



**DANE**

**SEN**<sup>™</sup>  
Sistema Estadístico  
Nacional-Colombia

Revisión de  
**REFERENTES  
INTERNACIONALES**



# **DIRECCIÓN DE REGULACIÓN, PLANEACIÓN, ESTANDARIZACIÓN Y NORMALIZACIÓN (DIRPEN)**

## **REVISIÓN DE REFERENTES INTERNACIONALES**

- (1) Buenas prácticas en el uso de la inteligencia artificial (IA) para la producción estadística.**
- (2) Webinar “Descubriendo voces ocultas: procesos inclusivos y prácticos de recopilación de datos LGBTQI+ para el monitoreo de los Objetivos de Desarrollo Sostenible (ODS)”.**
- (3) Evento: Datos en Acción: Conectando con la Ciudadanía a través de los Datos.**

**Febrero 2025**



## CONTENIDO

<b>Introducción.....</b>	<b>5</b>
1.1. <i>Buenas prácticas en el uso de la inteligencia artificial (IA) para la producción estadística .....</i>	<i>7</i>
<i>Resumen .....</i>	<i>7</i>
1.2. <i>Síntesis de hallazgos .....</i>	<i>8</i>
1.3. <i>Revisión de referentes .....</i>	<i>10</i>
1.3.1.    España.....	10
1.3.2.    Canadá.....	11
1.3.3.    Reino Unido .....	14
1.3.4.    Chile .....	15
1.4. <i>Conclusiones .....</i>	<i>16</i>
1.5. <i>Recomendaciones .....</i>	<i>17</i>
Anexo: Índice Latinoamericano de Inteligencia Artificial (ILIA) (CEPAL, 2024) - Comisión Económica para América Latina y el Caribe (CEPAL) .....	17
<b>2. Webinar: Descubriendo voces ocultas: procesos inclusivos y prácticos de recopilación de datos LGBTIQ+ para el monitoreo de los ODS .....</b>	<b>22</b>
<b>3. Evento: Datos en Acción: Conectando con la Ciudadanía a través de los Datos .....</b>	<b>27</b>
<b>Bibliografía.....</b>	<b>34</b>



**LISTA DE TABLAS**

<b>Tabla 1. Buenas prácticas en el uso de la inteligencia artificial (IA) para la producción estadística</b>	<b>8</b>
<b>Tabla 2. Estructura temática del Índice Latinoamericano de Inteligencia Artificial 2024</b>	<b>18</b>



## Introducción

Este informe presenta un análisis prospectivo de mejores prácticas internacionales en el ecosistema de datos, con el fin de generar recomendaciones y apoyar la toma de decisiones estratégicas en el Departamento Administrativo Nacional de Estadísticas (DANE) y el Sistema Estadístico Nacional (SEN). El análisis examina innovaciones en gestión de datos de organizaciones gubernamentales y de investigación, incluyendo oficinas nacionales de estadística, ministerios e institutos especializados. Las recomendaciones derivadas de este estudio se alinean con el Plan Estratégico Institucional del DANE y las agendas de trabajo e investigación prioritarias del SEN, contribuyendo así a fortalecer las áreas técnicas y la coordinación del sistema.

La selección de temas para cada informe se basa en una evaluación de las prioridades y necesidades identificadas a través de consultas directas con los grupos internos de trabajo, directores técnicos y coordinadores de las mesas estadísticas del SEN. La profundidad y el detalle de cada revisión se ajustan a las preguntas clave, el alcance definido y la información disponible, con el fin de proporcionar análisis pertinentes y valiosos para la entidad.

Esta versión del reporte se enfoca en el tema de “Buenas prácticas en el uso de la inteligencia artificial (IA) para la producción estadística”. Además, se presentan las reseñas del webinar titulado “Descubriendo voces ocultas: procesos inclusivos y prácticos de recopilación de datos LGBTIQ+ para el monitoreo de los Objetivos de Desarrollo Sostenible (ODS)” y del evento “Datos en Acción: Conectando con la Ciudadanía a través de los Datos”, realizado en Bogotá los días 12 y 13 de febrero de 2025.

Revisión de  
**REFERENTES  
INTERNACIONALES**

# 1

- **Buenas prácticas en el uso de la inteligencia artificial (IA) para la producción estadística**



## 1.1. Buenas prácticas en el uso de la inteligencia artificial (IA) para la producción estadística

### Resumen

El cambio tecnológico que impone el uso de inteligencia artificial (IA) en diferentes aristas de la vida humana, como el análisis de imágenes, audios, videos, procesamiento de lenguaje natural, servicios robotizados y automatización de procesos (BID, 2020) tiene como principal reto su desarrollo en un marco robusto de buenas prácticas y uso ético de los datos. A lo largo del tiempo, la IA ha permitido crear nuevos algoritmos de aprendizaje automático que se enfocan en aprender a resolver situaciones, por medio de datos y ejemplos preexistentes (BID, 2020).

Gracias al desarrollo de procesadores de bajo costo, se ha logrado impulsar la programación y el entrenamiento de algoritmos que detectan patrones y pueden hacer predicciones. Al realizar una revisión de literatura sobre las diferentes definiciones dadas al término IA, se observó cómo se ha logrado un consenso en su descripción. A continuación, se exponen algunas definiciones que evidencian la evolución del término IA:

- McCarthy et al. (1955) acuñó por primera vez el término de IA a la creación de máquinas inteligentes, especialmente de programas informáticos inteligentes y entendiendo la palabra inteligencia a la parte computacional de la capacidad para alcanzar objetivos en el mundo (McCarthy et. al, 1955).
- En la década de los ochenta la inteligencia artificial se definió como un área de las ciencias de la computación, específicamente de la tecnología de la información y las comunicaciones (TIC) que se caracteriza por el desarrollo de máquinas que simulan el comportamiento humano (Canbek, N.G. y M. E. Mutlu., 2016).
- Para 2022, la OCDE define la IA como *"un sistema basado en máquinas que puede, para un conjunto de objetivos definidos por el ser humano, hacer predicciones, recomendaciones o tomar decisiones que influyen en entornos reales o virtuales"* (OCDE, 2022).
- En 2024, la OCDE replanteó la definición de IA como *"un sistema basado en máquinas que, para un conjunto determinado de objetivos explícitos o implícitos, infiere, a partir de la entrada que recibe, cómo generar salidas de información como predicciones, contenido, recomendaciones o tomar decisiones que pueden influir en entornos físicos o reales"* (OCDE, 2024).
- La definición utilizada por los países de la Unión Europea se establece en la Ley de Inteligencia Artificial de la Unión Europea (2021) por el que se establecen normas armonizadas en materia de



inteligencia artificial y donde se define en su Artículo 3. la IA como *“un sistema basado en máquinas diseñado para funcionar con diversos niveles de autonomía y que puede mostrar capacidad de adaptación tras su despliegue y que, para objetivos explícitos o implícitos, infiere, a partir de la entrada que recibe, cómo generar salidas tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales; (...)”*. Esta Ley definió en su Anexo I. técnicas y estrategias de IA como: estrategias de aprendizaje automático; estrategias basadas en la lógica y el conocimiento, y estrategias estadísticas, estimación bayesiana, métodos de búsqueda y optimización.

Sin embargo, la necesidad de establecer un marco regulatorio en materia de uso ético para fines estadísticos es aún más reciente y requiere un mayor desarrollo por parte de los institutos nacionales de estadística (INE). Para el caso europeo, en 2024 se inició el proyecto de IA y aprendizaje automático para la estadística oficial en la Agenda de Innovación del Sistema Estadístico Europeo (SEE) titulado *“One stop shop for AI-ML for Official Statistics”* financiado por la Eurostat y que contó con la participación de 14 institutos nacionales de estadística. Este proyecto tuvo como objetivo proporcionar un marco de desarrollo de soluciones basadas en la IA y el aprendizaje automático para la producción de estadísticas oficiales, compartiendo los esfuerzos, los recursos y las herramientas en la innovación y la producción (INE E. , 2024).

Este documento, a partir de una robusta revisión internacional, permite identificar las buenas prácticas en el uso de IA por los INE a partir de orientaciones sobre la importancia y el uso de estas tecnologías en la producción estadística, como también identificar los procesos que se encuentran en desarrollo para garantizar el uso de la IA bajo un marco normativo en cuanto a la producción de estadísticas oficiales.

## 1.2. Síntesis de hallazgos

A continuación, se presenta una breve descripción de los principales hallazgos de la revisión de referentes internacionales sobre las buenas prácticas en el uso de la IA para la producción estadística.

**Tabla 1.** Buenas prácticas en el uso de la inteligencia artificial (IA) para la producción estadística

Referente	Buenas prácticas en el uso de la IA para la producción estadística
España	El INE participa en el proyecto de IA y aprendizaje automático para la estadística oficial en la Agenda de Innovación del Sistema Estadístico Europeo (SEE) titulado <i>“One stop shop for AI-ML for Official Statistics”</i> para incorporar la inteligencia artificial y el aprendizaje automático en la producción de estadísticas oficiales que contribuya a la transformación tecnológica, la innovación metodológica y la mejora en la calidad de sus productos.



Referente	Buenas prácticas en el uso de la IA para la producción estadística
	<p>Este proyecto tiene un horizonte de tiempo de cuatro (4) años, en los que se espera contar con acciones en el procesamiento de imágenes satélite y datos de observación terrestre, depuración (detección de errores) e imputación de datos, codificación automática a partir de texto en lenguaje natural, cadenas de suministro a nivel de empresas, uso de grandes modelos generativos de lenguaje y técnicas y aplicaciones de generación de datos sintéticos para la Estadística Oficial.</p> <p>Por otra parte, el INE recopila y publica estadísticas sobre el porcentaje de empresas que emplean tecnologías de IA, así como los diferentes tipos de tecnologías de IA utilizadas. Esta información se recopila desde el primer trimestre de 2021 y permite indagar, por medio de una Encuesta, sobre el uso de tecnologías de IA en las empresas de las diferentes comunidades autónomas de España. Se resalta esta buena práctica del INE, ya que permite utilizar variables para medir el uso de IA en las actividades económicas del país.</p>
Canadá	<p>Statistics Canada ha comenzado a explorar los beneficios potenciales de la IA haciendo hincapié en su uso responsable y ético. La agencia opera de acuerdo con los instrumentos y los marcos rectores que garantizan que la información de los individuos permanezca privada, segura y confidencial. Algunos ejemplos de la aplicación de esta tecnología pueden observarse en el procesamiento de datos, la prestación de servicios, la automatización y el análisis de datos.</p>
Reino Unido	<p>Desde 2019, la Oficina de Regulación de Estadísticas ha estado desarrollando el programa de automatización y ciencia de datos para mejorar la eficacia del trabajo de la oficina y comprender mejor sus actividades. Esto ha facilitado la exploración del entorno de la IA y la determinación del papel de la oficina en la regulación de la IA dentro del gobierno. Dentro de los proyectos del programa se encuentra la automatización de un calendario de publicaciones estadísticas, cuyo objetivo es incorporar todos los calendarios de publicaciones estadísticas para obtener una visión completa de todas las estadísticas oficiales.</p>
Instituto Nacional de Estadísticas de Chile	<p>El proyecto "Codificación automática de clasificadores CIUO-08.CL y CAENES a partir de técnicas de machine learning" (INE CHILE, 2022) tiene como propósito actualizar las bases de entrenamiento y optimizar los algoritmos de codificación para los clasificadores CAENES (Clasificación de Actividades Económicas Nacionales para Encuestas Sociodemográficas) y CIUO-08.CL (Clasificador Chileno de Ocupaciones) utilizando datos recolectados con dispositivos electrónicos. Este proyecto responde a la necesidad de mejorar la eficiencia y la calidad de los datos, reemplazando la práctica manual de codificación que históricamente ha sido implementada por el Instituto Nacional de Estadísticas de Chile para el análisis de respuestas a variables de preguntas abiertas. La codificación manual demanda tiempo y presenta limitaciones</p>



Referente	Buenas prácticas en el uso de la IA para la producción estadística
	en la uniformidad de criterios. Para abordar estos desafíos, se han implementado estrategias automatizadas de codificación basadas en aprendizaje de máquinas, que se han aplicado en diversas encuestas, como la Encuesta Nacional de Empleo (ENE) y la Encuesta Nacional de Seguridad Ciudadana (ENUSC). Sin embargo, los métodos de machine learning pueden implicar problemas de desactualización de datos, lo que compromete la calidad de los resultados. Además, se han puesto a disposición sets de datos de alta calidad y se ha desarrollado una interfaz para facilitar el uso de los modelos entrenados, contribuyendo así a un uso más eficiente de los recursos públicos.

Fuente: DANE a partir de las revisiones de referentes.

### 1.3. Revisión de referentes

En esta sección se presentará de forma sintetizada la revisión de referentes internacionales sobre los resultados de práctica internacional relacionados con el uso de la IA para la producción estadística.

#### 1.3.1. España

El Instituto Nacional de Estadística (INE) de España ha implementado varias prácticas en el uso de IA. Cada tres meses, el INE realiza la “Encuesta sobre el uso de TIC y del comercio electrónico en las empresas” sobre el uso de tecnologías de la información y la comunicación (TIC) y comercio electrónico en las empresas, incluyendo el uso de IA (INE E. , 2024).

El INE recopila y publica estadísticas sobre el porcentaje de empresas que emplean tecnologías de IA, así como los diferentes tipos de tecnologías de IA utilizadas, como: el análisis del lenguaje escrito, la conversión del lenguaje hablado en formato legible por máquina, la generación de lenguaje escrito o hablado, la identificación de objetos o personas en función de imágenes, el análisis de datos (aprendizaje automático) y la automatización de flujos de trabajo. Las aplicaciones enunciadas indican un enfoque sistemático y detallado en la recopilación y el análisis de datos relacionados con la IA (INE E. , 2024).

Además, el INE de España, es uno de los países que participa en el proyecto “*One stop shop for AI-ML for Official Statistics*” financiado por la Eurostat para incorporar la inteligencia artificial y el aprendizaje automático en la producción de estadísticas oficiales que contribuya a la transformación tecnológica, la innovación metodológica y la mejora en la calidad de sus productos.

Específicamente, el INE participa en el proyecto a través de tres subdirecciones generales (Metodología y Diseño de Muestras, Estadísticas Coyunturales y TIC) por medio de diversos paquetes de trabajo y roles. La Subdirección General de Metodología y Diseño de Muestras coordina el paquete de trabajo sobre el



uso de técnicas de imputación que permite los desarrollos metodológicos mediante el uso de técnicas de aprendizaje automático, como también el desarrollo de pruebas de concepto, prototipos y productos viables mínimos (INE E. , 2024).

Para el INE (2025) resulta importante implementar el uso de técnicas de imputación en tres líneas: i) imputación temprana para la producción de estimaciones estadísticas adelantadas que mejoren la oportunidad de la información; ii) imputación más allá de la muestra que permita incrementar el grado de granularidad de la información y disminuir la carga de respuesta, e iii) imputación tras la recogida para mitigar la ausencia de respuesta y mejorar la precisión de la información (INE E. , 2024).

También se resalta la participación del INE en los siguientes procesos:

- Creación de un marco metodológico para el uso de la inteligencia artificial y las técnicas de aprendizaje automático en la producción estadística oficial, que pueda ser implementado en la Eurostat.
- Desarrollo del paquete sobre codificación automática por medio del procesamiento de lenguaje natural, con el fin de desarrollar metodologías automáticas de procesamiento de texto que permitan obtener los códigos de clasificaciones estadísticas de forma masiva.
- Desarrollo de los paquetes de trabajo sobre la construcción de una plataforma tecnológica con herramientas de código abierto para el desarrollo y la ejecución de estas técnicas sobre los modelos de lenguajes grandes y sobre la generación de datos sintéticos (INE E. , 2024) que permita la implementación de la IA a través de una plataforma interactiva.

### **1.3.2. Canadá**

En la medida en que la IA y el aprendizaje automático (ML, por sus siglas en inglés) continúa evolucionando y ofreciendo oportunidades interesantes a las organizaciones para mejorar sus servicios, Statistics Canada ha comenzado a explorar los beneficios potenciales de esta herramienta, no sin hacer énfasis en su uso responsable y ético.

En este sentido, Statistics Canada continúa operando en concordancia con los instrumentos y marcos rectores, mientras garantiza que la información de los individuos permanezca privada, segura y confidencial.

### **IA para el procesamiento de datos**



Algunos ejemplos en los que la IA facilita el procesamiento de datos se mencionan a continuación. Uno de estos es el esfuerzo de modernización en curso en Statistics Canada que se centra en disminuir la carga de respuesta de los agricultores que participan en encuestas agrícolas mediante el uso de métodos innovadores de procesamiento de datos, como la visión por computadora.

Por ejemplo, el Programa de Estadísticas Agrícolas ha estado utilizando el ML supervisado para los siguientes proyectos:

- La División de Ciencia de Datos (DScD) de Statistics Canada completó recientemente un proyecto de investigación para la Serie de Informes sobre Cultivos de Campo (FCRS, por sus siglas en inglés) en el que se exploró el uso de técnicas de aprendizaje automático (más precisamente, técnicas de regresión supervisada) para la predicción temprana del rendimiento de los cultivos. El objetivo del proyecto era investigar si las técnicas de aprendizaje automático podrían utilizarse para mejorar la precisión del método existente de predicción del rendimiento de los cultivos (conocido como el método Baseline).
- Durante el Censo de Agricultura de 2021, Statistics Canada llevó a cabo un proyecto de ML para determinar la superficie total de invernaderos en Canadá utilizando imágenes satelitales.
- Uno de los proyectos de investigación de la División de Ciencia de Datos de Statistics Canada se centra en evaluar la efectividad de una técnica estadística llamada análisis de componentes principales funcionales (FPCA) como método de ingeniería de características para extraer tendencias temporales de datos de series temporales de satélite de Radar de Apertura Sintética (SAR). El propósito de este proyecto es utilizar las imágenes de 2018, 2019 y 2020 para determinar qué cultivos estaban creciendo en tierras agrícolas. Los tipos de cultivos se predicen utilizando imágenes satelitales y la aplicación de redes neuronales. Las nuevas estimaciones del modelo se utilizan luego para actualizar una base de datos, lo que permite a los usuarios finales obtener las estimaciones más recientes durante la temporada de cultivo. Los resultados iniciales muestran que este enfoque es mucho más rápido y reducirá la carga de respuesta a encuestas para los propietarios de granjas, especialmente durante los períodos ocupados del año. Los resultados fueron 95% concluyentes.

### **IA para la prestación de servicios, la automatización y el análisis de datos**

Statistics Canada está utilizando técnicas de IA de las siguientes maneras para proporcionar servicios y para analizar, categorizar y agrupar datos:

*Chatbot del Censo 2026:* actualmente se está desarrollando un nuevo chatbot para el Censo de 2026 con el objetivo de mejorar las comunicaciones y los servicios de apoyo proporcionados a la ciudadanía. Esta herramienta utilizará una combinación de soluciones de ML de código abierto y SaaS (software como servicio) para proporcionar respuestas automáticas, precisas y oportunas (redactadas y revisadas por



expertos de Statistics Canada para garantizar su precisión) a preguntas frecuentes de la ciudadanía y ofrecerá una vía para contactar con un agente en vivo, si es necesario.

*Base de datos canadiense de forenses y médicos examinadores (CCMED por sus siglas en inglés):* esta base de datos recopila información sobre las muertes reportadas a los forenses y examinadores médicos en Canadá. La IA se utiliza para analizar grandes volúmenes de datos y extraer patrones y tendencias que pueden ayudar a identificar y caracterizar peligros de seguridad emergentes y conocidos. así como mejorar la precisión y eficiencia en la identificación de causas de muerte y circunstancias relacionadas lo que permite a los investigadores y profesionales de la salud pública tomar medidas preventivas más efectivas para reducir las muertes evitables.

### **Marcos de gobernanza para el uso de IA**

- Statistics Canada implementa un marco de gobernanza e instrumentos para abordar las oportunidades y desafíos asociados con las nuevas tecnologías y que guían el uso responsable de la IA, incluyendo los siguientes: *Protección de la confidencialidad de acuerdo con la Ley de Estadísticas:* la Ley de Datos e Inteligencia Artificial de Canadá parte del proyecto de ley "C-27", presentada en 2022, tiene como objetivo asegurar el desarrollo responsable de la IA y proteger la confidencialidad de los datos personales. Esta ley establece regulaciones para garantizar que las tecnologías de IA se desarrollen y utilicen de manera ética y transparente, protegiendo la privacidad de los ciudadanos canadienses.
- *Evaluaciones de ciberseguridad:* cada nuevo proyecto que utiliza IA se somete a una evaluación de riesgo de ciberseguridad de todas las aplicaciones, los sistemas y el software que se utilizarán para procesar información sensible y personal. Esta evaluación identifica y gestiona cualquier amenaza, vulnerabilidad y riesgo potencial asociado con procesos, tecnologías (vulnerabilidades de software), personas y documentos.
- *Instrumentos de gobernanza:* Statistics Canada implementa las directivas del Gobierno de Canadá sobre el uso responsable de la IA, incluyendo la Directriz sobre la Toma de Decisiones Automatizada de 2020, la cual establece las normas y principios para el uso de la IA y otras tecnologías automatizadas en la toma de decisiones administrativas. Esta directriz tiene como objetivo garantizar que el uso de estas tecnologías sea transparente, responsable, legal y justo.
- *Comités de gobernanza:* Statistics Canada trabaja con varios mecanismos de gobernanza que aprueban proyectos de recopilación de datos. El Consejo Asesor sobre Ética y Modernización del Acceso a Microdatos complementa el papel orientador desempeñado por el Consejo Asesor de Estadísticas Canadienses.



- *Principios de necesidad y proporcionalidad:* la privacidad y la confidencialidad son principios centrales en cada paso de un proyecto que implique la recopilación de datos. En 2022, se desarrolló el Marco de Necesidad y Proporcionalidad en colaboración con la Oficina del Comisionado de Privacidad de Canadá, en total conformidad con la Ley de Estadísticas y la Ley de Privacidad (Privacy Act), para equilibrar la necesidad de la sociedad de estadísticas oficiales, con la necesidad de reducir la carga de respuesta para los canadienses, todo mientras se protege su privacidad. De esta manera, busca garantizar que cualquier implementación de IA sea necesaria para alcanzar un objetivo legítimo y que los beneficios superen los riesgos y posibles impactos negativos. El Marco se aplica para evaluar y mitigar los riesgos asociados con el uso de sistemas automatizados y algoritmos. Esto incluye considerar aspectos como la privacidad, la equidad, la transparencia, la rendición de cuentas y la calidad de datos.
- *Marco para el uso responsable del ML:* con el fin de guiar el uso ético de los datos procesados mediante técnicas de ML, Statistics Canada ha desarrollado el marco "Uso Responsable del Aprendizaje Automático en Statistics Canada" para programas y proyectos estadísticos que utilizan algoritmos de ML.

### 1.3.3. Reino Unido

La Oficina de Regulación de Estadísticas (OSR, por sus siglas en inglés) es un regulador independiente de estadísticas oficiales producidas en el Reino Unido. Su papel es mantener el Código de Prácticas de Estadísticas (Code of Practice for Statistics), proteger el papel de las estadísticas en el debate público y desarrollar una mayor comprensión del bien público que representan las estadísticas.

En octubre de 2019, la OSR nombró a un jefe de Datos y Automatización que ha estado desarrollando el programa de automatización y ciencia de datos para permitir a la oficina trabajar de manera más efectiva y obtener una mayor comprensión del trabajo que se realiza. De esta manera, se ha abierto el camino para explorar el panorama de la IA y determinar el papel de la oficina en la regulación de la IA dentro del gobierno.

El programa de automatización y ciencia de datos analiza cómo se puede automatizar parte del trabajo que se realiza en la OSR para permitir que los reguladores tengan más tiempo para interactuar con las personas con las que necesitan interactuar.

El primer proyecto que inició el programa de trabajo fue automatizar un calendario de publicaciones estadísticas que informará sobre próximos lanzamientos, publicaciones añadidas o eliminadas y cualquier cambio realizado en las fechas de lanzamiento. El programa obtiene la información de la página de investigación y estadísticas de gov.uk, aunque el objetivo es incorporar todos los calendarios de publicaciones estadísticas para obtener una visión completa de todas las estadísticas oficiales. Este



proyecto resultó ser muy útil durante la pandemia de COVID-19 debido al volumen de nuevas estadísticas que se deben seguir.

Aunque se titula "Automatización y Tecnología", el programa de trabajo abarca trabajo de Ciencia de Datos y Visualización de Datos. Después de que se recopilan datos de la web, se necesitan técnicas de minería de datos para estructurar los datos en algo utilizable. Después de eso, es necesario extraer significado o conocimiento, y una buena manera de hacerlo es mediante el Procesamiento del Lenguaje Natural (PLN) (disciplina dentro de la Ciencia de Datos que se ocupa del análisis de datos de texto).

Durante la pandemia de COVID-19, la OSR se dedicó a responder a las preocupaciones sobre estadísticas engañosas y a colaborar para producir revisiones rápidas de nuevas estadísticas o estadísticas cambiantes. Para ello, el programa de trabajo de Automatización y Tecnología creó un tablero de la red social X que abarcaba todas las técnicas mencionadas anteriormente para permitir observar de qué se estaba hablando en relación con las estadísticas de COVID-19. Funcionaba todos los días y recopilaba los tuits relacionados con un término de búsqueda proporcionado, y luego extraía los tuits para proporcionar los tuits más retuiteados, los hashtags principales, los enlaces populares y otras métricas útiles.

#### **1.3.4. Chile**

El Instituto Nacional de Estadísticas de Chile ha implementado la codificación manual para el análisis de respuestas a variables de preguntas abiertas, una práctica manual intensiva en términos de tiempos y costos y presenta limitaciones en la uniformidad de criterios. Para mejorar la eficiencia y la calidad de los datos, se han implementado estrategias automatizadas de codificación basadas en aprendizaje de máquinas. El proyecto "Codificación automática de clasificadores CIUO-08.CL (Clasificador Chileno de Ocupaciones) y CAENES (Clasificación de Actividades Económicas Nacionales para Encuestas Sociodemográficas), a partir de técnicas de machine learning" (INE CHILE, 2022) busca actualizar las bases de entrenamiento y optimizar los algoritmos de codificación para los clasificadores CAENES y CIUO-08.CL utilizando datos recolectados con dispositivos electrónicos y reduciendo significativamente el tiempo y la discrecionalidad en la asignación de códigos.

Estas técnicas se han aplicado en diversas encuestas, como la Encuesta Nacional de Empleo (ENE) y la Encuesta Nacional de Seguridad Ciudadana (ENUSC). El proyecto se centró en la generación de conjuntos de datos etiquetados tomados de encuestas realizadas mediante entrevistas personales asistidas por algún dispositivo electrónico, utilizando principalmente datos de la Encuesta Nacional de Empleo (ENE) y la encuesta Piloto de Presupuestos Familiares de 2020. Para garantizar la calidad de la codificación, se implementó un sistema de doble etiquetado y auditorías periódicas, que permitió corregir discrepancias y mejorar la precisión de los códigos asignados. Las auditorías finales mostraron una alta coincidencia entre los códigos asignados por el equipo de Servicios Compartidos y la Sección de Nomenclaturas, con una precisión del 91,1% para CIUO-08.CL y del 92,5% para CAENES.



Sin embargo, los métodos de machine learning han implicado problemas de desactualización de datos, lo que compromete la calidad de los resultados y afecta la precisión de las predicciones, por factores como el dinamismo de los fenómenos de estudio o los cambios metodológicos en la recolección de la información, que pueden generar diferencias entre datos de entrenamiento y datos nuevos, lo que compromete el rendimiento del modelo.

#### 1.4. Conclusiones

A partir de la revisión de referentes internacionales se concluye lo siguiente sobre las buenas prácticas en el uso de la IA para la producción estadística:

- Se observa que algunos INE vienen implementando procesos de IA para optimizar sus procesos estadísticos. Sin embargo, la existencia de un marco ético y legal establecido frente al uso de IA para la producción estadística está aún en desarrollo.
- Se resalta el compromiso de los INE y el trabajo articulado, como en el caso del proyecto liderado por la Eurostat, donde más de 14 países de la Unión Europea comparten esfuerzos para incorporar la IA y el aprendizaje automático en la producción de estadísticas oficiales que contribuya a la transformación tecnológica, la innovación metodológica y la mejora en la calidad de sus productos.
- Statistics Canada estableció su compromiso de utilizar la IA de manera responsable y ética. La institución adapta continuamente sus procesos a las oportunidades y los desafíos que plantean las nuevas tecnologías, al tiempo que cumple con los instrumentos y los marcos de gobernanza que aseguran el uso ético de la IA, tales como: protección de la confidencialidad de acuerdo con la Ley de Estadísticas, evaluaciones de ciberseguridad, instrumentos y comités de gobernanza, principios de necesidad y proporcionalidad, principios rectores para la consideración ética y marco para el uso responsable del aprendizaje automático.
- Desde 2019, la OSR ha trabajado en mejorar la eficiencia y la comprensión de sus operaciones mediante un programa de automatización y ciencia de datos. A través de este programa, la OSR abarca técnicas de minería de datos y Procesamiento del Lenguaje Natural (PLN) para estructurar y extraer información útil de los datos recopilados.
- El proyecto del INE de Chile constituye un avance significativo en la automatización de la codificación de datos, a través de la implementación de técnicas de machine learning y la actualización constante de las bases de entrenamiento. Sin embargo, es clave mantener un monitoreo continuo y auditorías que garanticen una alta precisión en la codificación. De esta manera no solo se optimiza el uso de recursos públicos, sino que también se asegura la producción de datos estadísticos de alta calidad, esenciales para la toma de decisiones informadas.



## 1.5. Recomendaciones

Se recomienda a nivel general para generar buenas prácticas en el uso de la IA para la producción estadística lo siguiente:

- Se deben aunar esfuerzos entre los INE de la región para generar espacios que permitan conocer la experiencia que han tenido en el uso de la IA y el aprendizaje automático en la producción de estadísticas oficiales que contribuya a la transformación tecnológica, la innovación metodológica y la mejora en la calidad de sus productos.
- Es fundamental definir un marco ético que refuerce las buenas prácticas en el uso de la IA para la producción estadística.
- A medida que los algoritmos se entrenan con nuevos datos, se recomienda monitorear y ajustar continuamente los modelos para mejorar la precisión. Al ajustarlos regularmente con nuevos datos, se pueden identificar y corregir sesgos, mejorar las predicciones y adaptarse a cambios en el entorno o en los datos.
- Se recomienda fomentar la colaboración entre expertos en estadística, ciencia de datos, ética y otras disciplinas relevantes para abordar de manera integral los desafíos y las oportunidades que presenta la IA.

### **Anexo: Índice Latinoamericano de Inteligencia Artificial (ILIA) (CEPAL, 2024) - Comisión Económica para América Latina y el Caribe (CEPAL)**

El Índice Latinoamericano de Inteligencia Artificial (ILIA) (CEPAL, 2024) es una iniciativa liderada por la Comisión Económica para América Latina y el Caribe (CEPAL) y el Centro Nacional de Inteligencia Artificial de Chile (CENIA). El informe del ILIA se lanzó en agosto de 2023 y presentó datos cuantitativos y cualitativos sobre el estado de avance de la IA en 19 países de América Latina y el Caribe<sup>1</sup> con el fin de identificar logros, brechas y oportunidades de mejora en los ecosistemas de IA para la región. El ILIA está estructurado en torno a tres dimensiones, nueve subdimensiones, 19 indicadores y 68 subindicadores que brindan un panorama de los ecosistemas de IA en el conjunto de países. La estructura temática del ILIA se presenta a continuación:

---

<sup>1</sup> Los países incluidos en la medición del ILIA son: Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, México, Panamá, Paraguay, Perú, República Dominicana, Uruguay y Venezuela.

**Tabla 2.** Estructura temática del Índice Latinoamericano de Inteligencia Artificial 2024

DIMENSIÓN	SUBDIMENSIONES	INDICADORES	SUBINDICADORES
FACTORES HABILITANTES	Infraestructura	Indicador de conectividad	Porcentaje de población que usa Internet, Promedio de velocidad de descarga móvil (Mbps), Implementación de 5G, Cobertura de redes móviles, Hogares con acceso a Internet, Suscripciones activas de banda ancha móvil, Suscripciones de banda ancha fija, Promedio de velocidad de descarga banda ancha fija, Promedio de latencia, Cesta básica de banda ancha fija.
		Indicador de cómputo	Nube, Capacidad de infraestructuras de HPC, Centros de datos certificados, IXP, Servidores de Internet seguros.
		Indicador de dispositivos	Hogares que tienen computadora, Asequibilidad a teléfonos inteligentes, Adopción de IPv6.
	Datos	Indicador de Barómetro de Datos	Disponibilidad, Capacidad, Gobernanza, Uso e impacto.
	Talento humano	Indicador de Alfabetización en IA	Educación temprana en ciencias, Educación temprana en IA, Habilidad en inglés.
		Indicador de Formación profesional en IA	Penetración de habilidades en IA, Licenciados en STEM.
		Indicador de Talento Humano Avanzado	Programas de magíster en IA en universidades del Ranking QS, Programas de doctorado en IA en universidades del Ranking QS, Programas de magíster en IA universidades acreditadas.
INVESTIGACIÓN, INNOVACIÓN Y DESARROLLO	Investigación	Indicador de Investigación	Cantidad de publicaciones en IA, Investigación activa, Productividad de investigadores (as), Impacto de la investigación en IA, Presencia de centros de investigación de IA. Proporción de autoras en esta disciplina, Investigación consistente en IA, Participación en main tracks y side events de conferencias A+.
	Innovación y Desarrollo	Indicador de Innovación	Número de inversiones privadas, Valor total estimado de la inversión privada, Empresas de IA, Empresas unicornio, Gasto en investigación y desarrollo en proporción al PIB, Desarrollo de aplicaciones, Entorno emprendedor.
		Indicador de Desarrollo	Productividad open source, Calidad open source, Cantidad de patentes.
	Adopción	Indicador de Industria	Trabajadores en sectores de alta tecnología, Fabricación de tecnología mediana y alta, Fabricación de tecnología mediana y alta Proporción del valor añadido de fabricación de tecnología mediana y alta, Gobierno digital.
GOBERNANZA	Visión e Institucionalidad	Indicador de Estrategia de IA	Existencia de estrategia, Presencia de institución encargada de ejecución, Mecanismos de evaluación, Mecanismos de coordinación interinstitucional, Ética y gobernanza de la IA, Infraestructura y Tecnología de la IA, Desarrollo de capacidades, Datos, Gobierno digital, Industria y emprendimiento, I+D, Cooperación regional e internacional.



DIMENSIÓN	SUBDIMENSIONES	INDICADORES	SUBINDICADORES
		Indicador de Involucramiento de la sociedad	Participación ciudadana, Metodología stakeholder.
		Indicador de Institucionalidad	Existencia de Institucionalidad.
	Vinculación Internacional	Indicador de Participación en estándares	Participación en ISO.
		Indicador de Participación en organismos Internacionales	Sin subindicador.
	Regulación	Indicador de Regulación en IA	Sin subindicador.
		Indicador de Ciberseguridad	Sin subindicador.
		Indicador de Ética y Sustentabilidad	Sin subindicador.

Fuente: ILIA 2024.

Luego de las mediciones para cada país, el informe del ILIA presentó los hallazgos más importantes:

- **Talento escaso:** aunque la concentración de talento en IA en la fuerza de trabajo de Latinoamérica ha venido en aumento en últimos años, ningún país de la región ha alcanzado los niveles del Norte Global, lo que indica que las brechas entre las dos regiones se mantienen.
- **Alfabetización al alcance de la mano:** la diferencia de competencias en ingeniería de IA también es significativa; no obstante, en términos de alfabetización, algunos países de la región muestran una penetración más alta. Eso indica que promover habilidades en IA es clave para mejorar las oportunidades laborales.
- **El desafío no es solo formar, también retener:** desde 2019, América Latina ha experimentado una fuga de talentos en IA, a excepción de Costa Rica y Uruguay en años específicos. Esto plantea un desafío adicional para los países, que no solo debe formar capacidades, sino también retener a sus profesionales, ya que la región no significa mayor atractivo de talentos en la materia.
- **La importancia de la matriz económica:** las características económicas y políticas públicas de cada país afectan la adopción de IA. Países liberales como Chile, Uruguay y Costa Rica tienen mejores



entornos emprendedores, mientras que otros más industrializados como México y Brasil se destacan en generación de patentes y en empresas emergentes, fuertes en tecnología avanzada.

- **Brecha de género:** la participación de mujeres en IA varía significativamente y en este sentido es crucial aprender de las mejores prácticas para promover la equidad de género y aprovechar las ventajas de disminuir esta brecha.
- **Ventajas de la multidisciplinaria:** el 80% de las publicaciones en IA en la región son multidisciplinarias, siendo la más relevante la medicina clínica y el desarrollo científico.
- **Interés legislativo:** según los resultados del índice, existen 38 iniciativas legales en discusión o aprobadas sobre el uso ético de la IA generativa, ampliando los marcos regulatorios amplios.
- **Sin sentido de urgencia:** a pesar de los avances, las políticas gubernamentales no adelantan iniciativas que comprometan recursos para fortalecer el progreso en IA.
- **Ecosistema de startups:** la creación de startups (o empresas emergentes) en IA es débil y se concentra en pocos países, reflejando la necesidad de fortalecer el apoyo y financiamiento para su escalamiento.

Los resultados del informe han sido insumo para iniciativas como, por ejemplo, la inversión en infraestructura de cómputo con énfasis en IA (Banco de Desarrollo de América Latina) la creación del Grupo de Trabajo por la Ética de la IA (que comenzó en la Cumbre de Santiago) y programas de formación para aumentar la adopción empresarial de la IA, impulsados por el Banco Interamericano de Desarrollo.

Revisión de  
**REFERENTES  
INTERNACIONALES**

# 2.

**Webinar: Descubriendo voces ocultas: procesos inclusivos y prácticos de recopilación de datos LGBTIQ+ para el monitoreo de los Objetivos de Desarrollo Sostenible (ODS).**



## 2. Webinar: Descubriendo voces ocultas: procesos inclusivos y prácticos de recopilación de datos LGBTIQ+ para el monitoreo de los ODS

El webinar “Descubriendo voces ocultas: procesos inclusivos y prácticos de recopilación de datos LGBTIQ+ para el monitoreo de los Objetivos de Desarrollo Sostenible (ODS)”, organizado por la Global Partnership, Colombia Diversa, Red Sin Violencia LGBTIQ+ y DataLat, realizado el 6 de enero de 2025, exploró la importancia de contar con metodologías inclusivas y prácticas para mejorar los datos sobre identidad de género y orientación sexual, fundamentales para el seguimiento de los ODS.

A través de estudios de caso, se identificaron cómo los datos inclusivos pueden reflejar las realidades de las personas LGBTIQ+<sup>2</sup>, y como por medio de estos se puede impulsar cambios en políticas públicas y mejorar la prestación de servicios.

### Contextualización

A principios de 2024, la Global Partnership for Sustainable Development Data, Red Sin Violencia LGBTIQ+ y Colombia Diversa crearon un espacio de diálogo y acción, abierto a todas las personas interesadas en participar en la creación y la implementación de una agenda regional de datos que promueva el uso de la información para garantizar derechos y mejorar la vida de las personas LGBTIQ+.

### Introducción

Esta sesión exploró la importancia de desarrollar metodologías y prácticas accesibles para mejorar la disponibilidad de datos sobre identidad de género y orientación sexual. Estos enfoques son fundamentales para monitorear el progreso hacia los Objetivos de Desarrollo Sostenible (ODS), asegurando que nadie quede excluido del ecosistema de datos. Se analizaron ODS clave como ODS 1: fin de la pobreza, ODS 3: salud y bienestar, ODS 5: igualdad de género y ODS 10: reducción de desigualdades, identificando oportunidades para incluir desagregaciones que reflejen a estas poblaciones y abordando los desafíos específicos que enfrentan las personas LGBTIQ+, como la estigmatización, la discriminación, las violaciones de derechos humanos y las barreras legales.

A través de la presentación de las organizaciones, la sesión demostró cómo las metodologías de datos inclusivos pueden capturar las realidades vividas por las personas LGBTIQ+ e integrarse en los indicadores de los ODS, impulsando cambios políticos significativos y mejorando la prestación de servicios.

### Presentación de las organizaciones invitadas

---

<sup>2</sup> LGBTIQ+: Lesbianas, Gays, Bisexuales, Transgénero, Intersexuales, Queer, y el símbolo (+) representa a todas las personas que no se identifiquen con las letras anteriores.



## 1. Red Sin Violencia LGBTI

La Red fue creada con el apoyo de la organización internacional Diakonia y actualmente está conformada por las siguientes organizaciones: Colombia Diversa (Colombia), CATRACHAS (Honduras), COMCAVIS TRANS (El Salvador), Lambda (Guatemala), LETRA S, Sida, Cultura y Vida Cotidiana (México), ADESPROC Libertad (Bolivia), LIFS (Perú) y Trans Siempre Amigas (TRANSSA) (República Dominicana).

El principal objetivo de la Red es erradicar las violencias contra las personas LGBTIQ+ en la región, en especial aquellas motivadas por su orientación sexual e identidad de género. Para ello, promueve la investigación y la documentación de casos de violencia, además de impulsar acciones que mejoren el acceso a la justicia y garanticen la no repetición de estos hechos.

### Metodología de la red

La red usa como fuente principal de datos los medios de comunicación y, de acuerdo con el acceso a la información en cada país, recurre a reportes de organizaciones sociales, información estatal y aportes directos de personas cercanas a las víctimas.

Los datos generados por la Red (por ejemplo, la tasa de homicidios de personas LGBTIQ+) son contrastados con los datos presentados por los Estados con el objetivo de solicitar a estas mejoras en los sistemas de información estatal.

## 2. Datalat: Datos Abiertos para la Inclusión y el Desarrollo

Datalat es una organización ecuatoriana comprometida con el uso de datos abiertos para promover la transparencia, la toma de decisiones basada en evidencia y el desarrollo inclusivo en Ecuador. Su labor se centra en la generación, el análisis y la aplicación de datos en diversos ámbitos, incluyendo derechos humanos, innovación cívica y equidad social.

Uno de sus enfoques clave es la recopilación y el análisis de información sobre la población LGBTIQ+ en Ecuador, una tarea fundamental para garantizar el reconocimiento de sus derechos y la formulación de políticas públicas efectivas.

Julio López Peña, director de proyectos e innovación y cofundador de la Fundación Datalat, resalta la importancia de contar con datos precisos y representativos de la población LGBTIQ+. La falta de información sobre esta comunidad perpetúa su invisibilidad y dificulta la implementación de estrategias que respondan a sus necesidades.

### ¿Por qué es esencial generar datos sobre la población LGBTIQ+ para DataLat?

Datalat identifica varias razones clave para la recopilación y el análisis de datos sobre la población LGBTIQ+ en Ecuador:



*Visibilización y reconocimiento:* la ausencia de datos oficiales contribuye a la invisibilización de la comunidad LGBTIQ+, lo que impide la identificación de sus problemáticas y la adopción de medidas adecuadas. Generar información precisa permite dimensionar su presencia en la sociedad y reconocer su diversidad.

*Diseño de políticas públicas basadas en evidencia:* la formulación de políticas inclusivas debe sustentarse en información confiable y representativa. Contar con datos desagregados permite a las instituciones públicas y privadas diseñar programas en áreas clave como salud, educación, empleo y acceso a la justicia, garantizando que las acciones gubernamentales sean efectivas y atiendan las necesidades reales de la comunidad LGBTIQ+.

*Seguimiento del cumplimiento de los ODS:* la disponibilidad de datos desagregados por identidad de género y orientación sexual es esencial para medir el avance en los ODS. Estos datos permiten evaluar el impacto de las políticas implementadas y asegurar que la comunidad LGBTIQ+ no quede atrás en los esfuerzos por alcanzar la igualdad y la equidad.

*Medición de la discriminación y la violencia:* sin datos concretos, es difícil dimensionar la magnitud de la violencia y la discriminación que enfrenta la población LGBTIQ+ en Ecuador. La recopilación de información permite visibilizar estos problemas, identificar patrones y diseñar estrategias para erradicarlos.

*Acceso equitativo a servicios y oportunidades:* con una mejor comprensión de la situación de la población LGBTIQ+, las instituciones pueden garantizar un acceso más justo a servicios esenciales, como salud, educación, vivienda y empleo, contribuyendo a reducir brechas de desigualdad.

### **3. Colombia Diversa**

Es la organización líder en la defensa de los derechos de lesbianas, gays, bisexuales y personas trans en Colombia. Su labor abarca la incidencia política, la defensa legal y la generación de información clave para la protección y la promoción de los derechos de la comunidad LGBTI.

Según Nikita Simonne Dupuis-Vargas, investigador de la organización y consultor en derechos humanos, una de las áreas clave en las que Colombia Diversa ha centrado sus esfuerzos es la investigación y la generación de datos sobre la comunidad LGBTIQ+. A través de informes detallados, la organización documenta y analiza problemáticas como la violencia, la discriminación y el acceso desigual a derechos fundamentales. Estos estudios no solo visibilizan las realidades que enfrenta esta población, sino que también sirven como herramientas fundamentales para la incidencia política y la formulación de políticas públicas inclusivas.



## Colombia Diversa y los ODS

En el marco de la Agenda 2030, el trabajo de Colombia Diversa en la recopilación y el análisis de datos es esencial para monitorear y promover el cumplimiento de metas relacionadas con la igualdad de género (ODS 5), la reducción de desigualdades (ODS 10) y la promoción de sociedades pacíficas e inclusivas (ODS 16).

Al proporcionar información precisa y actualizada sobre la situación de la comunidad LGBTIQ+ en Colombia, la organización contribuye a garantizar que las políticas y los programas alineados con los ODS sean verdaderamente inclusivos y respondan a las necesidades específicas de esta población.

Además de su labor en la investigación, Colombia Diversa es reconocida por su impacto en la defensa legal y la incidencia política. Su trabajo ha sido clave en la obtención de avances legislativos y en el reconocimiento de los derechos de las personas LGBTIQ+ en Colombia. A través de litigios estratégicos y acciones de incidencia, la organización ha impulsado cambios significativos en políticas y normativas, contribuyendo a la reducción de la discriminación y la promoción de una sociedad más equitativa.

Gracias a su compromiso con la justicia y los derechos humanos, Colombia Diversa continúa desempeñando un papel fundamental en la transformación social del país, garantizando que las voces y las necesidades de la comunidad LGBTIQ+ sean escuchadas y atendidas.

### Retos en la generación de datos sobre la población LGBTIQ+

A pesar de la importancia de estos datos, surgen diversos desafíos en su recopilación, como:

- Falta de inclusión de preguntas sobre identidad de género y orientación sexual en censos y encuestas oficiales.
- Resistencia social y estigmatización que pueden influir en la disposición de las personas a proporcionar información.
- Escasa coordinación entre el Estado, la academia y la sociedad civil para consolidar un sistema de datos confiable y sostenible.

### Conclusión del evento

El trabajo de organizaciones como Datalat, Colombia Diversa y la Red Sin Violencia LGBTIQ+ es clave para cerrar la brecha de información y promover políticas públicas más inclusivas. La generación de datos sobre la población LGBTIQ+ no solo contribuye a su reconocimiento, sino que también es una herramienta fundamental para garantizar sus derechos y mejorar su calidad de vida.

Revisión de  
**REFERENTES  
INTERNACIONALES**

**3**

● **Datos en Acción:  
Conectando con la  
Ciudadanía a través de los  
Datos**



### 3. Evento: Datos en Acción: Conectando con la Ciudadanía a través de los Datos

El evento "Datos en Acción: Conectando con la Ciudadanía a través de los Datos" organizado por el DANE y el Colaborativo de Datos Ciudadanos de Naciones Unidas, realizado el 12 y 13 de febrero de 2025, surgió como una iniciativa clave en el marco de la Estrategia de Datos Ciudadanos que desarrolla el país, que tiene como objetivo la construcción del Marco de Datos Ciudadanos para Colombia. Este esfuerzo, liderado por el DANE, se desarrolla con el apoyo de aliados estratégicos como el Colaborativo de Datos Ciudadanos de Naciones Unidas, la Global Partnership for Sustainable Development Data, así como la participación de Organizaciones de la Sociedad Civil (OSC), la academia y otras entidades gubernamentales. La visión de esta estrategia es consolidar un ecosistema de datos más incluyente, participativo y representativo, en el que la ciudadanía no solo sea receptora de información, sino también un actor activo en la producción de datos.

La importancia de esta estrategia radica en que el DANE, como Oficina Estadística de Colombia y ente rector del SEN, reconoce la relevancia de la participación de las OSC y la interacción con diversos actores del ecosistema de datos. Esta apertura responde al compromiso del DANE con la producción de datos bajo un enfoque inclusivo, donde las mediciones se realizan por y para los ciudadanos. En este sentido, los Datos Ciudadanos (DC) representan un enfoque transformador al permitir que individuos y comunidades participen activamente en la generación de datos, desempeñando un papel central en la identificación de problemas y la propuesta de soluciones. Este modelo no solo enriquece las estadísticas oficiales, sino que también promueve la inclusión y el empoderamiento al dar voz e influencia a quienes históricamente han sido subrepresentados.

El propósito de trabajar en esta estrategia es visualizar y propiciar la participación de la sociedad civil en toda la cadena de valor de los datos, desde su recolección hasta su análisis y uso. Se busca reconocer los esfuerzos realizados por todos los actores en la recopilación de información, fortaleciendo el proceso de generación de estadísticas oficiales que reflejen de manera más precisa las realidades y necesidades de toda la población colombiana. Este enfoque garantiza que las estadísticas nacionales no solo sean más representativas, sino también más pertinentes y precisas.

El evento del 12 y 13 de febrero de 2025 se consolidó como una plataforma clave para fortalecer las iniciativas de datos ciudadanos en Colombia. Con un enfoque en la implementación del Marco de Copenhague y el desarrollo del Marco de Datos Ciudadanos para el país, esta reunión ofreció espacios participativos, diálogos y talleres prácticos orientados al intercambio de experiencias y mejores prácticas. Además, el DANE presentó el Modelo de Madurez de Datos Ciudadanos, una herramienta creada en la Dirección de Regulación, Planación, Estandarización y Normalización (DIRPEN) para guiar el desarrollo y la integración de los datos ciudadanos en las estadísticas oficiales.



El evento fue un paso significativo en un proceso continuo de colaboración. Al reunir a actores clave del ecosistema de datos, como el DANE, las OSC, la academia, el sector privado y los gobiernos locales, se buscó fortalecer las redes de colaboración multisectorial y fomentar una participación inclusiva en los procesos de datos. Esta sinergia no solo contribuyó a mejorar la calidad y la relevancia de las estadísticas oficiales, sino que también permitirá construir un futuro más justo y equitativo a través de los datos, donde todas las voces sean escuchadas y consideradas en la toma de decisiones.

Para tener un buen entendimiento de la participación de la sociedad civil en el desarrollo, la recopilación, la construcción y la incorporación de información en las estadísticas oficiales, es preciso entender que los intereses y las motivaciones de los ciudadanos están alineados con sus objetivos y no necesariamente representan grandes volúmenes de información, por tal motivo es imprescindible evaluar el lenguaje del marco, su extensión y lograr la inclusión efectiva de la ciudadanía en su desarrollo. Este, además, es un ejercicio de fortalecimiento del ecosistema de datos, resultando un elemento clave para cerrar las brechas existentes en la información ciudadana en el marco de mapeo de necesidades hasta el direccionamiento de políticas, es vital para conocer las estrategias y las perspectivas de la experiencia de los diferentes actores del SEN.

### ***Primer día del evento***

#### **Experiencia e iniciativas de organizaciones con datos ciudadanos**

Se realizó un ejercicio de micrófono abierto, cuyo objetivo fue conocer la experiencia de las organizaciones con datos ciudadanos o las iniciativas que han generado. A continuación, se presenta una relación de experiencias por organización que participaron en la actividad:

*Emisora Comunitaria Víctimas:* la emisora es un medio comunitario que ha trabajado con las víctimas de desplazamiento por conflicto armado, combatiendo la desinformación, se requiere la creación de redes como un mecanismo para fortalecer la información, además del seguimiento a la política pública de víctimas del conflicto. Actualmente tiene una herramienta, el tablero de planes de acción territorial, que recolecta información de conflicto armado.

*Colombia Diversa:* es una organización que lleva 20 años produciendo anualmente el informe de violencias contra personas LGBTIQ+, que busca generar acciones estratégicas y alianzas que deriven en información eficiente y eficaz que sirva a todos los ciudadanos, indican que la información no es conocida por la institucionalidad y la que tiene la institucionalidad no corresponde con las realidades. Se espera que el análisis de datos se lleve a cabo en conjunto con las comunidades.

*Auditoría General de la República:* es un ente de control fiscal que vigila las contralorías del país por medio de los procesos de control, vigilancia y participación ciudadana, ayudando y fortaleciendo las acciones de la ciudadanía en el control social que requiere. Por ello se está promoviendo la territorialización en la información y su dinamismo para vigilar los recursos promoviendo los comités de control social.



*DANE – Estadísticas Vitales Chocó:* el proyecto Partera Vital se desarrolla principalmente en el Chocó bajo la coordinación del DANE y el Fondo de Población de las Naciones Unidas (UNFPA, por sus siglas en inglés) con el apoyo de la Asociación de Parteros y Parteras. Su objetivo es fortalecer el registro de nacimientos en comunidades étnicas, enfrentando desafíos como el subregistro en zonas rurales dispersas debido a dificultades en las vías de acceso.

A pesar de estos retos, el proyecto ha logrado un mayor reconocimiento y empoderamiento de las parteras en el departamento. Para facilitar el proceso, se implementó un formato físico de encuesta de fácil diligenciamiento. Las parteras completan el formulario y lo envían por WhatsApp a la encargada del DANE, junto con imágenes que permiten consolidar la información. Posteriormente, los datos se centralizan en el DANE. El proyecto ha ganado visibilidad a nivel nacional y se espera ampliar su cobertura en el futuro.

*Corporación de Unidad Popular del Occidente Colombiano:* su objetivo es la inclusión del campesinado como autorreconocimiento, en las encuestas del DANE, en un proceso de reingeniería institucional de estado. Los datos del DANE caracterizan a los campesinos, información que sirve para la formulación de la política pública campesina en los departamentos del Cauca y Valle del Cauca. Resaltan que es importante la coordinación interinstitucional, pues una lectura conjunta con las entidades oficiales va a permitir la formulación de política pública campesina y el fortalecimiento de la información que se recolecta.

*Autoridades Indígenas de Suroccidente (AISO):* manifiestan que se sienten excluidos por falta de información pública en las estadísticas nacionales sobre los pueblos indígenas e indican que debe haber una inclusión con una base de datos específica que permita la visibilización en materia de procesos sociales, económicos y ambientales con la información que tienen las propias comunidades para alimentar el proceso y hacer una construcción colectiva.

*Corporación de Sociólogos Universidad Autónoma Latinoamericana:* indica que se cuenta con sistemas de información y diferentes fuentes sesgadas o con información incompleta que no sirven en su totalidad para las políticas públicas y se realizan ejercicios de relacionamiento con comunidades a través de planes de trabajo buscando fuentes de inversión. Tiene por objetivo es la consolidación de una herramienta para hacer diagnóstico para los planes de mejoramiento de las organizaciones de la sociedad civil.

*Fundación Unidas, mujeres construyendo tejido social:* el programa “tertulia ser mujer hoy” se transmite por YouTube todos los martes desde hace 3 años, permite la equidad y el empoderamiento de las mujeres llevando temas clave a la mesa con mujeres de diferentes países, lo que permite la visibilización de los espacios de las mujeres (indígenas, rurales, migrantes, empresarias, profesionales, etc.). Aunque todavía no se tienen datos cuantitativos, a nivel cualitativo se evidencia un impacto positivo en la comunidad incluidos colegios y universidades.



*Unidad de Planeación Minero-Energética (UPME):* consideran que la información debe estar al servicio de las comunidades, por lo que la UPME la ha puesto a disposición en línea a través del Geoportal Sectorial (GPS). Esta plataforma ofrece datos abiertos e información clave para investigaciones, incluyendo datos locales.

Además, han desarrollado un Geovisor exclusivo para La Guajira, que abarca información sobre el sector energético y datos relevantes sobre cementerios y fuentes de agua. Es fundamental llevar la planeación al territorio para comprender cómo las comunidades desean acceder a la información, cuáles son sus necesidades y el alcance de los datos que requieren.

*Instituto Nacional de Medicina Legal y Ciencias Forenses:* realizan la consolidación de información médico legal de hechos violentos y accidentes en los casos que hay orden judicial para realizarlas y en la página de datos abiertos tienen información dispuesta para el público con variables demográficas (sexo, edad, escolaridad, contextos de violencia). Se entrega información veraz y de calidad para que se formulen políticas públicas.

*Mujeres decididas en red:* se requiere integralidad de derechos y los datos presentan cuellos de botella, en relación con la corresponsabilidad entre la sociedad civil y el Estado. Se sugiere el fomento de semilleros.

*Corporación desarrollo y paz, Canal del dique y zona costera:* proyecto desarrollado con UNFPA, que proporciona información sobre pescadores y pesca artesanal de la zona a partir de sus conocimientos y saberes, así como información sobre migraciones enfocada en la niñez, la juventud y las mujeres. Estos datos buscan generar iniciativas transformadoras y genere una ruta visible para la toma de decisiones en políticas públicas. Desarrollaron el proyecto *Tírame el dato* con metodologías innovadoras con jóvenes a través del arte, la pintura y la oratoria.

*Organización niñez ya:* generan información sobre los derechos de Profesional en comunicación social o áreas relacionadas la niñez a partir de fuentes de información oficial que no son oportunas y no están desagregadas por municipio. Proponen fichas técnicas e indicadores sobre la información faltante.

*Conferencia nacional de organizaciones afrocolombianas:* agrupa más de 300 organizaciones afrocolombianas realizando incidencia política y legislativa, así como investigación en participación política identificando brechas étnicas y de género. Manifiestan que se requiere litigio estratégico para la inclusión étnico racial en censos y registros administrativos de la población afrocolombiana buscando la garantía de derechos y la visibilización de las problemáticas de la población. Su objetivo es generar una inclusión real y efectiva, para ello, han llegado a levantar su propia información para demostrar a las instituciones que sí es posible contar con datos oportunos y en esa misma vía es posible refutar datos que no corresponden.



*Latinfoods*: es una organización que usa datos abiertos y aplica modelos analíticos, machine learning, ciencias de datos y redes neuronales. Ha creado el término gobierno de precisión, que significa que los gobernantes pueden tomar decisiones precisas, con base en los datos, para enfocar los recursos en las variables claves para resolver problemas. Resalta la importancia de toma datos a partir de problemas comunes para responder y desarrollar modelos analíticos. Resaltan las conexiones entre las organizaciones como punto clave para avanzar en la utilización de datos ciudadanos.

*Consejo Nacional de Datos de Colombia*: con 15 años de experiencia científica y académica en Latinoamérica, es una empresa y ONG que realiza investigación científica y técnica de temas de los datos de DD.HH. y seguridad, así como sobre la soberanía digital de los países. Enfatiza en la regulación de la utilización de los datos evitando el uso indiscriminado por parte de grandes empresas extranjeras porque la data de las comunidades se basa en la planificación de los territorios.

### **Retos y oportunidades para avanzar en datos ciudadanos en Colombia.**

Los panelistas invitados fueron Heather Page (División de Estadística de Naciones Unidas), Fredy Vanegas (Corporación Universo LGBTI), Luz Stella Cuero (Asociación de Parteras Unidas del Pacífico - ASOPARUPA), Federico Parra (WIEGO) y Harrison Cuero (Vicepresidencia de la República). Los principales aprendizajes del panel fueron:

- La producción de datos maneja poder y tiene un propósito fundamental, servir de base para la integración social y el desarrollo de políticas que beneficien a todos los ciudadanos. Es importante la articulación de saberes para la recolección de información.
- Existe tensión entre la academia y las comunidades, ya que aún no se le otorga el mismo valor a los datos que produce la comunidad. Por ello los datos comunitarios necesitan una arquitectura para que tengan una validación para su inclusión en los sistemas de registro; para esto se requiere presupuesto y apoyo técnico.
- Aunque se ha avanzado en la visibilización del trabajo de la partería aún existen niños que no han sido registrados y no se reconoce el aporte de la partera en el proceso; la comunidad se siente vulnerada porque a veces no son tenidos en cuenta para construcción social desde sus características y necesidades.
- En lugares como Chocó, el levantamiento de información es complejo por problemas de orden público, la geografía, las costumbres y la interacción con el Estado. Es importante la participación de las comunidades en el diseño y la recolección de los datos. El reto para construir políticas públicas efectivas para las comunidades del Pacífico colombiano es la obtención de datos oportunos y de calidad.
- Se hizo necesario el autorreconocimiento de los afrodescendientes en los censos y se hace necesario construir redes para la construcción de políticas.

### **Segundo día del evento**



El segundo día del evento estuvo enfocado en la importancia de la transparencia, el acceso a la información y la utilización de datos abiertos para fortalecer la toma de decisiones y la interacción entre las instituciones y la ciudadanía. Durante la jornada, se abordaron estrategias para mejorar la accesibilidad y comprensión de la información estadística, permitiendo que diversos actores, desde el sector público hasta la sociedad civil, puedan interpretar y utilizar los datos de manera efectiva. Se destacó la necesidad de que la estadística oficial no solo sea recopilada, sino también procesada y comunicada de forma clara y así garantizar que la ciudadanía cuente con herramientas adecuadas para evaluar y monitorear las políticas públicas con base en evidencia.

Uno de los ejes centrales de la discusión fue el papel de la **alfabetización estadística y digital** en la democratización de la información. Se resaltó que, si bien el acceso a datos ha mejorado gracias a las plataformas digitales, aún existen brechas significativas en su interpretación, lo que limita su aprovechamiento. En este sentido, expertos presentaron iniciativas orientadas a mejorar la divulgación de información mediante herramientas interactivas, visualizaciones dinámicas y narrativas de datos diseñadas para facilitar su comprensión. Asimismo, se subrayó la importancia de capacitar a distintos sectores de la sociedad, desde la ciudadanía hasta los tomadores de decisiones, en el uso adecuado de la información estadística, con el propósito de prevenir la desinformación y fomentar análisis rigurosos y fundamentados.

Otro tema de relevancia fue el impacto de los **datos en la formulación de políticas públicas y la asignación eficiente de recursos**. Se enfatizó que los datos abiertos pueden fortalecer la participación ciudadana al brindar mayor transparencia y fomentar la rendición de cuentas, permitiendo que la sociedad evalúe el impacto de las políticas implementadas. Además, se presentaron herramientas que facilitan la consulta y el análisis de datos en tiempo real, lo que contribuye a una respuesta más ágil y eficaz ante problemáticas emergentes.

Finalmente, el evento destacó la importancia de la **colaboración interinstitucional entre entidades gubernamentales, el sector académico, el sector privado y la ciudadanía** para consolidar un ecosistema de datos sólido y confiable. Se enfatizó que la información estadística debe ser considerada un bien público accesible y libre de barreras, de manera que su adecuada interpretación fortalezca la confianza en las instituciones. En este sentido, se hizo un llamado a la innovación continua en los procesos de recolección, análisis y divulgación de datos, con el objetivo de garantizar que su uso contribuya al desarrollo sostenible y a la toma de decisiones informadas en todos los ámbitos de la sociedad.

### **Conclusiones del evento:**

- Ante los desafíos de la ciudadanía frente a diferentes fenómenos y situaciones que impactan a sus comunidades, se hace necesario establecer redes y realizar control social a través de la recopilación



y el análisis de datos, asegurando que este proceso sea inclusivo, transparente y articulado con la sociedad civil. Para ello, es fundamental fortalecer la alianza interinstitucional con las comunidades para la construcción colectiva reconociendo las realidades de los contextos y las características de las comunidades.

- La información debe estar al servicio de la ciudadanía utilizando medios como los sistemas de información, datos abiertos del gobierno colombiano y herramientas como los geovisores, lo que permite tener una ciudadanía informada y activa en la toma de decisiones.
- A partir de las iniciativas de las organizaciones civiles para visibilizar la información de sus comunidades se puede fortalecer la información oficial para contribuir con la oportunidad y la desagregación a nivel local, municipal y departamental.
- El uso de modelos analíticos son una herramienta útil para el análisis de datos y la resolución de problemas de información, permitiendo enfocar los recursos en las necesidades priorizadas.
- Se requiere la creación de datos con propósito, que definan claramente la pertenencia y el público objetivo para lograr una incidencia real con la información que se produce.
- Se requiere construir redes propias entre OSC con el objetivo de generar un intercambio efectivo de información, experiencias y tener mayores beneficios a nivel técnico, estadístico y social.

El evento fue un espacio provechoso para reconocer las iniciativas de datos nacientes y en marcha, las cuales se fortalecen con las alianzas realizadas entre organizaciones sociales y su colaboración con otros actores los cuales potencian el desarrollo de estas importantes iniciativas.



## Bibliografía

- BID, B. I. (2020). Obtenido de <https://ie42003cgalbarracin.edu.pe/biblioteca/LIBR-NIV331012022134652.pdf>
- Canbek, N.G. y M. E. Mutlu. (2016). On the Track of Artificial Intelligence: Learning with Intelligent Personal Assistants. *Journal of Human Sciences*. 13(1), 592-601.
- CEPAL, C. E. (2024). *Índice Latinoamericano de Inteligencia Artificial*. Obtenido de <https://indicelatam.cl/>
- INE. (2025). Obtenido de [https://www.ine.es/ss/Satellite?L=es\\_ES&c=INEmasNoticia\\_C&cid=1259954337492&idp=1254736090719&pagename=masINE%2FmasLayout](https://www.ine.es/ss/Satellite?L=es_ES&c=INEmasNoticia_C&cid=1259954337492&idp=1254736090719&pagename=masINE%2FmasLayout)
- INE CHILE, I. N. (Octubre de 2022). *www.ine.gob.cl*. Obtenido de [www.ine.gob.cl](http://www.ine.gob.cl): <https://www.ine.gob.cl/docs/default-source/api-codificacion-automatica/codificacion-automatica-de-clasificadores-ciuo-08-cl-y-caenes-tecnicas-de-machine-learning.pdf>
- INE, E. (2024). Obtenido de [https://www.ine.es/ss/Satellite?L=es\\_ES&c=INEmasNoticia\\_C&cid=1259954337492&idp=1254736090719&pagename=masINE%2FmasLayout](https://www.ine.es/ss/Satellite?L=es_ES&c=INEmasNoticia_C&cid=1259954337492&idp=1254736090719&pagename=masINE%2FmasLayout)
- McCarthy et. al. (1955). Obtenido de <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- OCDE. (2022). Obtenido de [https://www.oecd.org/content/dam/oecd/es/publications/reports/2022/11/harnessing-the-power-of-ai-and-emerging-technologies\\_c74e020c/6e76bc18-es.pdf](https://www.oecd.org/content/dam/oecd/es/publications/reports/2022/11/harnessing-the-power-of-ai-and-emerging-technologies_c74e020c/6e76bc18-es.pdf)
- Office for Statistics Regulation. (s.f.). *Regulating artificial intelligence*. Recuperado el 19 de febrero de 2025, de <https://osr.statisticsauthority.gov.uk/regulating-artificial-intelligence/>
- Statistics Canada. (s.f.). *Advisory Council on Ethical, Social and Cultural Concerns* Recuperado el 19 de febrero de 2025, de <https://www.statcan.gc.ca/en/about/relevant/acemma>
- Statistics Canada. (s.f.). *Canadian Statistics Advisory Council* Recuperado el 19 de febrero de 2025, de <https://www.statcan.gc.ca/en/about/relevant/CSAC>
- Statistics Canada. (s.f.). *Collecting your data: Artificial intelligence*. Recuperado el 19 de febrero de 2025, de <https://www.statcan.gc.ca/en/trust/collecting-your-data/artificial-intelligence?form=MG0AV3>



Statistics Canada. (s.f.). *Extracting temporal trends*. Recuperado el 19 de febrero de 2025, de <https://www.statcan.gc.ca/en/data-science/network/extracting-temporal-trends>

Statistics Canada. (s.f.). *Greenhouse detection*. Recuperado el 19 de febrero de 2025, de <https://www.statcan.gc.ca/en/data-science/network/greenhouse-detection>

Statistics Canada. (s.f.). *Integrity*. Recuperado el 19 de febrero de 2025, de <https://www.statcan.gc.ca/en/trust/integrity>

Statistics Canada. (s.f.). *Machine learning*. Recuperado el 19 de febrero de 2025, de <https://www.statcan.gc.ca/en/data-science/network/machine-learning>

Statistics Canada. (s.f.). *Policy on administrative data*. Recuperado el 19 de febrero de 2025, de [https://www.statcan.gc.ca/en/about/policy/admin\\_data](https://www.statcan.gc.ca/en/about/policy/admin_data)

Statistics Canada. (s.f.). *Projects*. Recuperado el 20 de febrero de 2025, de <https://www.statcan.gc.ca/en/data-science/projects>

Statistics Canada. (s.f.). *Protecting your privacy*. Recuperado el 19 de febrero de 2025, de <https://www.statcan.gc.ca/en/trust/protecting-your-privacy>

Statistics Canada. (s.f.). *Yield prediction*. Recuperado el 19 de febrero de 2025, de <https://www.statcan.gc.ca/en/data-science/network/yield-prediction>

Treasury Board of Canada Secretariat. (s.f.). *Policy on Results*. Recuperado el 19 de febrero de 2025, de <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>



En la preparación del reporte de esta edición contó con la participación de:

Alexandra Jane Simpson Silva - [ajsimpsons@dane.gov.co](mailto:ajsimpsons@dane.gov.co)

Catherine Ávila Alvarado – [jcavilaa@dane.gov.co](mailto:jcavilaa@dane.gov.co)

María Paula Diaz Bejarano – [mpdiaz@dane.gov.co](mailto:mpdiaz@dane.gov.co)

María Jimena Vargas Mayo - [mjvargasm@dane.gov.co](mailto:mjvargasm@dane.gov.co)

Yuly Alexandra Mazo Suárez - [yamazos@dane.gov.co](mailto:yamazos@dane.gov.co)

Revisión de estilo por: Sonia Naranjo - [smnaranjom@dane.gov.co](mailto:smnaranjom@dane.gov.co)

Revisión de contenido por: Mauricio Valencia Amaya - [mgvalenciaa@dane.gov.co](mailto:mgvalenciaa@dane.gov.co)

Revisión de contenido por: Iván Rolando Castillo Prieto – [ircastillop@dane.gov.co](mailto:ircastillop@dane.gov.co)

Si tiene dudas comentarios o aportes sobre esta edición por favor no dude en comunicarse al correo:

María Jimena Vargas Mayo – [mjvargasm@dane.gov.co](mailto:mjvargasm@dane.gov.co)

Yuly Alexandra Mazo Suárez - [yamazos@dane.gov.co](mailto:yamazos@dane.gov.co)



@DANE\_Colombia



/DANEColombia



/DANEColombia



@DANEColombia

[www.dane.gov.co](http://www.dane.gov.co)