

 **CAPÍTULO VI**

CONCLUSIONES Y REFLEXIONES DE CIERRE

A lo largo de este libro hemos recorrido un camino que va más allá del entrenamiento de modelos o del uso de librerías específicas. Hemos aprendido que la ciencia de datos comienza mucho antes del modelado y que su verdadero valor reside en la comprensión profunda del dato: su origen, su estructura, su contexto y las decisiones que lo transforman en conocimiento.

Desde los datos estructurados y las series de tiempo, hasta el audio sintético y la detección de deepfakes de voz, cada capítulo ha mostrado que no existen soluciones universales ni recetas únicas. Existen, en cambio, procesos rigurosos, preguntas bien formuladas y una responsabilidad técnica y ética frente a los datos que utilizamos y a los sistemas que construimos.

Este recorrido también deja una idea central: un modelo es tan confiable como lo son los datos que lo alimentan y las decisiones que lo preceden. La ingeniería de datos, el análisis exploratorio y la ingeniería de características no son etapas accesorias, sino el núcleo donde se define la calidad, la interpretabilidad y el impacto real de cualquier solución basada en inteligencia artificial.

En un contexto donde los sistemas automáticos tienen cada vez mayor influencia sobre la sociedad, la ciencia de datos exige algo más que habilidades técnicas. Exige criterio, conciencia y una mirada crítica capaz de cuestionar resultados, reconocer sesgos y comprender las implicaciones del uso de la tecnología. Formar ingenieros y científicos de datos no consiste únicamente en enseñar a programar modelos, sino en cultivar una forma de pensar con datos.

Este libro cierra como comenzó: con una invitación. A explorar con curiosidad, a construir con rigor y a investigar con responsabilidad. Que cada lector continúe este camino entendiendo que los datos no son un fin en sí mismos, sino un medio para comprender mejor el mundo y aportar soluciones que sean, al mismo tiempo, técnicamente sólidas y humanamente conscientes.



REFERENCIAS

1. Ballesteros, D.; Moreno, J.
Highly transparent steganography model of speech signals using Efficient Wavelet Masking.
Expert Systems with Applications, vol. 39, **2012**.
Disponible en: <https://doi.org/10.1016/j.eswa.2012.02.066>
2. Ballesteros, D.; Moreno, J.
On the ability of adaptation of speech signals and data hiding.
Expert Systems with Applications, vol. 39, **2012**.
Disponible en: <https://doi.org/10.1016/j.eswa.2012.05.027>
3. Ballesteros, D.
COVID-19 Bogotá: código en Python para análisis de datos.
Video de YouTube, **2020**.
Disponible en: <https://youtu.be/kezqpilbKyA> (último acceso: 19 diciembre 2025).
4. Ballesteros L., D. M.; Rodríguez, Y. P.; Renza, D.
H-Voice: Fake voice histograms (Imitation + DeepVoice).
Mendeley Data, V4, **2020**.
Disponible en: <https://data.mendeley.com/datasets/k47yd3m28w/4>
(último acceso: 16 enero 2025).
5. Ballesteros, D. M.; Rodríguez, Y.; Renza, D.
A dataset of histograms of original and fake voice recordings (H-Voice).
Data in Brief, vol. 29, 105331, **2020**.
Disponible en: <https://www.sciencedirect.com/science/article/pii/S2352340920302250>

6. Ballesteros, D.; Rodríguez, Y.; Renza, D.; Arce, G.
Deep4SNet: deep learning for fake speech classification.
Expert Systems with Applications, vol. 184, **2021**.
Disponible en: <https://doi.org/10.1016/j.eswa.2021.115465>
7. Ballesteros, D.
Imitation-using-Signal-Processing.
Repositorio de GitHub, **2023**.
Disponible en: <https://github.com/doramariaballesteros/Imitation-using-Signal-Processing> (último acceso: 19 diciembre 2025).
8. Beltrán, M.; Ballesteros L., D. M.
Fake Audio Dataset (ElevenLabs & Respeecher).
Mendeley Data, V1, **2025**.
Disponible en: <https://data.mendeley.com/datasets/79g59sp69z/1>
(último acceso: 16 enero 2025).
9. González, B.; Ballesteros L., D. M.
TTS/V2V Audio Deepfake Dataset.
Mendeley Data, V2, **2025**.
Disponible en: <https://data.mendeley.com/datasets/h4zbs27tkr/2>
(último acceso: 16 enero 2025).
10. Pachón. C.; Ballesteros, D.
FakeVoiceFinder
Repositorio de GitHub, **2025**.
Disponible en: <https://github.com/DEEP-CGPS/FakeVoiceFinder>
(último acceso: 16 enero 2026).
11. Pachón, C.; Ballesteros, D.
FakeVoiceFinder: An Open-Source Framework for Synthetic and Deepfake Audio Detection.
Big Data and Cognitive Computing, vol. 10, no. 1, 2026.
Disponible en: <https://doi.org/10.3390/bdcc10010025>

