombres del este de Asia). La abrumadora mayoría de los algoritmos tiene como grupo más preciso a los hombres blancos europeos.

En tercer lugar, en cuanto a las **condiciones de contexto**, resulta necesario destacar que **el sistema bajo investigación presenta el mayor grado de complejidad relativa en el campo de utilización de las TRF**: un proceso de identificación (comparación de una muestra obtenida contra una galería de imágenes, 1:N) en que **la muestra es de baja calidad y se obtiene en condiciones no colaborativas**. Las imágenes obtenidas en el espacio público son sub-estándar, porque provienen de cámaras de videovigilancia; entre sus problemas se encuentran alteraciones de foco, sobre o subexposición, ángulos indeterminados, colocaciones no frontales de la cámara, oclusiones, distancia, detección de múltiples rostros, etc. **Por otro lado, la galería está conformada, a nuestro saber, por una sola imagen de buena calidad de cada sujeto; ello impide que el algoritmo pueda "aprender" variaciones posicionales.**

Las únicas pruebas de comparación exhaustivas y públicas de sistemas en estas condiciones, efectuadas en idénticas condiciones para todos los participantes, fueron llevadas a cabo por el subproyecto FIVE (Face in Video Evaluation) del proyecto FRVT del NIST antes mencionado. Una nueva ronda de pruebas se planea para el año en curso con conjuntos de datos y casos de uso adicionales: al aire libre con iluminación direccional, a larga distancia y potencialmente afectado por turbulencia atmosférica, desde plataformas elevadas, y con múltiples personas en escena. Los resultados que en el futuro se obtengan de estas pruebas serán una excelente guía para usos como el que aquí estudiamos, pero desafortunadamente aún no están disponibles. No obstante, es posible extraer datos de interés de la ronda previa de pruebas, documentada en Grother et al., 2017,[4] en particular para el contexto de prueba en los pasillos de un estadio deportivo cerrado, bien iluminado y con 11 cámaras (Dataset P). Estos datos indican claramente que existen diferencias de órdenes de magnitud entre distintos sistemas de reconocimiento en cuanto a la capacidad de detección de rostros (entre 206 por minuto para el mejor caso y 11 para el peor), proporción de falsos negativos (12.9 % para el mejor y 99.1 % para el peor, relativos a una cámara cercana ubicada a 1.83 m de altura — los resultados para otras posiciones de cámara son siempre peores), y cantidad de falsos positivos en el total de rostros detectados (entre 1 cada 1200 rostros detectados y 1 cada 50, respectivamente, para la mejor cámara).

ii. <u>Sobre la evaluación que debe llevarse a cabo</u>

[4] P. Grother, G. Quinn y M. Ngan, *Face In Video Evaluation (FIVE) Face Recognition of Non-Cooperative Subject*, Interagency Report NISTIR 8173, marzo 2017. Gaithersburg, MD: National Institute of Standards and Technology.

Siendo este el escenario, a falta de evaluaciones fiables e independientes para el sistema bajo análisis, surge como alternativa llevar a cabo una propia. Debe tenerse en cuenta muy especialmente que la realización de estas pruebas requiere un gran volumen de trabajo y costos significativos, con el número suficiente de iteraciones para obtener resultados consistentes. La evaluación se facilitará -y se aumentará la precisión- si se cuenta **con la cooperación del proveedor** en el suministro de dos insumos básicos para la investigación, a saber:

a) Conjuntos de datos: El requisito más aceptable en este sentido es que elproveedor proporcione una copia de su conjunto de datos (entrenamiento, testeo y validación); si por alguna razón se hallara que no es posible, deberá suministrar como mínimo una muestra al azar estadísticamente representativa (intervalo de confianza 0.95, margen de error menor a 1 %) y la documentación asociada. Adicionalmente deberá proveer la siguiente información sobre los conjuntos de datos empleados en el entrenamiento previo del sistema, si se tratara de un conjunto no disponible de manera pública y general: el número de imágenes; el número de personas a quienes corresponden esas imágenes; la mediana, la media aritmética, la media geométrica y el desvío estándar de imágenes por persona; la distribución por origen étnico y grupos de edad.

b) Algoritmo: El requisito más aceptable en este sentido es que el proveedor proporcione una copia del código fuente y de todas las bibliotecas asociadas, más las instrucciones de compilación y configuración. De no resultar posible, deberá autorizar (si resultara pertinente, bajo acuerdo de confidencialidad) la realización de pruebas de ingeniería inversa sobre el código ejecutable. Proveerá además la siguiente información sobre el algoritmo: técnica empleada (por ejemplo, redes neuronales convolucionales profundas); desarrollador; pruebas independientes realizadas, tipo de pruebas y su resultado (por ejemplo, las del proyecto FRTE (antes FRVT) del National Institute of Standards and Technology); descripción general de funcionamiento; tiempo de ejecución para incorporar una galería de 40000 imágenes; estructura de búsqueda en la galería; mecanismo de comparación entre muestra y galería (1:1 secuencial / n:1 / n;n / otros (describir, en su caso));tiempo de ejecución (mediana y 95 percentil) para generar una plantilla (template); longitud de plantilla en bytes; tiempo de búsqueda de similitudes en una galería.

Sin estos insumos una auditoría sería incompleta y no cumpliría el fin para el que se está ordenando: verificar su funcionamiento en el caso concreto, y no en forma de "caja negra".

En el mismo sentido, cabe destacar que también debería realizarse una **evaluación de seguridad considerando la criticidad del sistema y los perjuicios**

graves que causaría su uso indebido. Una guía inicial para esta actividad puede extraerse de la taxonomía formulada por Le Roux et al.[5]

II. Petitorio

Por los motivos expuestos, solicitamos se tenga por contestado el traslado conferido y se tengan en cuenta los lineamientos mínimos ofrecidos a la hora de delinear la auditoría correspondiente.

Proveer de conformidad,

ES DERECHO

[5] Q. Le Roux, E. Bourbao, Y. Teglia y K. Kallas, "A Comprehensive Survey on Backdoor Attacks and Their Defenses in Face Recognition Systems," en *IEEE Access*, vol. 12, pp. 47433-47468, 2024, doi: 10.1109/ACCESS.2024.3382584.