

# Utilización de la Inteligencia Artificial en la Salud:

## Lecciones aprendidas al enfrentar el brote de COVID-19

Por Edson Amaro Jr<sup>1</sup>, Michel Fornaciali<sup>2</sup>, Andre Batista<sup>3</sup>, Murilo Gazzola<sup>4</sup>, Livia Paiva da Silva<sup>5</sup>, Diogo F. C. Patrão<sup>6</sup> y Marcos Freitas Jr<sup>7</sup>

En el contexto de la pandemia de COVID-19, se ha hecho más evidente la necesidad de interacción entre los profesionales de la salud y las tecnologías de análisis de datos. Las aplicaciones directas incluyen los escenarios de atención (ayudando a asignar puntuaciones de riesgo a los pacientes en relación con el pronóstico y mejorando las decisiones de atención médica), a la planificación a corto plazo (organizando equipos y fuentes de atención) y a la evaluación de políticas públicas y acciones macro-regionales a largo plazo (aplicaciones de modelos epidemiológicos y simulaciones).

Cada uno de estos escenarios requiere enfoques específicos. En la asistencia sanitaria, las soluciones son fundamentales para los profesionales de la primera línea y se aplican cuando los pacientes están en la atención primaria, secunda-

ria o terciaria<sup>8</sup>. En la planificación a corto plazo, se necesitan herramientas para la gestión de los equipos o de la maquinaria que puedan encontrarse operativas en un máximo de dos semanas, aplicadas a la gestión de hospitales, ciudades y estados o provincias. Para la planificación a largo plazo, es necesario elaborar estrategias para microrregiones y macro-regiones basadas en escenarios ajustables, no sólo a nivel local, sino también a escala nacional. En todos estos frentes, el objetivo es utilizar la tecnología para prevenir las muertes evitables y las derivadas de las complicaciones. Esto es un reto, sobre todo por la falta de recursos; sin embargo, en un contexto como éste, las tecnologías de análisis de datos pueden ayudar a hacer más eficientes los sistemas sanitarios, especialmente el trabajo de los equipos de salud.

<sup>1</sup> Neurorradiólogo y superintendente de Ciencia de Datos y Analytics - Big Data del Hospital Israelita Albert Einstein (HIAE) y profesor de la Universidad de São Paulo (USP).

<sup>2</sup> Doctor y magíster en Ingeniería Computacional por la Universidad Estatal de Campinas (Unicamp) y licenciado en Ciencias Computacionales por la Unicamp. Es científico de datos en el equipo de Big Data Analytics del HIAE.

<sup>3</sup> Doctor en Ingeniería Computacional por la Escuela Politécnica de la USP y miembro del equipo de Big Data Analytics del HIAE.

<sup>4</sup> Estudiante de doctorado, tiene una maestría en Ciencias Computacionales y Matemáticas Computacionales de la USP. Es miembro del Centro Interinstitucional de Lingüística Computacional de la USP y científico de datos en el equipo de Big Data Analytics de HIAE.

<sup>5</sup> Estudiante de posgrado en Gestión de la Salud en el HIAE, es posgraduada en Gestión de Proyectos e Ingeniería Eléctrica con énfasis en Telecomunicaciones por la Universidad Presbiteriana Mackenzie. Es Tecnóloga en Telecomunicaciones por la Unicamp y Técnica en Análisis y Procesamiento de Sistemas y Datos por la Escuela Técnica Federal de São Paulo.

<sup>6</sup> Doctor en Oncología por la Fundación Antônio Prudente (FAP/A.C. Camargo Cancer Center) y licenciado en Física por la USP. Es científico de datos en el equipo de Big Data Analytics de HIAE.

<sup>7</sup> Magíster en Sistemas de Información por la USP y licenciado en Análisis y Desarrollo de Sistemas por la Facultad de Tecnología Termomecánica (FTT). Es científico de datos en el equipo de Big Data Analytics de HIAE.

<sup>8</sup> La atención primaria se centra en acciones relacionadas con la reducción del riesgo de enfermedades y la protección de la salud. La atención secundaria la prestan los hospitales y ambulatorios encargados de ofrecer a la población tratamientos de pediatría, cardiología y otras especialidades médicas, además de garantizar la estructuración de los servicios de urgencia y emergencia. Por último, la atención terciaria ofrece servicios de alta complejidad, representados por grandes hospitales y clínicas especializadas.

Si los modelos de IA suelen depender de un gran volumen de datos, ¿cómo hacer frente a la escasez de datos en la fase inicial de una pandemia?

### Experiencia en análisis de Big Data en el Hospital Israelita Albert Einstein

La organización de equipos dedicados al estudio, discusión y desarrollo de herramientas de apoyo basadas en el análisis de datos y, en particular, de algoritmos de Inteligencia Artificial (IA), no es tarea fácil. El reto es aún mayor cuando se realiza en contextos de trabajo remotos, para cumplir con las medidas de distanciamiento social impuestas para frenar el nuevo coronavirus.

El grupo de Big Data Analytics del Hospital Israelita Albert Einstein (HIAE) pasó por esta experiencia única. Situado en la ciudad de São Paulo, este hospital trató el primer caso de la enfermedad en Brasil. La implementación de soluciones de IA en las rutinas de los hospitales presentaba muchos desafíos, en particular la siguiente cuestión paradójica: si los modelos de IA suelen depender de un gran volumen de datos, ¿cómo hacer frente a la escasez de datos en la fase inicial de una pandemia? Los aspectos prácticos de la formación de un equipo para abordar los aspectos específicos de esta situación, en la que no hay suficientes datos o éstos carecen de calidad, pueden aportar ideas para los periodos postpandémicos.

Basándose en un modelo ágil de gestión de proyectos (Hidalgo, 2019), el equipo multifuncional de Big Data Analytics del HIAE se subdividió en “escuadras” (equipos de profesionales dedicados a tareas o proyectos específicos) con perfiles según tipos específicos de experiencia. Se designó a una persona como punto focal para la intercomunicación dentro del grupo, y como portavoz en los foros para definir el alcance, la estrategia, la gobernanza y la presentación de la evolución de las soluciones para las áreas que las utilizarían.

### Iniciativa para involucrar a la comunidad abierta: AntennaCovid

La participación de la comunidad que se ocupa del procesamiento y análisis de datos es crucial en el contexto de una pandemia. Por ello, es importante crear espacios que incluyan varios frentes, como investigadores, gestores y científicos (Marston et al., 2020). La gestión de estos espacios participativos requiere encontrar personas con áreas de formación variadas, con sensibilidad para reconocer y utilizar los diferentes conocimientos y experiencias de los individuos, unidos en torno al objetivo de crear soluciones de apoyo para los profesionales de la salud.

¿Cómo es posible mantener una coproducción constructiva en el contexto de una pandemia, con la presión de generar resultados prácticos en poco tiempo? En este contexto surgió AntennaCovid, un foro de interacción entre profesionales de la salud, científicos de datos, ingenieros de *machine learning* y matemáticos, que trabajan juntos para combatir la COVID-19. Apoyado por la Red Nacional de Educación e Investigación y programado para existir mientras sea necesario, AntennaCovid fue dividido en varias plataformas para permitir a los participantes trabajar y compartir sus proyectos – como GitHub ([github.com/antennaCovid](https://github.com/antennaCovid)), para compartir códigos de código abierto; Slack ([antennaCovid.slack.com](https://antennaCovid.slack.com)), para facilitar la comunicación, integrada con la codificación; y la página web ([antennaCovid.org](https://antennaCovid.org)), para la difusión al público.

Además, se celebran reuniones virtuales semanales para que los miembros compartan sus preocupaciones, ideas y trabajo. También hay rondas

(*roundups*), programadas según sea necesario para discutir el desarrollo de los proyectos y la presentación de los resultados prácticos. Es importante destacar que todo se difunde abiertamente en GitHub.

Por lo tanto, AntennaCovid gestiona un espacio participativo abierto a la comunidad para aumentar el conocimiento a través del trabajo en red, para que todos puedan plantear sus objetivos, controlar su desarrollo e interactuar con equipos de diferentes instituciones. Cabe destacar que en Brasil se han creado varias iniciativas similares, algunas de las cuales interactúan directamente con AntennaCovid, lo que pone de manifiesto el gran potencial que tiene el uso de herramientas colaborativas en red en el país.

## Soluciones predictivas durante la atención: D+0<sup>9</sup>

Las interacciones entre los profesionales de la salud y las tecnologías de análisis de datos y de *machine learning* durante la prestación de los cuidados son un terreno fértil para el desarrollo de modelos predictivos. Es posible crear soluciones que asignen puntuaciones de riesgo a los pacientes en relación con el pronóstico y mejoren las decisiones relativas a la necesidad y la intensidad de la atención médica – contribuciones esenciales para enfrentar la COVID-19, ya que la identificación temprana de los pacientes de alto riesgo puede, por ejemplo, reducir el uso de la ventilación mecánica invasiva (Sun et al., 2020).

En el caso de nuevas enfermedades, como la COVID-19, los algoritmos son capaces de aprender patrones referentes a las interacciones de las características de los pacientes que conducen tanto a un mayor riesgo de un diagnóstico positivo de la enfermedad (Batista et al., 2020) como a la evolución de la condición o incluso a la muerte, ayudando así a los equipos a tomar decisiones clínicas y a asignar recursos físicos para los casos más graves (Alimadadi et al., 2020)<sup>10</sup>. Por lo tanto, el desarrollo de modelos de predicción del pronóstico puede servir de apoyo en las rutinas médicas para el triaje de los pacientes y la determinación del tratamiento, entre otros<sup>11</sup>.

En una revisión sistemática de publicaciones en el área de la salud, Wynants et al. (2020) evaluaron los procesos de construcción, evaluación y difusión científica de 66 modelos de predicción de COVID-19. Un número importante de modelos ha sido diseñado para predecir la enfermedad a través de imágenes. Los autores también identificaron como predictores más utilizados para diagnosticar la COVID-19 factores como: la edad, el sexo, la temperatura corporal, los signos vitales, los síntomas, la presión arterial y la creatinina. Entre los modelos de pronóstico también se han destacado los resultados de muerte, deterioro clínico y duración de la estancia del paciente.

Las interacciones entre los profesionales de la salud y las tecnologías de análisis de datos y de *machine learning* durante la prestación de los cuidados son un terreno fértil para el desarrollo de modelos predictivos.

<sup>9</sup> Soluciones para la actividad en el mismo día - "D+0".

<sup>10</sup> Más concretamente, Batista et al. (2020) presentaron un modelo predictivo para la COVID-19 basado en información como la edad y el sexo, así como en análisis de sangre básicos. Se construyeron y evaluaron cinco modelos predictivos supervisados, donde el mejor (*support vector machine*) demostró que el uso de pruebas de laboratorio es prometedor para diagnosticar el nuevo coronavirus. Esta conclusión fue corroborada por estudios posteriores en instituciones de otros países (véase, en particular, Gao et al., 2020 y Brinati et al., 2020). La importancia de construir modelos predictivos para el pronóstico de COVID-19 también se confirma en un estudio de Hirsch et al. (2020) sobre la incidencia de la insuficiencia renal aguda en pacientes con la enfermedad.

<sup>11</sup> También en este sentido, los esfuerzos se han centrado en la predicción de la insuficiencia renal aguda en pacientes hospitalizados. Los modelos se construyen a partir de arquitecturas computacionales de redes neuronales recurrentes, en particular para la memoria a corto plazo (LSTM), discutidas en Swapnarekha et al. (2020).

Tanto la calidad de los datos como sus condiciones de recogida y su uso en las rutinas asistenciales son esenciales para que las soluciones analíticas aporten valor a las prácticas de los profesionales de la salud.

También se han realizado estudios dedicados a la construcción de herramientas en línea para la predicción de las puntuaciones de riesgo de la COVID-19 (DeCaprio et al., s.f.) y al desarrollo de aplicaciones móviles para la predicción diagnóstica de la enfermedad a partir de la inserción de los resultados de las pruebas de laboratorio y los síntomas de los pacientes (Meng et al., s.f.). Estos estudios han reforzado el importante concepto de compartir datos en la ciencia (*eScience* y *open data science*). Otros autores han compartido sus códigos fuente, permitiendo a la comunidad científica conocer las especificaciones técnicas de los modelos construidos.

A pesar de los esfuerzos por difundir los resultados y los modelos predictivos resultantes, es importante tener en cuenta, según Wynants et al., que estos modelos necesitan pasar por un proceso de validación externa antes de ser incorporados a las prácticas médicas (Cosgriff et al., 2019). Tanto la calidad de los datos como sus condiciones de recogida y su uso en las rutinas asistenciales son esenciales para que las soluciones analíticas aporten valor a las prácticas de los profesionales de la salud.

Los avances en el uso de las técnicas de *machine learning*, y la abundancia de publicaciones científicas que han presentado resultados optimistas sobre cómo los modelos predictivos pueden contribuir a la profesión médica, abren un espacio de reflexión y llaman la atención sobre la importancia de las buenas prácticas en la difusión de estos modelos. La adopción de estas prácticas permite el análisis crítico por parte de los pares y el detalle de las premisas, los procesos de selección de muestras y las transformaciones aplicadas a los datos, así como las consideraciones que pueden mejorar la comprensión de la comunidad científica y optimizar el trabajo de validación externa<sup>12</sup>.

## Soluciones predictivas para la gestión de los servicios: D+15

Las predicciones a corto plazo – hasta 15 días, o D+15 – han asumido papeles diferentes entre los contextos y las etapas de progresión de la COVID-19. Al principio de la pandemia, se observó un crecimiento exponencial, no sólo en Brasil, del número de personas infectadas (Remuzzi & Remuzzi, 2020). En ese momento, era esencial entender cuándo habría un cambio en la pendiente de la curva, especialmente para planificar los recursos físicos y humanos necesarios para proporcionar la mejor atención.

En esta etapa se utilizaron modelos de predicción con técnicas tradicionales que indicaban un crecimiento lineal y/o exponencial (Martínez et al., 2020). Los países donde la propagación del virus estaba más avanzada fueron *proxys* importantes para entender el comportamiento de las curvas (Barmparis & Tsiornis, 2020), permitiendo incluso la proyección inicial de la evolución de los casos en Brasil a través de métodos de agrupamiento. Aunque las proyecciones a mediano y largo plazo han sido más solicitadas por los gestores, principalmente por el mayor impacto de las decisiones urgentes, los modelos a corto plazo con un enfoque matemático simplificado han seguido siendo relevantes. Esto se debe a que – aunque parezca contradictorio – los modelos matemáticos más

---

<sup>12</sup> La directriz *Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD)*, por ejemplo, fue uno de los primeros instrumentos para verificar los requisitos mínimos para la publicación y difusión de modelos predictivos (Collins et al., 2015), y sigue siendo bastante útil.



complejos no son necesariamente más fiables y no siempre dan los mejores resultados (Roda et al., 2020).

Por lo tanto, abordar la pluralidad y las limitaciones de los modelos predictivos presentados (Roda et al., 2020) requiere comprender los diferentes contextos, así como los retos y resultados de varios grupos de investigadores en todo el mundo. Algunos ejemplos de escenarios típicos son la escasez de datos (común al comienzo de una pandemia), la inconsistencia en el proceso de recolección y la falta de parámetros bien establecidos. Otros son la adopción de criterios variados para las pruebas de los individuos infectados, las decisiones gubernamentales contradictorias y, específicamente en Brasil, las diferencias socioeconómicas entre las regiones, que limitan la posibilidad de generalizar los resultados y se reflejan en la calidad de los registros de datos<sup>13</sup>.

Los intensos intercambios que se produjeron entre grupos de investigadores a raíz de la pandemia permitieron aprender que el desarrollo de soluciones predictivas – aunque sea a corto plazo, con un impacto positivo para los gestores – implica algo más que enfoques matemáticos y requiere una comprensión de los contextos y los límites de uso. Comprender los retos actuales es esencial para interpretar con precisión lo que dicen las predicciones y para reducir los riesgos cuando se utilizan para apoyar las decisiones, especialmente en el caso de las políticas públicas en el sector sanitario.

## Soluciones a largo plazo: Modelos D+30 y ciclos epidemiológicos

Aunque implican cierta variabilidad, los modelos de predicción a largo plazo son importantes, no sólo como instrumentos para obtener estadísticas de los meses futuros, sino especialmente para entender cómo las intervenciones no farmacológicas, las decisiones gubernamentales y la dinámica natural de las macro-regiones se reflejan en el comportamiento de las curvas de predicción a lo largo del tiempo. Entender de forma heurística (por construcción) que existe un conjunto de proyecciones que indican tendencias con una pendiente más acentuada puede aportar conocimientos relevantes en cuanto a la eficacia de las acciones.

La epidemiología es, por definición, el estudio en términos cuantitativos de la distribución de los fenómenos de salud y de sus factores determinantes en la población. Por ello, es la ciencia que más utiliza los modelos de predicción a largo plazo. En el caso de la COVID-19, su uso se centró en simulaciones de la propagación del virus en 60 días o más, así como en la identificación y caracterización de los grupos de riesgo, acciones cruciales para una mejor gestión de la pandemia en aras de la salud pública.

## Simulaciones a largo plazo

Toda epidemia es la manifestación de una enfermedad caracterizada por su transmisión rápida, que infecta simultáneamente a muchas personas en el

La epidemiología es, por definición, el estudio en términos cuantitativos de la distribución de los fenómenos de salud y de sus factores determinantes en la población. Por ello, es la ciencia que más utiliza los modelos de predicción a largo plazo.

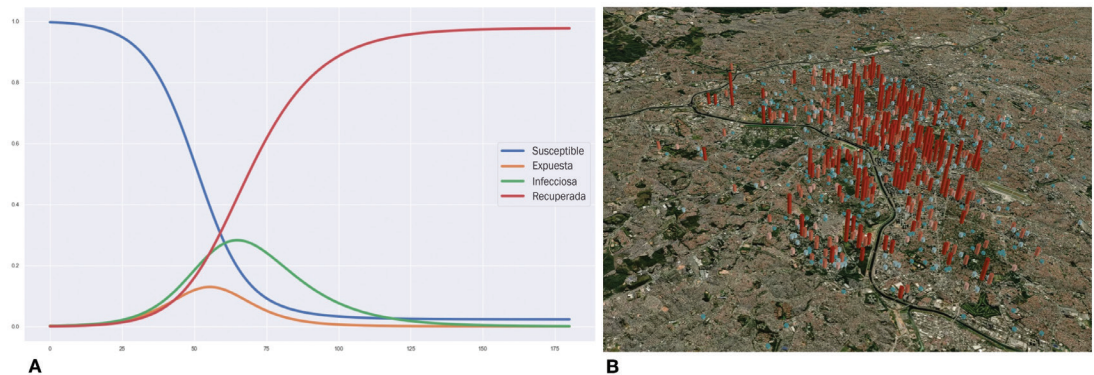
---

<sup>13</sup> Los retrasos en la notificación de las hospitalizaciones y las defunciones, por ejemplo, son un problema al que se enfrentan muchos países y que repercute en el uso de técnicas de actualización de datos (*nowcasting*), destinadas a mejorar la precisión de las soluciones predictivas desarrolladas (Altmejd et al., 2020; Schneble et al., 2020; Puca & Buonanno, s.f.).

mismo lugar. Entender el ciclo epidemiológico específico, desde el inicio del brote hasta su extinción, es una tarea compleja, especialmente en el caso de las nuevas enfermedades. Por ello, se han utilizado modelos que describen la transmisión entre individuos de forma simplificada.

Entre la amplia gama de modelos epidemiológicos existentes (Britton, 2009; Chowell et al., 2016), los autores del presente artículo se centraron en modelos compartimentados deterministas. Se adoptó el modelo SEIR, que caracteriza a una población en una epidemia en cuatro etapas: susceptible, expuesta, infectada y recuperada (que incluye las muertes). La Figura 1 muestra la evolución teórica de la enfermedad e indica el porcentaje de la población en cada etapa a lo largo del tiempo.

**Figura 1 - REPRESENTACIÓN TEÓRICA DE UNA PANDEMIA UTILIZANDO EL MODELO SEIR (A) Y REPRESENTACIÓN DE LOS TESTEOS PARA COVID-19 EN SÃO PAULO EL 12 DE MARZO DE 2020 (B)**



Fuente: Elaboración propia.

Como toda simplificación matemática, el modelo SEIR tiene limitaciones. Dos de las más notables son la sobreestimación de los casos y la dependencia de fuertes premisas relacionadas con la probabilidad uniforme de que las personas se infecten. En el caso de la COVID-19, hay que tener en cuenta además que se trata de una enfermedad desconocida, con pocos estudios publicados y un contexto muy dinámico para su comprensión. Estas complicaciones impiden aplicar el modelo para determinar con precisión cuándo se producirá el fenómeno de interés (el pico de la curva de infecciones y el final de la epidemia, por ejemplo) y cuál será su magnitud (cuántos casos habrá en el pico de la curva y cuántas muertes habrá a los 90 días).

Para reducir las limitaciones técnicas y contextuales, se introdujeron dos innovaciones en los modelos: trabajar con múltiples escenarios para identificar incertidumbres; y considerar acciones reales de mitigación de la COVID-19 para aumentar la asertividad de las curvas. En primer lugar, se centró la atención en la curva de infección, ya que de ella se derivan las cifras de nuevos casos, hospitalizaciones y muertes. Al modelar esta curva, se incorporó el concepto de incertidumbre a través de una ilustración visual de varios escenarios a lo largo del tiempo, representando las fluctuaciones entre contextos más optimistas y más pesimistas. Este enfoque hace que los usuarios sean conscientes de que los modelos serán naturalmente inexactos, evitando así una interpretación literal de las cifras presentadas en las simulaciones. Al comparar las diferencias entre los escenarios del mismo modelo, se pueden inferir los impactos en los

mejores y peores casos, lo que se ha abordado con éxito en otros estudios sobre el nuevo coronavirus (López y Rodo, n.d.; Prem et al., 2020).

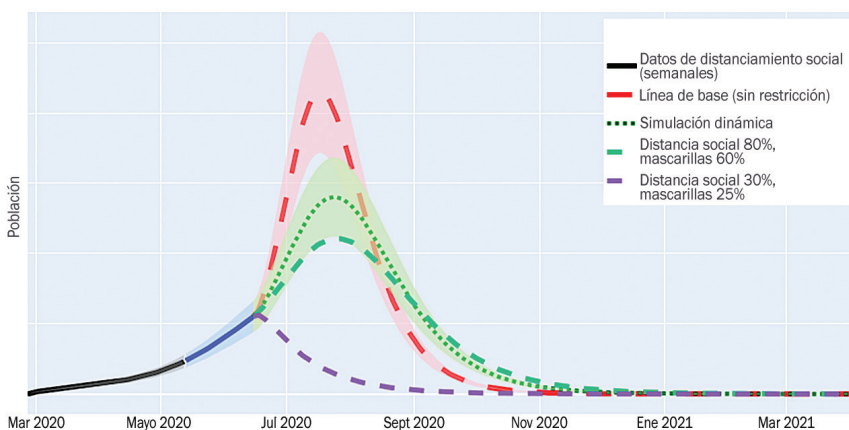
El modelo SEIR asume que toda la población puede ser infectada y que la transmisión de una enfermedad ocurre libremente, sin acciones de mitigación, lo que lleva a sobreestimar el número de casos. Para mitigar este fenómeno, se incorporaron al modelo iniciativas para controlar la propagación del virus. Dado que no existen vacunas ni fármacos específicos para la COVID-19, estas medidas fueron en su mayoría no farmacológicas (Lai et al., 2020), como la higiene personal y el distanciamiento social.

Uno de los principales parámetros de la modelización es el número de reproducción (RO), que determina el potencial de transmisión de una enfermedad, es decir, cuántos casos nuevos produce una persona infectada. En general el RO se trata como una tasa única a lo largo del tiempo. Sin embargo, durante una pandemia, se aplica una serie de intervenciones para mitigar la transmisión del virus. Los modelos del presente estudio trataron de estimar matemáticamente en qué medida cada intervención y la adhesión a la misma por parte de la población afectan al RO, creando un RO comparable a través de escenarios.

Dado que las acciones con mayor impacto en la mitigación de la COVID-19 son el distanciamiento social (Hellewell et al., 2020) y el uso de mascarillas (MacIntyre & Chughtai, 2020), se implementaron estas medidas para ajustar el RO<sup>14</sup>. El resultado fue una simulación más realista, que pone de manifiesto la incertidumbre del enfoque. Los usuarios pueden navegar por el simulador, creando situaciones de adherencia a intervenciones no farmacológicas y comparando sus efectos a largo plazo. Por ejemplo, en la Figura 2 se comparan cuatro escenarios con diferentes tasas de adhesión de la población al distanciamiento social y al uso de mascarillas, así como una simulación actual y un escenario sin ninguna intervención (línea de base) – para estas dos últimas categorías, las zonas sombreadas ilustran la incertidumbre.

Los modelos del presente estudio trataron de estimar matemáticamente en qué medida cada intervención y la adhesión a la misma por parte de la población afectan al RO [número de reproducción], creando un RO comparable a través de escenarios.

**Figura 2 - HERRAMIENTA RESULTANTE DE UNA SIMULACIÓN DE LA COVID-19 A LARGO PLAZO**



Fuente: Elaboración propia.

<sup>14</sup> Los índices de distanciamiento social proceden de estimaciones de movilidad basadas en la triangulación de antenas de telefonía móvil. El uso de mascarillas no se mide oficialmente, por lo que se utiliza una estimación empírica del porcentaje de población que se adhiere a la medida. También se definió un porcentaje de eficacia de las mascarillas, teniendo en cuenta la calidad del producto y su uso incorrecto por parte de los individuos.

Para analizar datos complejos, como los del sector sanitario, es necesario recopilar información que permita comprender fenómenos de base cuantitativa para realizar análisis de predicción, prescripción o clasificación.

En relación con la comprensión de los grupos de riesgo – una tarea relevante en términos de salud pública –, el primer paso es reconocer quiénes son los más susceptibles de experimentar los graves efectos de la COVID-19. En términos generales, se trata de personas de edad avanzada y con comorbilidades (hipertensión, obesidad y diabetes etc.) (Jordan et al., 2020). A medida que la pandemia se extiende, se observa que la población de mayor riesgo pasa a ser la más propensa a infectarse y la que tiene un acceso restringido al sistema de salud, es decir, las personas más pobres y las que no pueden auto-aislarse por diversas razones, especialmente por la falta de empleo formal y la imposibilidad de trabajar a distancia.

Esta población fue identificada y geolocalizada, lo que constituye una actividad esencial en la priorización de las regiones para las intervenciones de los gestores sanitarios. Además, es posible seleccionar lugares para la construcción de centros de testeos masivos, o incluso definir parámetros para la reapertura de ciudades en periodos de postcuarentena. Se espera que estas herramientas interactivas permitan a los gestores comparar las mediciones; aunque no hay certeza en cuanto a la magnitud absoluta de sus efectos, es bastante razonable considerar su comparación relativa.

Los autores del presente artículo también creen que las acciones guiadas por la epidemiología generarán soluciones más dinámicas para que los gestores sanitarios comprendan y evalúen cómo las medidas adoptadas en el presente repercutirán en el futuro de la pandemia de COVID-19. Asimismo, la identificación y localización de las poblaciones vulnerables agiliza la toma de decisiones y promueve una mejor gestión de la salud pública.

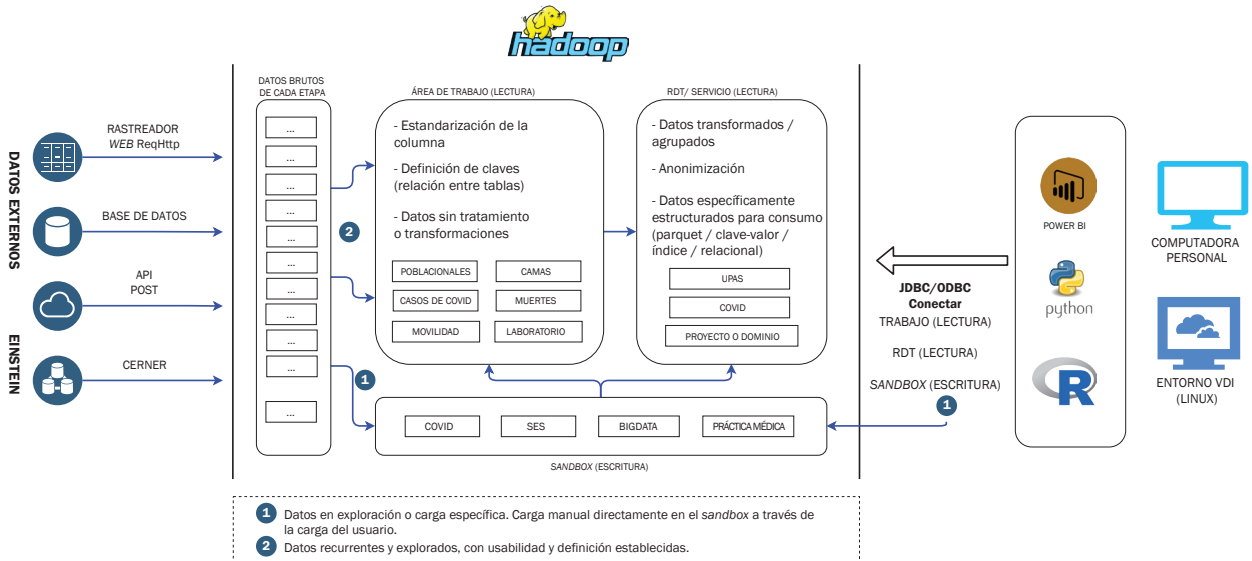
## Soluciones de *data lake*: Organizar los datos de múltiples fuentes

Para analizar datos complejos, como los del sector sanitario, es necesario recopilar información que permita comprender fenómenos de base cuantitativa para realizar análisis de predicción, prescripción o clasificación. En vista de la falta de datos disponibles al comienzo de la pandemia de COVID-19, esta necesidad se hizo aún más urgente. Cada una de las tres principales bases de datos que contienen las cifras diarias de casos de la enfermedad por países tiene una metodología diferente, lo que genera diferencias en el formato, la granularidad de las localizaciones y las direcciones (URL) de recuperación.

El trabajo de identificación, extracción, preparación y convergencia de los datos puede ser arduo y llevar mucho tiempo. En un equipo de científicos, es posible que muchos miembros necesiten los mismos datos, generando redundancia o incluso incoherencia en los resultados. Para minimizar este problema, es habitual crear un *data lake*, o lago de datos (a veces llamado océano de datos, o variaciones del mismo). En general, un lago de datos es un repositorio de datos en formato crudo, que contiene datos estructurados y no estructurados. Se trata de un entorno que puede realizar tareas como la extracción, la transformación y la carga (ETL, *Extract, Transform, Load*), el almacenamiento de datos a gran escala y el suministro de estos datos de manera uniforme.

Para apoyar la creación de herramientas de análisis para la pandemia de COVID-19 en Brasil, el equipo del presente artículo estableció una base de datos específica (COVID Lake), un lago de datos con datos de diversos alcances, como se muestra en la Figura 3.

Figura 3 – ESQUEMA ARQUITECTÓNICO DE UN LAGO DE DATOS<sup>15</sup>



Fuente: Elaboración propia.

Esta estructura sirvió para apoyar las actividades de los diferentes equipos. Se consideraron 25 bases de datos públicas o de proyectos específicos, muchos de ellos realizados en asociación, cuya naturaleza incluye: comportamiento (cuestionarios aplicados a la población brasileña, con el objetivo de comprender las aglomeraciones y la exposición a la contaminación); movimiento (datos relacionados con la movilidad mediante teléfonos móviles); clima (estaciones meteorológicas e índices de contaminación atmosférica); composición demográfica; e información de otros países.

## Conclusión

A pesar de la compleja experiencia que trajo la lucha contra el brote de COVID-19, ésta puede servir de acelerador en cuanto a las contribuciones a la adopción de soluciones de ciencia de datos, incluida la aplicación de la Inteligencia Artificial. Se espera que los resultados sean cada vez más prometedores para que en el futuro, si se produce una situación similar, se disponga de herramientas analíticas más maduras, capaces de prevenir daños que todavía no son evitables.

<sup>15</sup> Estructura basada en Hadoop, HDFS y HBase, con autenticación Kerberos, Active Directory (AD) corporativo. El acceso está restringido a la red interna de la institución.

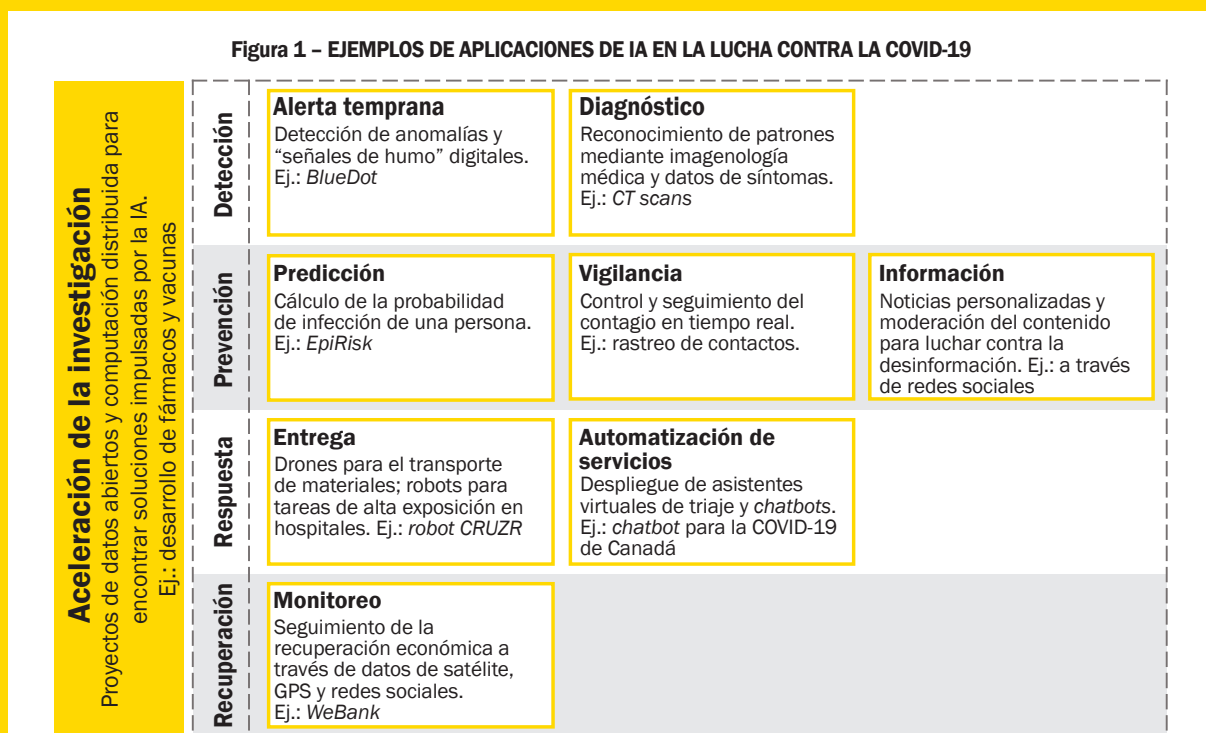


## Referencias

- Akselrod-Ballin, A. et al. (2019). Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology*, 292(2), 331–342.
- Alimadadi, A. et al. (2020). Artificial intelligence and machine learning to fight COVID-19. *Physiological Genomics*, 52(4), 200–202.
- Altmejd, A., Rocklöv, J., & Wallin, J. (2020). Nowcasting COVID-19 statistics reported with delay: A case-study of Sweden. *arXiv.org*, q-bio.PE.
- Barmparis, G. D., & Tsironis, G. P. (2020). Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. *Chaos, Solitons & Fractals*, 135, 109842.
- Barrozo, L. V. et al. (2020). GeoSES: A socioeconomic index for health and social research in Brazil. B. Lanza Queiroz, ed. *PLoS one*, 15(4), e0232074.
- Batista, A. F. de M. et al. (2020). COVID-19 diagnosis prediction in emergency care patients: A machine learning approach. *medRxiv*, 1–8.
- Brinati, D. et al. (2020). Detection of COVID-19 infection from routine blood exams with machine learning: A feasibility study. *Journal of Medical Systems*, 44(8).
- Britton, T. (2009). Stochastic epidemic models: A survey. *arXiv.org*, math.PR.
- Chowell, G. et al. (2016). Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*, 18, 66–97.
- Collins, G. S. et al. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. 1–10.
- Cosgriff, C. V. et al. (2019). Developing well-calibrated illness severity scores for decision support in the critically ill. *npj Digital Medicine*, 1–8.
- DeCaprio, D. et al. (n.d.). Building a COVID-19 Vulnerability Index. *arXiv.org*.
- Gao, L. et al. (2020). Prognostic value of NT-proBNP in patients with severe COVID-19. *Respiratory Research*, 21(1), 83–7.
- Hellewell, J. et al. (2020). Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet. Global health*, 8(4), e488–e496.
- Hidalgo, E. S. (2019). Adapting the scrum framework for agile project management in science: Case study of a distributed research initiative. *Heliyon*, 5(3), e01447.
- Hirsch, J. S. et al. (2020). Acute kidney injury in patients hospitalized with COVID-19. *Kidney International*, 98(1), 209–218.
- Hornik, R. (2004). Some reflections on diffusion theory and the role of Everett Rogers. *Journal of Health Communication*, 9(sup1), 143–148.
- Jordan, R. E., Adab, P., & Cheng, K. K. (2020). COVID-19: Risk factors for severe disease and death. *British Medical Journal. Clinical Research Edition*, 368, m1198.
- Lai, S. et al. (2020). Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. *medRxiv*.
- Lopez, L. R., & Rodo, X. (n.d.). A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics. *medRxiv*.
- MacIntyre, C. R., & Chughtai, A. A. (2020). A rapid systematic review of the efficacy of face masks and respirators against coronaviruses and other respiratory transmissible viruses for the community, healthcare workers and sick patients. *International Journal of Nursing Studies*, 108, 103629.
- Marston, C., Renedo, A., & Miles, S. (2020). Community participation is crucial in a pandemic. *Lancet*, 395(10238), 1676–1678.
- Martinez, E. Z., Aragon, D. C., & Nunes, A. A. (2020). Short-term forecasting of daily COVID-19 cases in Brazil by using the Holt's model. *Revista da Sociedade Brasileira de Medicina Tropical*, 53(10223), 102433–35.
- Meng, Z. et al. (n.d.). Development and utilization of an intelligent application for aiding COVID-19 diagnosis. *medRxiv*.
- Nadim, S. S., Ghosh, I., & Chattopadhyay, J. (2020). Short-term predictions and prevention strategies for COVID-19: A model based study. *arXiv.org*, q-bio.PE.
- Prem, K. et al. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *The Lancet. Public health*, 5(5), e261–e270.
- Puca, M., & Buonanno, P. (n.d.). Using newspaper obituaries to nowcast daily mortality: Evidence from the Italian COVID-19 hot-spots. *medRxiv*.
- Remuzzi, A., & Remuzzi, G. (2020). COVID-19 and Italy: What next? *Lancet*, 395(10231), 1225–1228.
- Roda, W. C. et al. (2020). Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling*, 5, 271–281.
- Santos, H. G. D. et al. (2020). Machine learning to predict 30-day quality-adjusted survival in critically ill patients with cancer. *Journal of Critical Care*, 55, 73–78.
- Schneble, M. et al. (2020). Nowcasting fatal COVID-19 infections on a regional level in Germany. *arXiv.org*, stat.AP.
- Sun, Q. et al. (2020). Lower mortality of COVID-19 by early recognition and intervention: Experience from Jiangsu Province. *Annals of Intensive Care*, 10(1), 33–34.
- Swapnarekha, H. et al. (2020). Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review. *Chaos, Solitons & Fractals*, 138, 109947–48.
- Wynants, L. et al. (2020). Prediction models for diagnosis and prognosis of COVID-19 infection: Systematic review and critical appraisal. *British Medical Journal. Clinical Research Edition*, 369, m1328.

## Usos de la Inteligencia Artificial en el contexto de la COVID-19

Las tecnologías de Inteligencia Artificial (IA) se están empleando de varias maneras para combatir la COVID-19. La Figura 1 presenta ejemplos de aplicaciones de IA en distintas fases de la crisis y con distintos fines, como la producción de información personalizada, el seguimiento de las infecciones en tiempo real o incluso el transporte y la entrega de materiales.



Fuente: OECD, Using artificial intelligence to help combat COVID-19 (2020).

## Salud pública y vigilancia: IA para el rastreo de contactos

El rastreo de contactos es una estrategia de vigilancia pública que se ha aplicado como mecanismo de prevención en el sector sanitario. Se refiere al proceso utilizado para identificar, informar y supervisar a las personas que estuvieron en contacto cercano con alguien infectado por el virus. Este proceso depende cada vez más del uso de tecnologías, especialmente de la IA.

En el contexto de la COVID-19, la IA se ha utilizado en estrategias destinadas a disminuir la transmisión del virus, a saber: la detección de la fiebre por infrarrojos en los espacios públicos; drones y robots basados en la IA para detectar el movimiento, las reuniones sociales y las personas sin mascarillas; y el uso de algoritmos de rastreo de contactos guiados por la IA para enviar mensajes de texto personalizados a los ciudadanos.

En China, un ejemplo de rastreo de contactos basado en la IA se apoya en complementos instalados en aplicaciones ampliamente utilizadas por la población. El *plug-in* recoge datos de los usuarios y alimenta una base de datos central, cuya información se analiza mediante herramientas de IA. A continuación, el algoritmo emite códigos de colores para indicar a los usuarios si están restringidos en sus idas y venidas, además de advertir a los contactos cercanos recientes en caso de un diagnóstico positivo de COVID-19.

## Artículo II

# Rastreo digital de contactos para luchar contra la COVID-19

Por Sacha Alanoca<sup>16</sup>, Nicolas Guetta Jeanrenaud<sup>17</sup>, Nyasha Weinberg<sup>18</sup>, R. Buse Çetin<sup>19</sup> y Nicolas Miaillhe<sup>20</sup>

## Una crisis de salud pública sin precedentes

Desde el brote de COVID-19 en Wuhan (China) a finales de 2019 y su clasificación como pandemia por la Organización Mundial de la Salud (OMS) en marzo de 2020, el número de muertes e infecciones ha seguido aumentando a nivel mundial. Hasta el 10 de julio de 2020, el recuento fue de más de 12,3 millones de casos diagnosticados y más de 555.000 muertes. En América Latina, las cifras se han disparado recientemente, alcanzando más de 3,3 millones de infecciones y 143.000 muertes, con la mayor concentración de casos en Brasil, Perú y Chile<sup>21</sup>. Ante esta crisis sin precedentes, la respuesta más inmediata de la mayoría de los gobiernos ha sido la aplicación de estrictas normas de distanciamiento social junto con cierres parciales o totales.

Estas medidas ralentizan el ritmo de las infecciones, o “aplanan la curva”, en referencia a la forma plana de los gráficos que muestran una tendencia reducida de los casos diarios, lo que permite a los sistemas sanitarios tratar el flujo de pacientes entrantes. Sin embargo, sin capacidad para realizar pruebas a gran escala, por un lado, y sin perspectivas de despliegue de tratamientos y vacunas en un futuro próximo, por otro, estos esfuerzos sólo pueden retrasar la propagación de la pandemia, no detenerla.

Ante esta crisis sin precedentes, la respuesta más inmediata de la mayoría de los gobiernos ha sido la aplicación de estrictas normas de distanciamiento social junto con cierres parciales o totales.

<sup>16</sup> Investigadora principal de políticas de IA y responsable de desarrollo comunitario en The Future Society. Tiene un máster por la Sciences Po París, donde se especializó en Economías Emergentes e Innovación Digital.

<sup>17</sup> Estudiante de posgrado en Tecnología y Política en el Instituto de Tecnología de Massachusetts (MIT), donde se especializa en ciencias sociales computacionales y sistemas de información. También tiene un máster en Ingeniería Civil por la l'Ecole des Mines de París.

<sup>18</sup> Investigadora asociada a The Future Society, actualmente con una beca del programa *re:constitution*. Es miembro del colegio de abogados del Reino Unido, experta en derecho y en políticas públicas, y posee una maestría de la Harvard Kennedy School.

<sup>19</sup> Investigadora de políticas de IA en The Future Society. Tiene un máster en Gestión Pública Internacional por la Escuela de Asuntos Internacionales Sciences Po de París (PSIA), donde se especializó en política económica y gobernanza global. Es licenciada en Ciencias Sociales y Estudios de Oriente Medio por la Sciences Po París.

<sup>20</sup> Cofundador de The Future Society. También es miembro del Comité Directivo de la asociación AI Commons, miembro del Grupo de expertos en IA de la OCDE y del Consejo Global de Inteligencia Extendida (CXI), entre otras organizaciones. Imparte clases en la Escuela de Asuntos Internacionales de París (Sciences Po), en la IE School of Global and Public Affairs de Madrid y en la Mohammed bin Rashid School of Government de Dubai. También es investigador asociado en Harvard y miembro de la IE Business School de Madrid.

<sup>21</sup> La OMS actualiza periódicamente estos datos. Disponible en: <https://COVID19.who.int>

Además, estas medidas tienen un coste socioeconómico muy importante. Desde el estallido de la pandemia, se calcula que más de 13 millones de personas están en situación de desempleo en Francia<sup>22</sup> y más de 33 millones en Estados Unidos<sup>23</sup>. En toda América Latina hay millones de desempleados. En Brasil, un récord de 7,8 millones de personas perdió su empleo entre marzo y mayo de 2020, dejando el porcentaje de la población que trabaja por debajo del 50%<sup>24</sup>. También es necesario tener en cuenta los impactos considerables del escenario actual en la salud mental de la población, con niveles crecientes de estrés y ansiedad, así como los costes aún desconocidos de las suspensiones o retrasos de los tratamientos sanitarios rutinarios (Ornell et al., 2020).

## La localización de contactos como técnica de mitigación rápida y precisa

Ante la urgente necesidad de mitigar la crisis, los gobiernos de todo el mundo han explorado una combinación de herramientas médicas y tecnológicas. Entre ellas, el rastreo de contactos describe una variedad de técnicas utilizadas para identificar a las personas que pueden haber estado en contacto con un individuo que recibió un diagnóstico positivo de COVID-19, tomando las medidas apropiadas para informarlas, aislarlas y tratarlas.

En el pasado, el rastreo manual de contactos se ha utilizado sistemáticamente para mitigar epidemias como la tuberculosis, el sarampión, el VIH y, más recientemente, para limitar la propagación del ébola y el SARS (Hart et al., 2020). Sin embargo, en comparación con otras enfermedades, la COVID-19 se considera altamente contagiosa<sup>25</sup>. El rastreo manual de contactos, que se basa en la memoria humana y en personal médico capacitado, carece de la velocidad y la precisión necesarias para adaptarse a la propagación del virus.

Mediante la movilización de herramientas digitales de planificación y análisis, el rastreo digital de contactos puede mejorar este escenario. Los teléfonos y las pulseras inteligentes automatizan este proceso, permitiendo el seguimiento de proximidad, la supervisión digital de la salud o incluso la identificación del paciente. Cuando se combina con otras medidas adecuadas, como el distanciamiento social y la realización de pruebas, el rastreo digital de contactos puede ayudar a romper las cadenas de contaminación y prevenir nuevos casos alertando a las poblaciones expuestas al riesgo. Varios países que han “aplanoado la curva” de la primera ola de brotes de COVID-19, como Taiwán, Corea del Sur, Singapur y China, han aplicado con éxito esta estrategia. Según el *MIT Technology Review*<sup>26</sup>, actualmente hay 47 aplicaciones de rastreo de contactos desplegadas en más de 30 países.

(...) el rastreo de contactos describe una variedad de técnicas utilizadas para identificar a las personas que pueden haber estado en contacto con un individuo que recibió un diagnóstico positivo de COVID-19, tomando las medidas apropiadas para informarlas, aislarlas y tratarlas.

<sup>22</sup> Más información en: <https://www.rtl.fr/actu/conso/coronavirus-pourquoi-y-a-il-autant-de-salaries-en-chomage-partiel-en-france-7800466953>

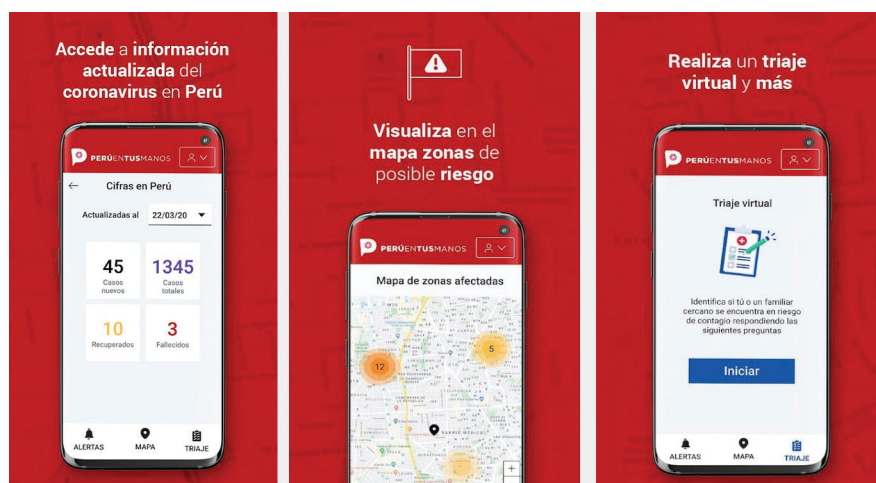
<sup>23</sup> Más información en: <https://www.bbc.com/news/business-52570600>

<sup>24</sup> Más información en: <https://edition.cnn.com/2020/07/01/americas/latin-america-coronavirus-unemployment-intl/index.html>

<sup>25</sup> La mediana del periodo de incubación es de cinco días, y el 97,5% de los que desarrollan síntomas lo hacen en 11,5 días (Lauer et al., 2020; Li et al., 2020). Las estimaciones de la proporción asintomática de individuos infectados oscilan entre el 15,5% (Mizumoto et al., 2020) y el 56% (Arons et al., 2020).

<sup>26</sup> Más información en: <https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-Covid-tracing-tracker>

Figura 1 – LA APLICACIÓN PERÚENTUSMANOS, DESPLEGADA POR EL GOBIERNO PERUANO



Fuente: Presidencia Del Consejo de Ministros, Perú.

Entre las cuestiones que se han planteado figuran su rendimiento y su impacto en la privacidad de los datos, los derechos humanos, la estigmatización de las personas, la desconfianza en las autoridades públicas y, por último, el temor a establecer una vigilancia masiva.

Sin embargo, en todo el mundo ha surgido rápidamente la preocupación por la adopción generalizada y extendida del rastreo digital de contactos. Entre las cuestiones que se han planteado figuran su rendimiento y su impacto en la privacidad de los datos, los derechos humanos, la estigmatización de las personas, la desconfianza en las autoridades públicas y, por último, el temor a establecer una vigilancia masiva. La desinformación y la falta de garantías éticas, legales y técnicas claras han polarizado el debate público en varios países en un momento en el que, más que nunca, es necesario generar confianza y evitar que se sacrifiquen las libertades civiles en nombre de la salud pública.

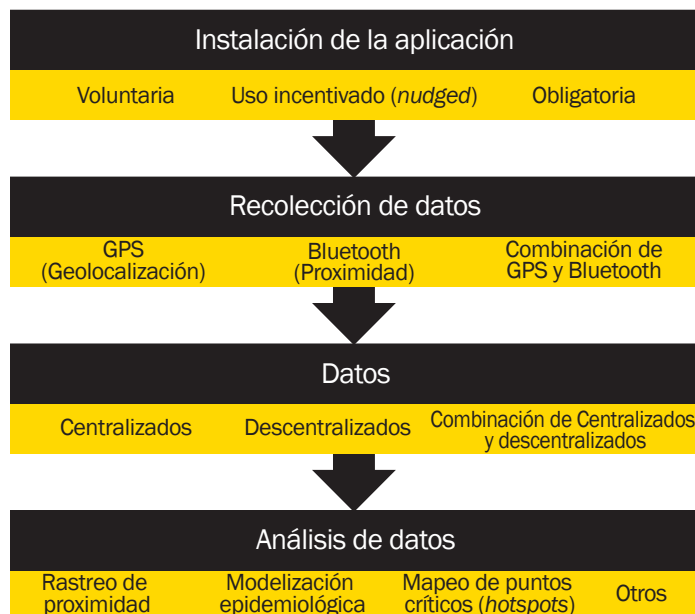
Hay una necesidad apremiante de co-diseñar mecanismos de gobernanza que maximicen los beneficios para la salud de las aplicaciones de rastreo de contactos al tiempo que mitiguen sus posibles efectos adversos. La urgencia de la situación sanitaria no debe repercutir en nuestra capacidad colectiva para tomar la decisión correcta y sortear con responsabilidad las posibles tensiones entre la salud pública, la seguridad y las libertades civiles. La evidencia sugiere que la adopción generalizada de estas aplicaciones depende de la confianza pública, lo que refuerza la necesidad de un referencial ético que rijas su desarrollo.

## Alternativas técnicas y *tradeoffs*

Las aplicaciones de rastreo de contactos se basan en una serie de tecnologías y protocolos de preservación de la privacidad. Pueden ser voluntarias (disponibles en las tiendas de aplicaciones para móviles), obligatorias o de uso incentivado (*nudged*). El proceso de recogida de datos se basa en la proximidad (normalmente a través de Bluetooth), en los datos de localización del GPS o en un modelo mixto. Los datos pueden almacenarse localmente en los dispositivos móviles de los individuos en un enfoque descentralizado, en un servidor centralizado o con un enfoque mixto. Por último, el análisis que realizan las autoridades competentes varía, y va desde el simple cálculo de la proximidad hasta la elaboración de modelos epidemiológicos, pasando por el mapeo de puntos críticos (*hotspots*) y mucho más.



Figura 2 – CARACTERÍSTICAS TÉCNICAS DEL RASTREO DIGITAL DE CONTACTOS



Fuente: The Future Society.

## Instalación de la aplicación

Un imperativo técnico del éxito de una aplicación de rastreo de contactos es la instalación y adopción generalizada por parte de los individuos. Aunque todavía no hay consenso entre los epidemiólogos sobre la tasa de transmisión de la COVID-19, un estudio realizado en la Universidad de Oxford estimó que, con una tasa de adopción mínima del 60%, estas aplicaciones podrían detener efectivamente la epidemia<sup>27</sup>. No obstante, se trata de un umbral mínimo: los autores del informe señalaron posteriormente que niveles de adopción más bajos ya serían vitales para frenar la propagación del virus<sup>28</sup>.

Aunque la instalación de estas aplicaciones es de gran importancia para la lucha de los gobiernos contra la COVID-19, las autoridades de varios países han adoptado un enfoque radical, haciéndolo obligatorio para todos los ciudadanos. En China, un “detector de contacto” se conecta a las aplicaciones existentes y ampliamente utilizadas WeChat y Alipay. El programa analiza los datos de salud y de viaje para asignar “colores de riesgo” a las personas, pudiendo negarles el acceso a tiendas y servicios esenciales<sup>29</sup>. En Kuwait y Bahréin, la aplicación es independiente de las funciones telefónicas existentes, pero su instalación es obligatoria por ley.

La gran mayoría de los países ha instituido aplicaciones voluntarias, aunque en sus funcionalidades se pueden incluir incentivos personalizados o genéricos. Algunos ofrecen información adicional sobre dónde hacerse la prueba de

Un imperativo técnico del éxito de una aplicación de rastreo de contactos es la instalación y adopción generalizada por parte de los individuos.

<sup>27</sup> Disponible en: <https://www.research.ox.ac.uk/Article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown>

<sup>28</sup> Más información en: <https://www.technologyreview.com/2020/06/05/1002775/Covid-apps-effective-at-less-than-60-percent-download>

<sup>29</sup> Más información en: <https://www.nytimes.com/2020/05/26/technology/china-coronavirus-surveillance.html>

COVID-19 (Australia) o la disponibilidad de hospitales (Turquía), mientras que otros publican noticias relacionadas con la pandemia (Bulgaria) o consejos sanitarios (Qatar). La medida en que estos incentivos adicionales lleven a la adopción obligatoria de la aplicación dependerá de si existen vías alternativas para acceder a servicios idénticos.

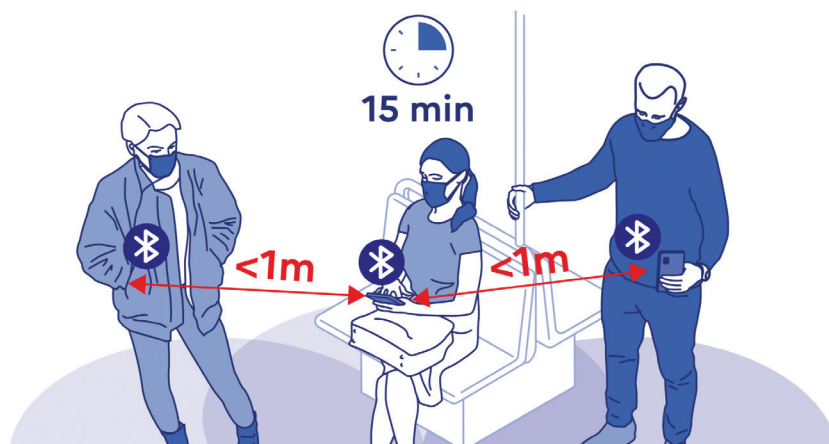
## Recolección de datos

Según la información del *COVID-19 Digital Rights Tracker*, de la empresa británica Top10VPN, en lo que respecta al proceso de recolección de datos, el 34% de las aplicaciones de rastreo de contactos existentes utilizan el GPS, el 35% el Bluetooth y el 24% una combinación de ambos. Esta distribución pone de manifiesto la falta de consenso en torno a la tecnología más eficaz para esta tarea.

Los protocolos Bluetooth se basan únicamente en la detección de proximidad. Un dispositivo móvil registra otras señales Bluetooth dentro de un rango dado durante un tiempo determinado, en lo que se denomina “apretón de manos digital”. Durante el intercambio, cada dispositivo recoge *tokens* seudonimizados<sup>30</sup> – generalmente cadenas de números que cambian aleatoriamente varias veces al día – como identificadores de los otros usuarios. Un mecanismo de notificación informa a los individuos cuando están cerca de un usuario señalado como infectado.

Países como México, Italia, Japón y el Reino Unido han adoptado protocolos *Bluetooth Low Energy* (BLE) de igual a igual, mientras que Singapur utiliza el protocolo *BlueTrace* basado en Bluetooth. En la mayoría de los casos, la proximidad se mide en un metro y los *tokens* seudonimizados se intercambian al cabo de 15 minutos, como muestra el ejemplo francés en la Figura 3.

**Figura 3 – APLICACIÓN STOPCOVID PROXIMITY-TRACING, DESPLEGADA POR EL GOBIERNO FRANCÉS**



Fuente: Gobierno de Francia (2020).

<sup>30</sup> Dispositivos físicos que generan una contraseña de protección temporal para las cuentas utilizadas por el usuario, ayudando a su seguridad personal.

Los esquemas GPS, por otro lado, se basan en los datos móviles para reconstruir el historial de localización de los usuarios. Cuando los individuos son marcados como infectados, la aplicación re-rastrea a todos los demás usuarios con los que se han cruzado. El prototipo de GPS tiene la característica adicional de resaltar los puntos geográficos del virus, permitiendo tomar medidas específicas, lo que no es posible utilizando Bluetooth. Sin embargo, esta técnica es menos precisa a la hora de registrar la proximidad, especialmente en interiores o durante los desplazamientos subterráneos. En cuanto a la privacidad, los protocolos GPS se han asociado con algunas de las aplicaciones más intrusivas, como BeAware en Bahrein, Shlonik en Kuwait y Smittestopp en Noruega<sup>31</sup>.

Por último, algunos protocolos de aplicación combinan datos Bluetooth y GPS para el rastreo de contactos. Un ejemplo de ello es la plataforma Private Kit Safe-Paths desarrollada en el Instituto de Tecnología de Massachusetts (MIT), que utiliza ambos tipos de datos para registrar la ubicación de las personas. Este protocolo, gratuito y de código abierto, ha sido adoptado en varios países, como Chipre<sup>32</sup>.

## Almacenamiento de datos

Además de las aplicaciones de rastreo de contactos, muchas otras tecnologías solicitan información personal adicional del usuario. Ketju, en Finlandia, y Hayat Eve Siğar, en Turquía, exigen números de teléfono válidos para la autenticación, mientras que en Islandia (sólo para las personas a las que se les ha diagnosticado COVID-19), Qatar y Kuwait, la autenticación se realiza mediante los números del DNI. Algunas aplicaciones, como ViruSafe en Bulgaria, recogen la edad y el historial médico de los usuarios en el momento del registro. Todas las aplicaciones de rastreo de contactos también recogen los datos del diagnóstico de COVID-19 para notificar a las personas en situaciones de riesgo.

Dada la naturaleza sensible de los datos en cuestión, la arquitectura de almacenamiento de datos ha sido objeto de mucho debate entre los desarrolladores de aplicaciones y los expertos en privacidad. En las soluciones centralizadas, los datos recogidos por las aplicaciones de rastreo de contactos se registran directamente en un servidor principal, que suele ser muy seguro y de fácil acceso para los organismos gubernamentales. Las soluciones descentralizadas, en cambio, mantienen registros de proximidad y localización en los dispositivos de los usuarios, que se almacenan localmente<sup>33</sup>.

Independientemente del protocolo de almacenamiento seleccionado, la mayoría de las aplicaciones de rastreo de contactos mitigan los problemas de privacidad mediante la desidentificación de los datos. Aunque la anonimización completa de los datos nunca es alcanzable en la práctica (Rocher, Hendrickx & Montjoye, 2019), diferentes métodos de seudonimización de datos son aspectos integrales de la mayoría de las aplicaciones de rastreo de contactos. Aplicaciones como Stopp Corona en Austria, StopCovid en Francia y TraceTogether en

Dada la naturaleza sensible de los datos en cuestión, la arquitectura de almacenamiento de datos ha sido objeto de mucho debate entre los desarrolladores de aplicaciones y los expertos en privacidad.

<sup>31</sup> Más información en: <https://www.amnesty.org/en/latest/news/2020/06/bahrain-kuwait-norway-contact-tracing-apps-danger-for-privacy>

<sup>32</sup> Más información en: <https://www.financialmirror.com/2020/04/06/technology-recruited-in-fight-against-coronavirus>

<sup>33</sup> Como ejemplos de estos enfoques, la solución descentralizada conocida como DP3T, desarrollada por Google y Apple, aparece en aplicaciones de Finlandia, Malasia y Suiza. Francia y Alemania también han adoptado lo que llaman un "protocolo de rastreo de contactos centralizado-descentralizado". El protocolo ROBERT (*ROBust and privacy-presERving proximity Tracing*) almacena los datos inicialmente en los dispositivos, pero los envía a un servidor centralizado una vez que los usuarios han dado positivo.

Singapur atribuyen identificaciones únicas a los usuarios, que pueden cambiar con el tiempo para mayor seguridad. Tanto el Private Kit SafePath del MIT como los protocolos de Google-Apple utilizan la privacidad diferencial para garantizar que los datos agregados puestos a disposición por la aplicación no permitan reidentificar a los usuarios individuales.

## Análisis de datos

Los datos recogidos por las aplicaciones de rastreo de contactos pueden servir para diferentes objetivos de los organismos gubernamentales. El uso mínimo de los datos de rastreo de contactos consiste en señalar la proximidad entre los individuos infectados y los usuarios de la aplicación (como medido por Bluetooth o GPS). Sin embargo, varios países aprovechan los datos de las aplicaciones de rastreo de contactos para diferentes fines. El plan inicial del Reino Unido para el rastreo digital de contactos, por ejemplo, permitía que los datos recogidos a partir de la aplicación COVID-19 del Sistema Nacional de Salud alimentaran también modelos epidemiológicos para estudiar la propagación del virus<sup>34</sup>.

## Riesgos éticos que plantea el rastreo digital de contactos

La adopción de aplicaciones de rastreo digital de contactos plantea una serie de cuestiones éticas relevantes. Dada la amplitud y la naturaleza sensible de los datos recogidos, estas tecnologías tienen el potencial de socavar libertades civiles fundamentales y valores humanos como la privacidad, la protección de datos, la autonomía humana y la equidad.

## Privacidad y protección de datos

Todo rastreo digital de contactos requiere cierto grado de acceso a información que podría vulnerar la privacidad, como el estado de salud, la ubicación y la información de las tarjetas de crédito. Por lo tanto, los riesgos para la privacidad son inherentes al uso de esta tecnología. Por ejemplo, otras aplicaciones instaladas en un *smartphone* pueden captar la aplicación de rastreo digital de contactos y enviar datos a terceros, y los datos anonimizados son susceptibles de ser reidentificados. Según el *COVID-19 Digital Rights Tracker*, el 20% de las aplicaciones de rastreo digital de contactos no tienen políticas de privacidad. Por lo tanto, la falta de información y de salvaguardias legales, técnicas y políticas suscita grandes preocupaciones.

En este contexto, hay dos aspectos importantes: las características reales de preservación de la privacidad de las propias aplicaciones y la medida en que ofrecen a los usuarios el control de la privacidad. La privacidad real implica limitar la exposición de la información en la mayor medida posible, mientras que el control de esa privacidad significa permitir a los individuos tomar decisiones transparentes sobre el uso de sus datos (Loi et al., 2020). La lucha contra la

El uso mínimo de los datos de rastreo de contactos consiste en señalar la proximidad entre los individuos infectados y los usuarios de la aplicación (como medido por Bluetooth o GPS). Sin embargo, varios países aprovechan los datos de las aplicaciones de rastreo de contactos para diferentes fines.

---

<sup>34</sup> Más información en: <https://social.techcrunch.com/2020/05/05/nhs-COVID-19-the-uks-coronavirus-contacts-tracing-app-explained>

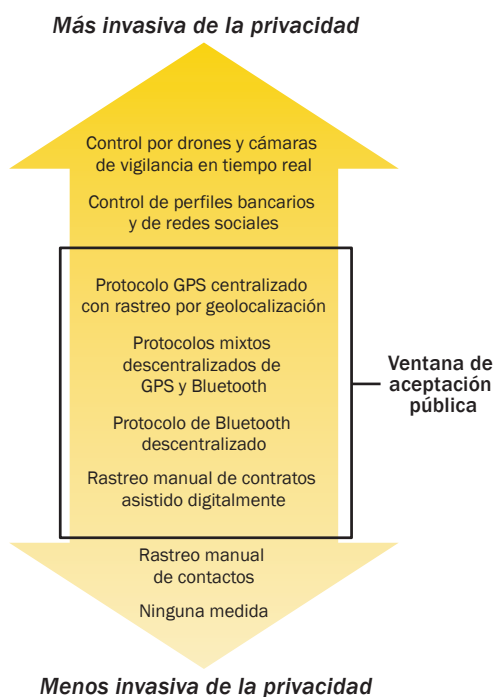
COVID-19 a través del rastreo digital de contactos puede implicar ajustes temporales en cuanto a la privacidad real; sin embargo, los individuos siempre deben estar empoderados para tomar estas decisiones por sí mismos.

Un aspecto importante es que la seudonimización de los datos y las medidas técnicas adoptadas en este sentido por las aplicaciones de rastreo de contactos pueden no mitigar totalmente el riesgo. Se ha afirmado que varias aplicaciones, incluida la francesa StopCovid, no han preservado suficientemente el anonimato, a pesar del cifrado de los usuarios. En 2013, los investigadores estudiaron los datos de localización de 1,5 millones de personas y descubrieron que eran tan específicos de los hábitos individuales que el 95% de ellos podían ser identificados (Rocher, Hendrickx, & Montjoye, 2019). Cuando las aplicaciones de rastreo de contactos combinan los datos de localización con otras informaciones sensibles, como en Taiwán y Corea del Sur, existe un mayor riesgo de violación de la privacidad e incluso de vigilancia masiva.

En este sentido, preocupa el uso del rastreo digital de contactos para el establecimiento de infraestructuras de vigilancia por parte de gobiernos y de grandes empresas tecnológicas. Atrapados entre el colapso económico y el desastre social, algunos gobiernos han desplegado herramientas de rastreo digital de contactos a gran velocidad y sin pasar por procedimientos de consentimiento democrático e informado. Se teme que las medidas excepcionales que infringen las libertades civiles seguirán existiendo después de la pandemia y se convertirán en “la nueva normalidad”. Es crucial garantizar que el rastreo digital de contactos sea voluntario y temporal, para evitar que el alcance de la vigilancia se extienda más allá de sus objetivos iniciales.

(...) la seudonimización de los datos y las medidas técnicas adoptadas en este sentido por las aplicaciones de rastreo de contactos pueden no mitigar totalmente el riesgo.

**Figura 4 – ESPECTRO DE PRIVACIDAD DE LAS APLICACIONES PARA LUCHAR CONTRA LA COVID-19**



Fuente: The Future Society.



Si no se adoptan medidas específicas para reducir la brecha digital, la base de usuarios no reflejará a toda la población, y determinados subgrupos se verán privados de los servicios de las aplicaciones.

## Estigma y discriminación

Se ha planteado la preocupación por el impacto de las aplicaciones de rastreo de contactos en grupos sociales que históricamente han sufrido discriminación, estigmatización y abuso. Existen tres riesgos importantes, especialmente en los Estados que imponen sanciones penales a la transmisión de la COVID-19, o que utilizan los datos recogidos en solicitudes para tomar decisiones en tanto “pasaporte” a determinados servicios. El primer riesgo es que los datos se utilicen para estigmatizar a determinados grupos, como las mujeres (Davis, 2020). El segundo es que los datos de localización se utilicen para estigmatizar y culpar a determinados grupos minoritarios<sup>35</sup>. El tercer riesgo es que los grupos eviten utilizar las aplicaciones de localización de contactos, prohibiendo el acceso a servicios o lugares.

El riesgo de estigmatización es intrínseco al seguimiento digital de los contactos, ya que implica la recopilación de datos sobre poblaciones enteras. Utilizados fuera del propósito de vigilar la exposición a la COVID-19, los datos pueden llevar a señalar “zonas críticas” del virus. Los barrios de minorías étnicas podrían convertirse en un blanco fácil para la estigmatización y la discriminación, así como los trabajadores de primera y segunda línea, los repartidores de comida, los cajeros y los conductores. Cuando estas profesiones son ejercidas por grupos marginados, el impacto del rastreo digital de contactos es aún más grave. La estigmatización crea alienación y perjudica la cohesión social.

## Accesibilidad

Las aplicaciones de rastreo de contactos también dependen del acceso tecnológico y la alfabetización digital para su adopción masiva. Si no se adoptan medidas específicas para reducir la brecha digital, la base de usuarios no reflejará a toda la población, y determinados subgrupos se verán privados de los servicios de las aplicaciones. Aunque está cada vez más extendido, el acceso a la tecnología móvil sigue siendo incompleto en todo el mundo: Según el Pew Research Center, en 2019, dos tercios de la población mundial no poseían un *smartphone*<sup>36</sup>.

La alfabetización digital suele estar muy correlacionada con otras características socioeconómicas. Un estudio reciente sobre los usuarios de Twitter en Italia reveló que son, en promedio, más jóvenes y con mayor nivel de educación que la población en general (Vaccari et al., 2013). También hay que señalar que las personas con discapacidad históricamente han sido marginadas por Internet y en las plataformas digitales<sup>37</sup>.

Según Floridi (2020), estas aplicaciones funcionarán mejor cuando estén más extendidas. Sin embargo, este será el caso donde los niveles de alfabetización digital y de posesión de teléfonos móviles sean más altos. Por lo tanto, existe un riesgo muy concreto de privilegiar a los ya privilegiados, así como a sus zonas de residencia. En este contexto, la brecha digital puede convertirse en una brecha biológica.

<sup>35</sup> Más información en: <https://news.un.org/en/story/2020/03/1060602>

<sup>36</sup> Más información en: <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally>

<sup>37</sup> Más información en: <http://www.theguardian.com/technology/2015/jun/29/disabled-people-internet-extra-costscommission-scope>

## El caso de un referencial ético para el rastreo digital de contactos

Es nuestra responsabilidad colectiva desarrollar y adoptar un referencial ético para navegar por los conflictos entre la salud pública, la seguridad, la actividad económica y las libertades civiles. La urgencia de la situación provocada por la COVID-19 no debería afectar a nuestra capacidad para construir un marco de referencia integral que oriente las decisiones de los ciudadanos, los desarrolladores de aplicaciones y los formuladores de políticas públicas. Los individuos informados están más dispuestos a aceptar sacrificios potencialmente significativos pero necesarios cuando las elecciones son voluntarias y están éticamente justificadas. Por lo tanto, la creación de un marco ético fiable es fundamental para garantizar la cohesión social y la eficacia de las aplicaciones de rastreo de contactos.

<b>Cuadro de principios éticos</b>	
<b>Propósito y rendimiento</b>	El propósito de las aplicaciones debe ser claro, comprensible dentro del contexto más amplio, medible y auditable de forma independiente.
<b>Voluntad y reversibilidad</b>	Los individuos deben poder elegir si instalan las aplicaciones por su propia voluntad, sin consecuencias negativas si deciden no hacerlo. Los usuarios deben poder desactivar las aplicaciones temporal o permanentemente en cualquier momento, sin que los desarrolladores de aplicaciones o terceros almacenen los datos personales o la información de proximidad restante.
<b>Privacidad por diseño (privacy by design)</b>	Las aplicaciones de rastreo de contactos deben alcanzar los niveles más altos de protección de la privacidad. El almacenamiento de datos debe ser seguro y con seudónimo.
<b>Uso mínimo de datos y tecnología</b>	La recolección de datos debe ser proporcionada, estar justificada y tener una fecha de caducidad definida. Sólo deben utilizarse y almacenarse los datos mínimos necesarios para cumplir la finalidad de las aplicaciones.
<b>Transparencia y verificabilidad</b>	El código fuente completo de las aplicaciones y los protocolos de rastreo principales deben estar disponibles de forma gratuita y ser reproducibles, sin restricciones de acceso para las auditorías.
<b>No discriminación y no estigmatización</b>	Los desarrolladores de aplicaciones y los responsables políticos deben garantizar que las aplicaciones de rastreo de contactos no estigmaticen o discriminen a las personas que han dado positivo en el testeo de COVID-19 o a sus familiares, a las categorías de trabajadores sociales, a los barrios o a las personas que no deseen utilizar las aplicaciones.
<b>Accesibilidad</b>	Hay que reconocer que las aplicaciones para teléfonos inteligentes y la conectividad a Internet no son accesibles para toda la población. Algunos ciudadanos pueden no tener teléfonos inteligentes, y las personas con discapacidades, los ancianos o las personas con menos conocimientos tecnológicos podrían no ser capaces de utilizar dichas aplicaciones. Deben desarrollarse soluciones complementarias y alternativas para garantizar la accesibilidad.
<b>Aviso y consentimiento informado</b>	La información sobre los fines y las características, así como la recolección de datos, debe presentarse claramente a los usuarios. El consentimiento informado y explícito debe ser un requisito previo para las aplicaciones. Deben evitarse los patrones de diseño oscuro ( <i>dark design</i> ) (por ejemplo, los “empujones” – <i>nudges</i> – a través de las notificaciones <i>push</i> , las aplicaciones pre-instaladas de fábrica en los <i>smartphones</i> y las funciones ocultas para desactivar o eliminar las aplicaciones).
<b>Rendición de cuentas</b>	Las aplicaciones de rastreo de contactos deben ser evaluadas y auditadas continuamente por entidades independientes legítimas en las que el público pueda confiar plenamente. Todas las partes implicadas en el diseño y el despliegue de las aplicaciones deben rendir cuentas con arreglo a marcos jurídicos claros en cuanto a sanciones y responsabilidades.

# Conclusiones

A medida que el número de casos positivos de COVID-19 sigue aumentando en todo el mundo, es una tarea colectiva no sólo explorar todas las soluciones médicas y tecnológicas, sino también gestionarlas de forma responsable. En los últimos años, la confianza del público en las instituciones políticas y las empresas tecnológicas se ha ido erosionando, lo que hace aún más relevante y necesario buscar proactivamente la concientización y el consentimiento del público antes de desplegar aplicaciones de rastreo de contactos. Un marco ético puede ayudar a construir un entendimiento común y guiar las decisiones de las diferentes partes interesadas. Para que este marco sea factible, los principios tendrán que traducirse en criterios específicos y acoplarse a parámetros de auditoría desarrollados y aplicados por terceros, con el fin de cultivar la confianza entre los usuarios y otras partes interesadas.

# Referencias

- Arons, M. M. et al. (2020). Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *New England Journal of Medicine*, 382, 2081.
- Davis, S. (2020, 29 Abril). *Contact tracing apps: Extra risks for women and marginalized groups*. Health and Human Rights Journal. Recuperado de <https://www.hhrjournal.org/2020/04/contact-tracing-apps-extra-risks-for-women-and-marginalized-groups>
- Floridi, L. (2020, 18 Abril). *Mind the app: Considerations on the ethical risks of COVID-19 apps*. Onlife. Recuperado de <https://thephilosophyofinformation.blogspot.com/2020/04/mind-app-considerations-on-ethical.html>
- Gobierno de Francia (2020). *ProjetStopCovid: Dossier de presse (21 mai 2020)*. Recuperado de <https://decryptageo.fr/wp-content/uploads/2020/05/Dossier-de-presse-Pojet-StopCovid.pdf>
- Hart, V. et al. (2020). Outpacing the virus: Digital response to containing the spread of COVID-19 while mitigating privacy risks. *Edmond J. Safra Center for Ethics*. Recuperado de <https://ethics.harvard.edu/outpacing-virus>
- Lauer, S. A. et al. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172, 577.
- Li, Q. et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*, 382, 1199.
- Loi, M. et al. (2020). *Ethical framework for human-centric public health digital surveillance*. Recuperado de [https://docs.google.com/document/d/19F\\_hXlIPvDKCk8JTfxNX0uex0r\\_GpmPEUuoSUMEj-zY/edit](https://docs.google.com/document/d/19F_hXlIPvDKCk8JTfxNX0uex0r_GpmPEUuoSUMEj-zY/edit)
- Mizumoto, K. et al. (2020). Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*, 25, 2000180.
- Ornell, F., Schuch, J. B., Sordi, A. O., & Kessler, F. H. P. (2020). “Pandemic fear” and COVID-19: mental health burden and strategies. *Brazilian Journal of Psychiatry*, 42(3), 232-235.
- Rocher, L., Hendrickx, J. M., & Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10, 3069.
- Vaccari, C. et al. (2013). Social media and political communication: A survey of Twitter users during the 2013 Italian general election. *Rivista Italiana di Scienza Politica*.

## Entrevista I

**P.S.I.\_** *Cada vez se habla más de los posibles aportes de la Inteligencia Artificial (IA) al ámbito de la salud. ¿Cuáles son los retos a los que se enfrenta su implantación en el sector?*

**E.M.\_** El uso de algoritmos y herramientas de IA para ayudar a médicos, enfermeras y otros profesionales en sus prácticas ha ido creciendo constantemente. Estas soluciones tienen el potencial de mejorar la eficiencia, la eficacia y la equidad de acceso en el ámbito de la salud. Sin embargo, aunque las perspectivas son optimistas, es importante tener en cuenta los retos que conlleva la tarea.

La discrepancia entre los esfuerzos por desarrollar modelos y soluciones de IA, por un lado, y su uso satisfactorio, por otro, refleja las dificultades de la aplicación de las herramientas de apoyo a la toma de decisiones en general. Aunque existen obstáculos técnicos, la mayoría de los retos tienen que ver con: la complejidad de adaptar las aplicaciones para integrarlas en sistemas como las historias clínicas electrónicas; la falta de alineación en relación con las necesidades reales de los usuarios; la falta de gestión de las expectativas; y los aspectos éticos y legales. Todos estos factores deben añadirse a la ecuación para obtener verdaderos beneficios.

La IA ofrece muchas oportunidades de soluciones que ayudan al diagnóstico de problemas clínicos, a la terapia de enfermedades y a la reducción de costos, generando un impacto significativamente positivo en el área de la salud en su conjunto. Sin embargo, es necesario ser consciente de la capacidad y las limitaciones actuales de estas técnicas para evitar la frustración y el consiguiente movimiento antagonista que obstaculizaría el progreso de la IA en el sector. Lo ideal es un equilibrio realista entre lo que se puede hacer y lo que es adecuado. Informar y educar a todos los públicos implicados – trabajadores de la informática, ingenieros, gestores, profesionales de la salud y la población en general – en relación con los principios y el uso de estas herramientas es fundamental.

**P.S.I.\_** *En el contexto del ecosistema de datos en el sector sanitario, donde existe una gran heterogeneidad de fuentes y se dispone de grandes volúmenes de datos, ¿cómo debe gestionarse la gobernanza de la IA?*

**E.M.\_** El compromiso de la comunidad científica, los profesionales de la salud y los ciudadanos es esencial para la integración y el uso equitativo y transparente de los datos. La gobernanza de la IA para las prácticas clínicas debería incluir, no sólo a los desarrolladores de tecnología, sino también a los representantes de los proveedores de servicios, los gestores, los usuarios y los pacientes. Debe tener en cuenta el análisis del rendimiento de la aplicación y su potencial costo-beneficio para mejorar la salud y reducir los costos sanitarios.

Para tener éxito en el ámbito de la salud, los algoritmos de IA deben entrenarse con datos que sean representativos de las características de la



**Dra. Eneida A. Mendonça**

Profesora de Pediatría y Bioestadística en la Universidad de Indiana, vicepresidenta de Desarrollo de Investigación y directora interina del Centro Clem McDonald de Informática Biomédica en el Instituto Regenstrief.

"Los métodos de aprendizaje automático tienden a funcionar bien en cuanto a la extracción de características, patrones y reglas a partir de datos complejos y en grandes volúmenes (Big Data), pero no producen significado, propósito, sentido de justicia o equidad."

población real: sólo así los modelos generados lograrán resultados adecuadamente precisos. Ante el aumento de la cantidad y la diversidad de datos relacionados con el sector (historias clínicas electrónicas, datos genéticos y determinantes sociales de la salud, por ejemplo), hay que crear un ecosistema que permita almacenar, integrar y analizar estos datos.

Ya existen varias opciones de estructuras de datos y estándares de representación de la información biomédica, como los códigos CID, LOINC y SNO-MED. Sin embargo, su uso no es universal, lo que reduce sustancialmente la capacidad de uso de la IA. Hay que dar prioridad a la adopción de modelos de datos comunes – por ejemplo, *Observational Medical Outcomes Partnership* (OMOP) e *Informatics for Integrating Biology and the Bedside* (i2b2) –, así como a estándares como el HL7 FHIR, que permiten la transferencia estandarizada de datos, junto con la continuidad de los debates sobre políticas para compartir datos, gobernanza, mecanismos para evaluar sistemas y algoritmos.

**P.S.I.\_** *Dado que las soluciones de IA se basan en los datos, ¿cuáles son los riesgos que conlleva su aplicación en el ámbito de la salud? ¿Cómo se pueden evitar estos riesgos?*

**E.M.\_** Aparte de la complejidad de los datos utilizados para entrenar los modelos, el ciclo de desarrollo, validación e implementación de las herramientas de IA requiere adoptar modelos muy controlados y rigurosos desde el punto de vista técnico-científico. En la práctica, en términos de riesgo para el paciente, estos algoritmos son similares a los medicamentos y equipos médicos.

Los modelos y las herramientas también deben evaluarse en relación con el rendimiento, la utilidad, las vulnerabilidades y los sesgos. Las variaciones en las normas locales de atención y terapia, las poblaciones con características diferentes y los sesgos en la selección de los datos de aprendizaje y validación son ejemplos de factores que pueden afectar significativamente al uso de los algoritmos y provocar consecuencias no deseadas.

Los métodos de aprendizaje automático tienden a funcionar bien en cuanto a la extracción de características, patrones y reglas a partir de datos complejos y en grandes volúmenes (Big Data), pero no producen significado, propósito, sentido de justicia o equidad. Es esencial reconocer que los algoritmos “aprenden” basándose únicamente en los datos que los entrenan. Si el conjunto de datos contiene sesgos, el algoritmo tendrá el potencial de amplificarlos. También es importante que haya transparencia en cuanto a cómo se utilizarán los datos y cómo se validarán los modelos. Además de la reproducibilidad en diferentes poblaciones, es necesario respetar el equilibrio entre la innovación, la seguridad y la privacidad de los datos de los individuos.

Por último, sugiero la lectura de una reciente publicación de la Academia Nacional de Medicina de los Estados Unidos<sup>38</sup> sobre la adecuada implantación de aplicaciones de IA en el sector sanitario. Los autores del documento presentan recomendaciones detalladas y ofrecen interesantes reflexiones para la comunidad.

---

<sup>38</sup> Disponible en: <https://nam.edu/artificial-intelligence-special-publication>



## Entrevista II

**P.S.I.\_** *¿Existen marcos propuestos por la OMS para regular, comparar y certificar los métodos basados en la IA en el ámbito de la salud digital? En caso afirmativo, ¿cómo se aplican estos procesos en la práctica?*

**B.M.J.\_** En la lucha contra la COVID-19, las tecnologías como la IA se utilizan para ayudar al triaje de la población, el seguimiento de los casos de infección, el control de los recursos y, sobre todo, para definir los determinantes sociales de la salud, que son el elemento fundamental de la equidad y los derechos humanos en la lucha contra la COVID-19. Sin embargo, aunque la IA y las tecnologías similares tengan un enorme potencial para el bien, también debemos dejar espacio para debatir las cuestiones éticas que rodean su uso en la vida real. Debemos asegurarnos de que la IA se utiliza correctamente en los sistemas sanitarios, especialmente en lo que respecta al acceso equitativo y a la privacidad de los pacientes.

La OMS está desarrollando la Estrategia Mundial de Salud Digital. Propone un marco de regulación, evaluación comparativa (*benchmarking*) y certificación de la IA y los dispositivos médicos de salud digital para su aplicación por los Estados Miembros. Junto con la Unión Internacional de Telecomunicaciones (UIT), la OMS está trabajando en la elaboración de una guía de *benchmarking* sobre la IA para los programas de salud. Fueron creados grupos de expertos sobre la ética de la IA y la regulación del área para desarrollar el marco de referencia.

El marco sobre IA para la salud incluirá las siguientes áreas de interés:

- Garantizar un acceso equitativo a la IA;
- Determinar de qué forma la IA interactúa con la problemática de la brecha digital;
- Preservar los derechos individuales de autonomía, privacidad, consentimiento informado y ausencia de prejuicios y discriminación;
- Garantizar la educación sobre el funcionamiento y la toma de decisiones de la IA;
- Garantizar la igualdad de acceso a las bases de datos, independientemente de los medios económicos;
- Mantener el control humano de la IA; y
- Reforzar la supervisión pública y la regulación del sector privado.

Además, la OMS y la UIT establecieron un Grupo focal sobre Inteligencia Artificial para la Salud (FG-AI4H) en julio de 2018. El FG-AI4H está desarrollando un proceso de *benchmarking* para los modelos de IA en el ámbito de la salud, que puede actuar como un marco de evaluación internacional estándar e independiente.



**Sr. Bernardo  
Mariano Junior**

*Chief Information Officer (CIO) y director del Departamento de Salud Digital e Innovación de la Organización Mundial de la Salud (OMS).*

"En el centro de la iniciativa [EIOS] está la creación de una comunidad de práctica para la inteligencia en salud pública que incluya a los Estados miembros, organizaciones internacionales, institutos de investigación y otros socios y colaboradores. Su objetivo último es salvar vidas."

### **P.S.I.\_ ¿Existen iniciativas específicas de IA que la OMS haya puesto en marcha o esté supervisando en el contexto de la pandemia de COVID-19?**

**B.M.J.\_** En septiembre de 2017, la OMS aceptó el liderazgo de la iniciativa Inteligencia Epidemiológica de Fuentes Abiertas (EIOS, *Epidemic Intelligence from Open Sources*), una colaboración única entre diversas partes interesadas en la salud pública de todo el mundo. Reúne iniciativas, redes y sistemas nuevos y existentes para crear un enfoque unificado *all-hazards* (centrado en todo el espectro de emergencias o desastres) *One Health* (acciones colaborativas, multisectoriales y transdisciplinarias) para la detección temprana, la verificación, la evaluación y la comunicación de las amenazas para la salud pública utilizando la información disponible públicamente. En el centro de la iniciativa está la creación de una comunidad de práctica para la inteligencia en salud pública (PHI, *Public Health Intelligence*) que incluya a los Estados miembros, organizaciones internacionales, institutos de investigación y otros socios y colaboradores. Su objetivo último es salvar vidas mediante la detección temprana de amenazas y la posterior intervención. La comunidad de prácticas se apoya en un sistema EIOS en evolución para la inteligencia en salud pública que no solo conecta otros sistemas y actores – como ProMed, HealthMap y la Red Mundial de Inteligencia Sanitaria (GPHIN, *Global Public Health Intelligence Network*) –, sino que también promueve y cataliza el desarrollo de colaboraciones nuevas e innovadoras mediante el uso de inteligencia artificial y aumentada, principalmente en el ámbito de la lingüística computacional y el procesamiento del lenguaje natural. El sistema EIOS se basa en una larga colaboración entre la OMS y el Centro Común de Investigación (CCI) de la Comisión Europea (CE) para responder a la necesidad de una iniciativa mundial que aglutine los esfuerzos de la IPS. Desde el comienzo de la pandemia de COVID-19, los miembros de la comunidad de EIOS han estado trabajando en el desarrollo de funciones avanzadas de IA y aprendizaje automático para ayudar a gestionar el volumen sin precedentes de información procedente de fuentes oficiales y no oficiales. Esto implica buscar formas mejoradas de filtrar, contextualizar y visualizar todo el flujo de contenido, como la introducción de datos adicionales a partir de las redes sociales (Twitter), así como el reconocimiento de la fiabilidad de los artículos de noticias mediante el análisis del tono y el estilo de escritura.

# Informe de Dominios

## Dinámica del registro de dominios en Brasil y en el mundo

El Centro Regional de Estudios para el Desarrollo de la Sociedad de la Información (Cetic.br) realiza un seguimiento mensual del número de nombres de dominio de alto nivel con código de país (ccTLD) registrados entre los países del G20<sup>39</sup>. En conjunto, superan los 79,50 millones de registros. En agosto de 2020, los dominios registrados en .de (Alemania) alcanzaron los 16,53 millones, seguidos por China (.cn), el Reino Unido (.uk) y Rusia (.ru), con 16,05 millones, 9,50 millones y 4,96 millones de registros, respectivamente. Brasil tuvo 4,41 millones de registros bajo el .br, ocupando el 5º lugar de la lista, como se muestra en la Tabla 1<sup>40</sup>.

**Tabla 1 – REGISTRO DE NOMBRES DE DOMINIO ENTRE LOS PAÍSES DEL G20 – AGOSTO DE 2020**

Posición	Países del G20	Número de dominios	Fuente
1	Alemania (.de)	16.531.825	www.denic.de
2	China (.cn)	16.048.918	research.domaintools.com/statistics/tld-counts/
3	Reino Unido (.uk)	9.501.094	www.nominet.uk/news/reports-statistics/uk-register-statistics-2020/
4	Rusia (.ru)	4.960.220	cctld.ru
<b>5</b>	<b>Brasil (.br)</b>	<b>4.408.632</b>	<b>registro.br/dominio/estadisticas/</b>
6	Francia (.fr)	3.566.512	www.afnic.fr/en/resources/statistics/detailed-data-on-domain-names/
7	Unión Europea (.eu)	3.531.400	research.domaintools.com/statistics/tld-counts/
8	Italia (.it)	3.315.268	nic.it
9	Australia (.au)	3.205.126	www.auda.org.au/
10	Canadá (.ca)	2.935.244	www.cira.ca
11	India (.in)	2.200.000	www.registry.in/
12	Estados Unidos (.us)	1.685.195	research.domaintools.com/statistics/tld-counts/
13	Japón (.jp)	1.596.328	jprs.co.jp/en/stat/
14	Sudáfrica (.za)	1.264.123	www.zadna.org.za
15	Corea del Sur (.kr)	1.107.229	krnic.or.kr/jsp/eng/domain/kr/statistics.jsp
16	México (.mx)	910.012	research.domaintools.com/statistics/tld-counts/
17	Argentina (.ar)	612.656	nic.ar/es/dominios/estadisticas
18	Indonesia (.id)	434.525	pandi.id/?lang=en
19	Turquía (.tr)	430.555	www.nic.tr/index.php?USRACTN=STATISTICS
20	Arabia Saudita (.sa)	69.293	www.nic.sa/en/view/statistics

Fecha de recolección de datos: 31/08/2020.

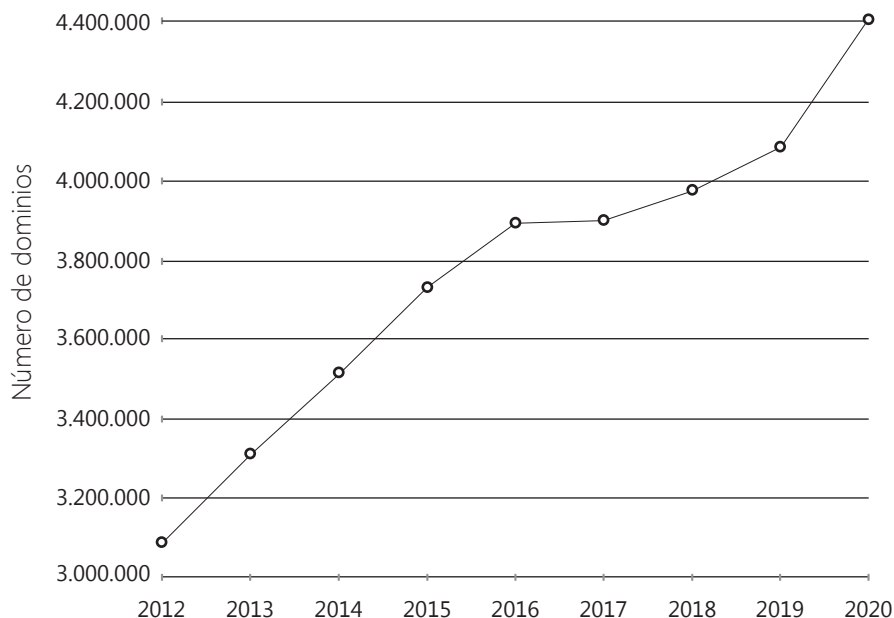
<sup>39</sup> Grupo de las 19 mayores economías del mundo y la Unión Europea. Más información en: <https://g20.org/en/Pages/home.aspx>

<sup>40</sup> La Tabla 1 presenta el número de dominios ccTLD, según las fuentes indicadas. Las cifras corresponden al registro publicado por cada país del G20. En el caso de los países que no presentan o publican estadísticas oficiales proporcionadas por la autoridad de registro de nombres de dominio, las cifras se obtuvieron de: <https://research.domaintools.com/statistics/tld-counts>. Es importante señalar que existen variaciones entre los períodos de referencia, aunque siempre es el más actualizado para cada país. El análisis comparativo del rendimiento de los nombres de dominio también debe tener en cuenta los diferentes modelos de gestión del registro de ccTLD. Además, al observar los rankings, es necesario tener en cuenta la diversidad de modelos de negocio existentes.

## /Panorama Sectorial de Internet

El Gráfico 1 muestra el rendimiento del .br desde 2012.

**Gráfico 1 - NÚMERO TOTAL DE REGISTROS DE DOMINIOS POR AÑO PARA .br - 2012 A 2020\***



\* Datos referidos a agosto de 2020.

Fuente: Registro.br

En agosto de 2020, los cinco dominios genéricos de alto nivel (del inglés, *generic Top-Level Domain* – gTLD) sumaban más de 182 millones de registros. Con 149,28 millones de registros, el dominio .com ocupó el primer lugar, como se muestra en la Tabla 2.

**Tabla 2 - PRINCIPALES GTLDS - AGOSTO 2020**

Posición	gTLD	Dominios
1	.com	149.279.349
2	.net	13.246.836
3	.org	10.240.183
4	.icu	5.618.105
5	.info	4.298.038

Fuente: DomainTools.com

Extraído de: [research.domaintools.com/statistics/tld-counts](https://research.domaintools.com/statistics/tld-counts)

## ¿SABE CÓMO SE REGISTRA LA INFORMACIÓN CLÍNICA DE LOS PACIENTES EN LOS CENTROS DE SALUD BRASILEÑOS? ¿QUÉ DATOS SE RECOGEN EN FORMATO ELECTRÓNICO?

Conozca estos y otros indicadores sobre la producción de datos en el sector sanitario<sup>41</sup>, que sirven de materia prima para aplicaciones de Inteligencia Artificial.

92%

DEL TOTAL DE LOS CENTROS DE SALUD UTILIZARON INTERNET EN LOS ÚLTIMOS 12 MESES<sup>42</sup>



DE ESTOS ESTABLECIMIENTOS:

82%

DISPONÍAN DE SISTEMAS ELECTRÓNICOS PARA REGISTRAR LA INFORMACIÓN DE LOS PACIENTES

64%  
EN LA INFORMACIÓN CLÍNICA Y DE REGISTRO DE LOS PACIENTES EN LAS HISTORIAS CLÍNICAS SE MANTENÍA EN PARTE EN PAPEL Y EN PARTE EN FORMATO ELECTRÓNICO

18%  
EN SÓLO EN PAPEL

18%  
EN SÓLO EN FORMATO ELECTRÓNICO

<sup>41</sup> Basado en datos de la encuesta TIC Salud 2019, realizada por Cetic.br|NIC.br. Los indicadores se refieren a los 12 meses anteriores a la encuesta. Más información: <https://cetic.br/pt/pesquisa/saude/indicadores/>

<sup>42</sup> Del total de centros de salud.

# ¿QUÉ DATOS SOBRE LOS PACIENTES ESTÁN DISPONIBLES ELECTRÓNICAMENTE EN LOS CENTROS DE SALUD?

**[89%]** DATOS DEMOGRÁFICOS DEL PACIENTE

**[65%]** HISTORIA CLÍNICA O NOTAS SOBRE LA ATENCIÓN

**[64%]** DIAGNÓSTICO Y CONDICIONES DE SALUD

**[54%]** ALERGIAS

**[46%]** VACUNAS DADAS

**[61%]** RESULTADOS DE LAS PRUEBAS DE LABORATORIO

**[53%]** LISTA DE MEDICAMENTOS PRESCRITOS

**[56%]** INGRESO, TRASLADO Y ALTA

**[64%]** PRINCIPALES MOTIVOS QUE MOTIVARON LA ATENCIÓN/CONSULTA

**[51%]** NOTAS DE ENFERMERÍA

**[38%]** INFORMES DE IMÁGENES DE RAYOS X

**[24%]** IMÁGENES DE RAYOS X





# /Créditos

## TEXTO

### ARTÍCULO I

Edson Amaro Jr, Michel Fornaciali,  
Andre Batista, Murilo Gazzola,  
Lívia Paiva da Silva, Diogo F. C. Patrão,  
Marcos Freitas Jr  
(Hospital Israelita Albert Einstein - HIAE)

### ARTÍCULO II

Sacha Alanoca, Nicolas Guetta  
Jeanrenaud, Nyasha Weinberg,  
R. Buse Çetin, Nicolas Mialhe  
(The Future Society)

### INFORME DE DOMINIOS

José Márcio Martins Júnior  
(Cetic.br | NIC.br)

### TRADUCCIÓN AL ESPAÑOL

Programa de Políticas de Ciencia,  
Tecnología e Innovación Oficina  
Regional de Ciencias de la UNESCO  
para América Latina y el Caribe  
(Oficina de Montevideo, Uruguay)

### REVISIÓN DEL ESPAÑOL

Stefania L. Cantoni,  
Javiera F.M. Macaya  
(Cetic.br | NIC.br)

### APOYO A LA EDICIÓN EN ESPAÑOL

Luiza Carvalho  
(Cetic.br | NIC.br)

### MAQUETACIÓN EN ESPAÑOL

Grappa Marketing Editorial

### DISEÑO GRÁFICO E INFOGRAFÍA

Giuliano Galvez, Klezer Uehara,  
Maricy Rabelo  
(Comunicação | NIC.br)

### COORDINACIÓN EDITORIAL

Alexandre Barbosa, Tatiana Jereissati,  
Stefania L. Cantoni  
(Cetic.br | NIC.br)

### AGRADECIMIENTOS

Edson Amaro Jr, Michel Fornaciali, Andre  
Batista, Murilo Gazzola, Lívia Paiva da  
Silva, Diogo F. C. Patrão, Marcos Freitas Jr  
(Hospital Israelita Albert Einstein - HIAE)  
Sacha Alanoca, Nicolas Guetta  
Jeanrenaud, Nyasha Weinberg, R. Buse  
Çetin, Nicolas Mialhe (The Future Society),  
Eneida A. Mendonça (Indiana University  
/Instituto Regenstrief),  
Bernardo Mariano Junior (Organización  
Mundial de la Salud - OMS)



Organización  
de las Naciones Unidas  
para la Educación,  
la Ciencia y la Cultura

cetic.br

Centro Regional de Estudios  
para el Desarrollo de la  
Sociedad de la Información bajo  
los auspicios de la UNESCO

nic.br

Núcleo de Informação  
y Coordinación del  
Punto BR

25 years  
cgi.br

Comité Gestor de  
Internet en Brasil

### CREATIVE COMMONS

Atribución  
NoComercial  
SinDerivadas  
(by-nc-nd)







# POR UNA INTERNET CADA VEZ MEJOR EN BRASIL

CGI.BR MODELO DE GOBERNANZA MULTISECTORIAL

[www.cgi.br](http://www.cgi.br)

[nic.br](http://nic.br) [cgi.br](http://cgi.br)