Instituto de Física de Cantabria

CSIC Consejo Superior de Investigaciones Científicas

UC Universidad de Cantabria

Jornadas sobre Información de Biodiversidad y Administraciones Ambientales

19/11/2025

Evolución de la IA en el contexto de la Biodiversidad

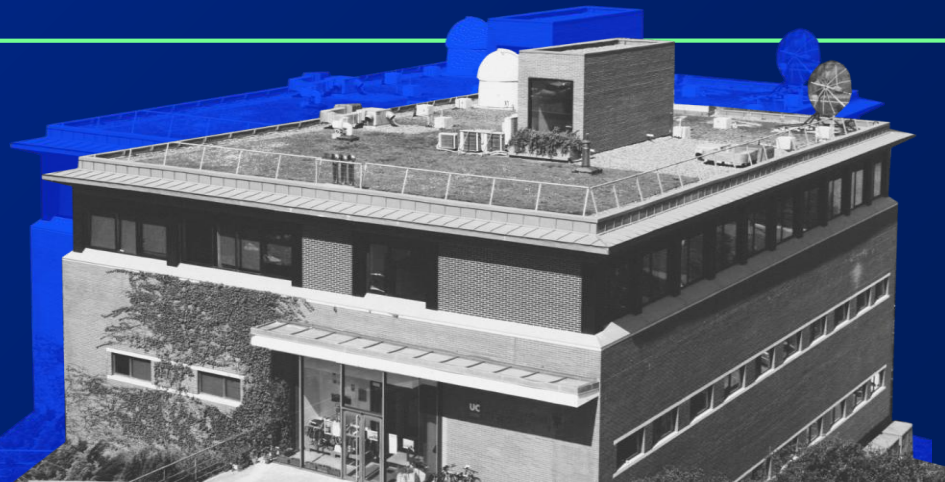
(más allá de ChatGPT)

Fernando Aguilar Gómez

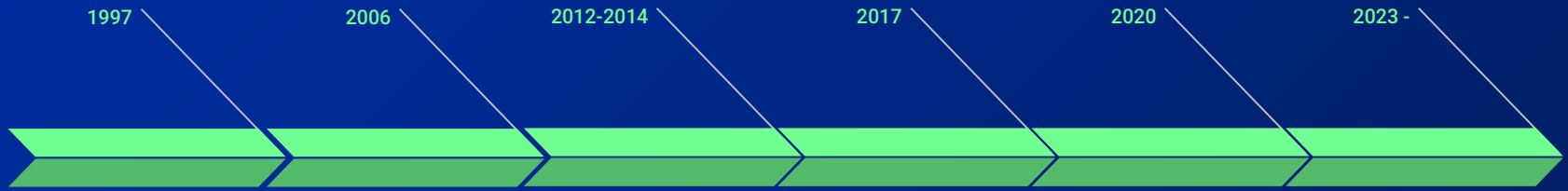
aguilarf@ifca.unican.es

Advanced Computing and e-Science Group

https://advancedcomputing.ifca.es

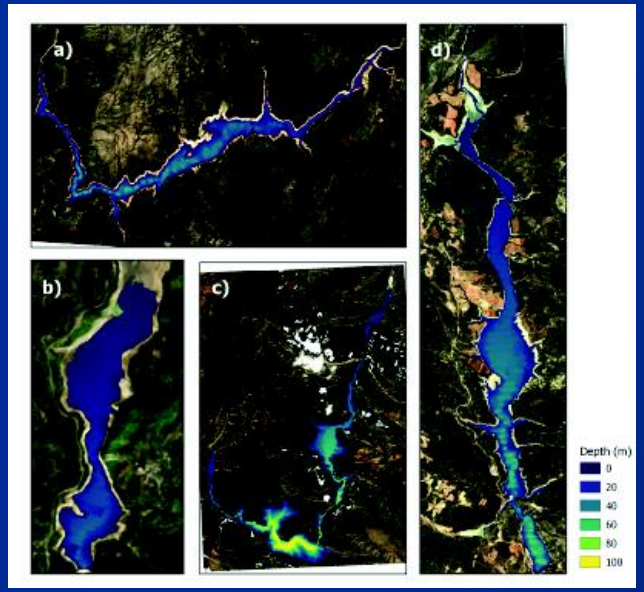


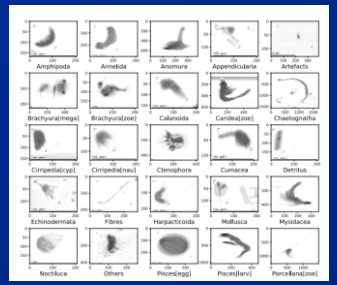
Evolución IA

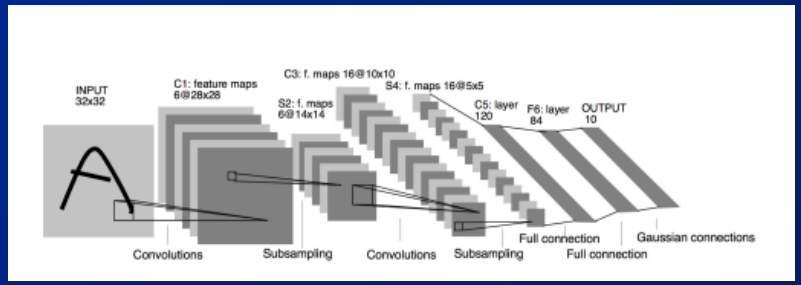


LSTM	Deep Learning	CNN - GANS	Transformers	LLMs	Foundations
Memoria	Multicapas	Convolucionales	Atención	GPTs	Multimodales
		Adversarias	Sin rec. o conv.(+ datos x -)		+Agentes

En capítulos anteriores…







Los nuevos asistentes cognitivos aceleran (y mejoran?) la investigación

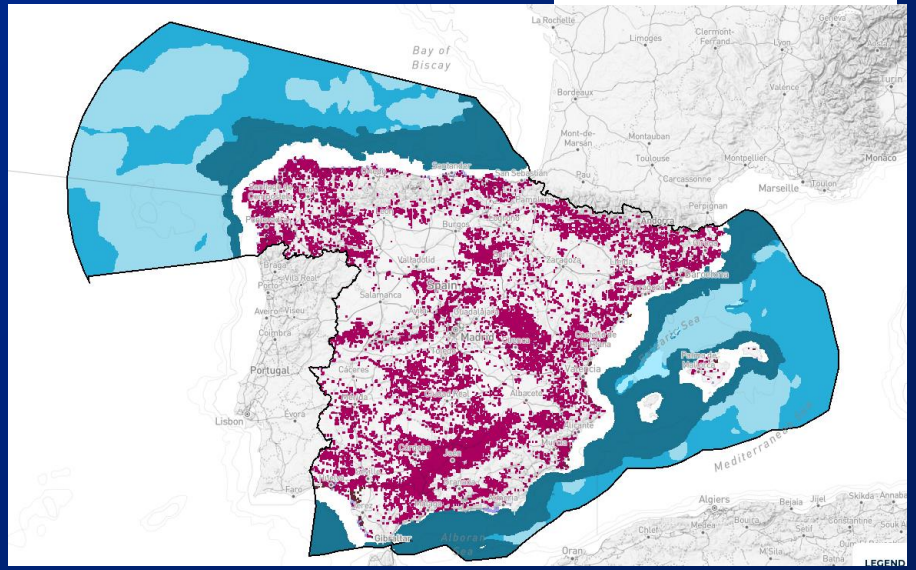
Lo que hay detrás

- **Transformers**: arquitectura central basada en atención, capaz de entender contexto a gran escala.
- **Modelos fundacionales**: entrenados con billones de ejemplos → reutilizables para miles de tareas científicas.
- **Multimodalidad**: combinan texto, PDFs, figuras, tablas, imágenes y datos.
- **RAG**: buscan información en tus documentos y la integran en las respuestas.
- **Fine-tuning**: se adaptan a dominios como biodiversidad, clima, genómica o teledetección.
- **Escalabilidad**: corren en clusters masivos con GPUs/TPUs → análisis que un humano no puede realizar.

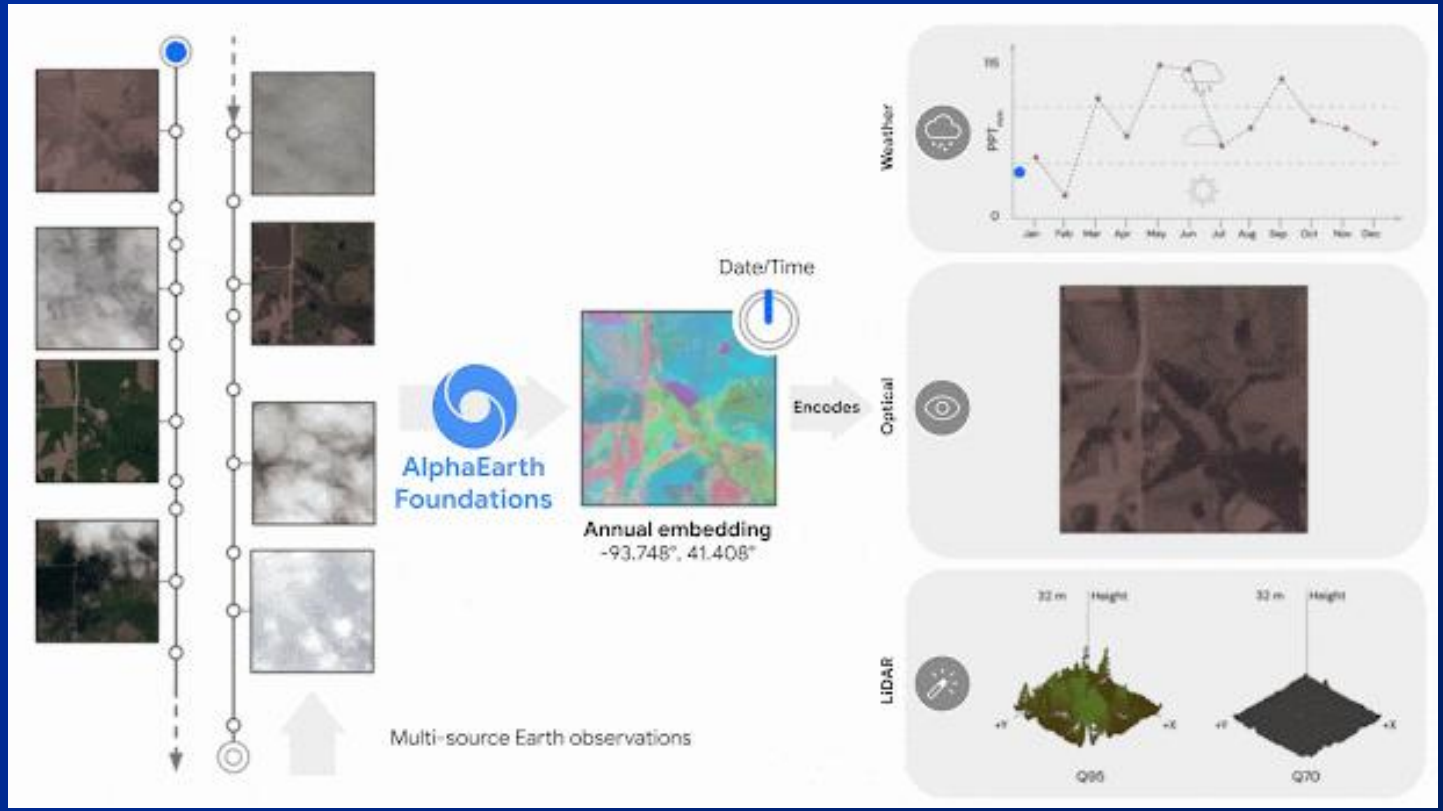
Alpha Earth Foundation

- Modelo fundacional Geoespacial
- Entrenamiento masivo
- Embeddings de 64 valores
- Clasificación, Detección
- Distribución de especies
- Estimación variables

GLOBAL ECOSYSTEMS ATLAS



Integrando Petabytes…



El tratamiento e integración de (buenos) datos es la clave

Utilizando un modelo fundacional. El trabajo “pesado” ya está hecho

Ejemplo: Distribución especie invasora

- Descargas puntos de presencia de GBIF.
- Descargas embeddings AlphaEarth para 2017-2024.
- Creas modelo de presencia/ausencia simple.
- Predices mapa riesgo de invasión.
- Comparas cómo cambia la idoneidad del hábitat con el tiempo.
- Identificas corredores de dispersión (carreteras, ríos).
- Priorizas áreas para intervención.

Arquitecturas más “clásicas” también funcionan

- Transformers: MUCHOS datos
- Conjuntos masivos auto-supervisados

Perch 2.0 - DeepMind

- Entrenado con casi el doble de datos que la primera versión, incluyendo aves, mamíferos, anfibios y ruido antropogénico.
- Mejor rendimiento y mayor capacidad de adaptación a nuevos entornos, incluso submarinos.
- Código abierto y disponible en Kaggle; ampliamente adoptado por la comunidad.
- Ya ha contribuido a descubrimientos relevantes (p. ej., nueva población del Plains Wanderer).

Modelos fundacionales: poderosísimos… pero siguen siendo cajas negras

Aprenden patrones, no leyes
Capturan correlaciones en los datos (clima, suelo, observaciones), pero no entienden, por ejemplo “ecología”: si hay sesgos en los datos, los heredarán.

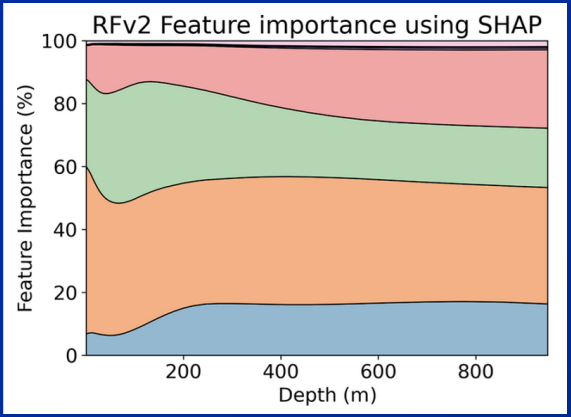
Difíciles de interpretar
Millones de parámetros → no podemos explicar fácilmente por qué una celda tiene alta probabilidad de invasora o por qué un mapa sale como sale.

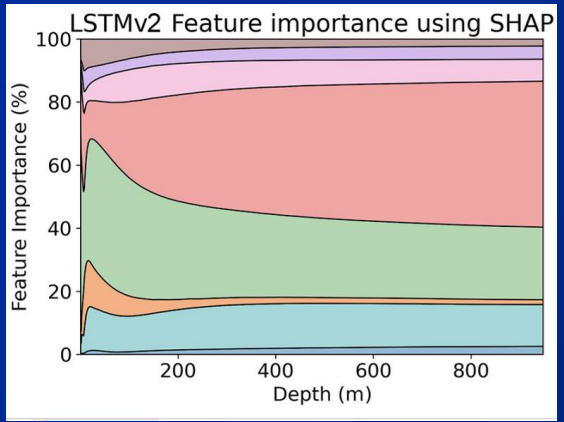
Necesitan validación y contexto experto
Los usamos para **generar hipótesis y priorizar**, pero siempre contrastando con muestreos de campo y conocimiento local.

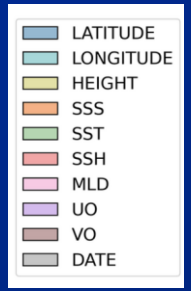
Tres enfoques para una IA explicable

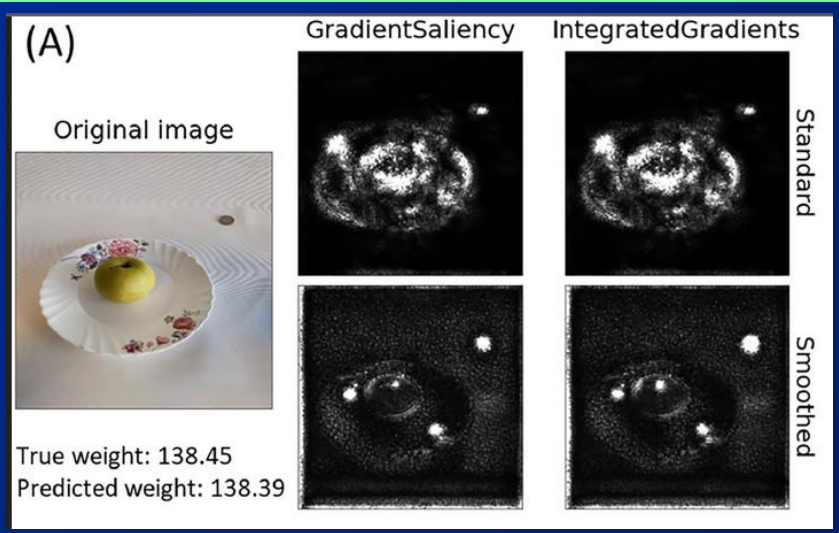
- **Explicabilidad post-hoc**: SHAP, Saliency maps, etc. Se aplican tras entrenar el modelo. Rápidos y flexibles, pero aproximados.
- **Explicabilidad intrínseca**: modelos diseñados para ser transparentes (árboles, modelos modulares). Interpretabilidad real pero menos potencia.
- **Modelos híbridos guiados/informados**: integran procesos ecológicos/físicos y mecanismos interpretables dentro del modelo. Más robustos y alineados con la realidad.

https://doi.org/10.5194/os-21-2579-2025

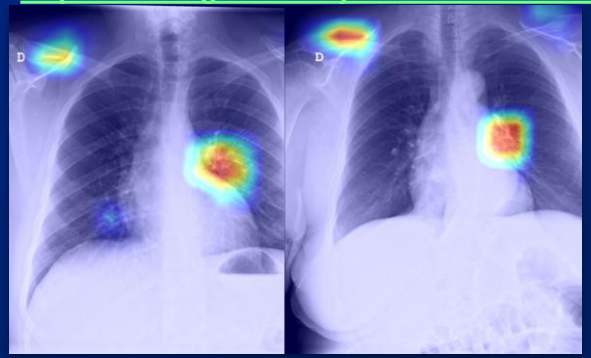








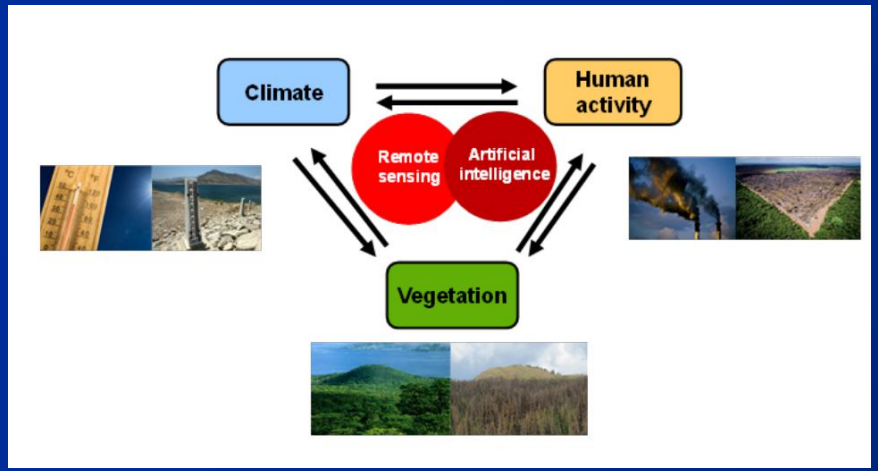
https://doi.org/10.1016/j.nexres.2025.101069



PROFECIA

PID2023-152114OB-I00

Centre d'Investigacions sobre Desertificació (CIDE)

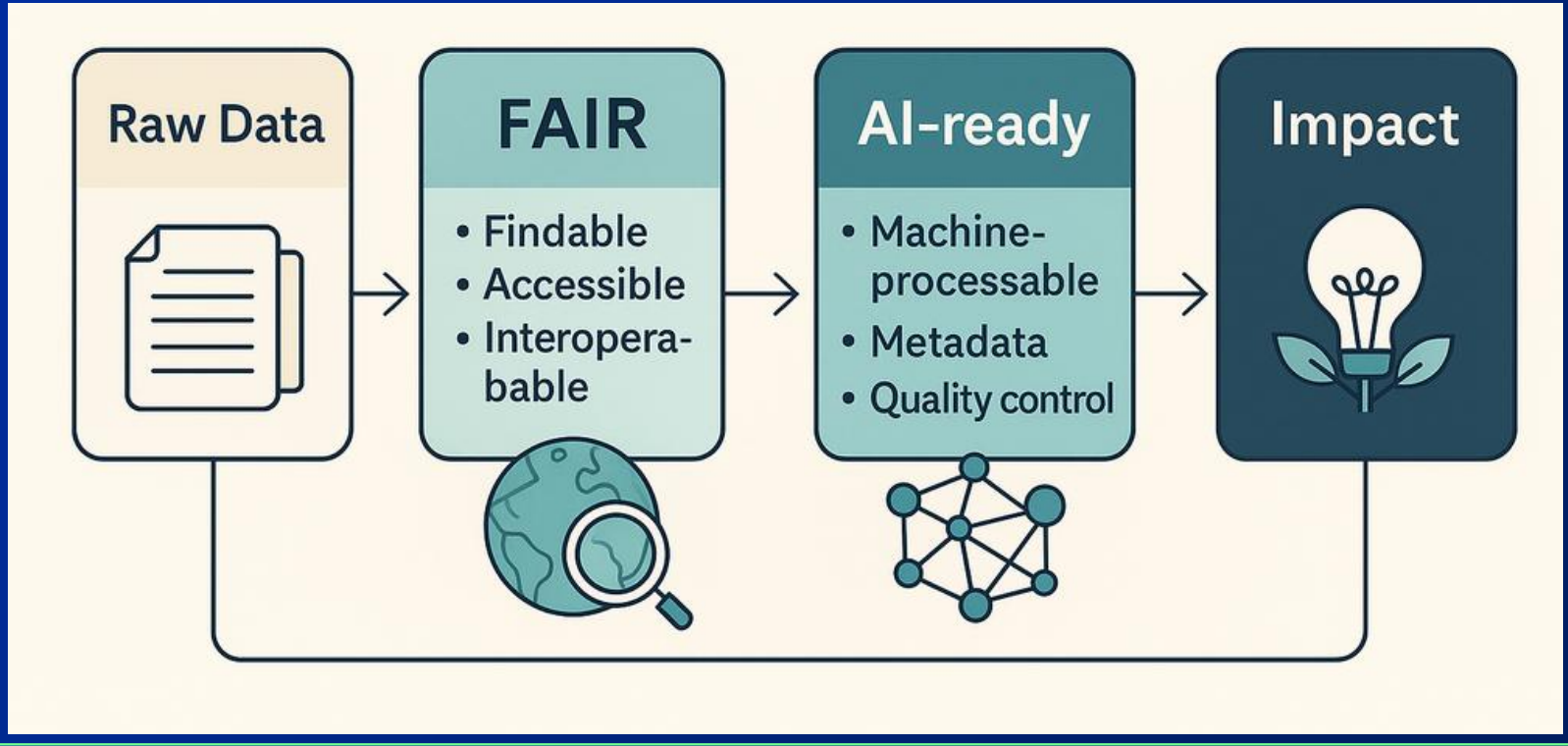


- Facilitando la explotación de gran volúmen de datos en la nube
- FAIR, Reproducibilidad y Explicabilidad
- Diseño modular para configuración de componentes
- Distintos modelos explicables
- Diferentes inputs/outputs

Evaluación de la respuesta de la productividad y fenología de los ecosistemas terrestres al cambio climático mediante teledetección e Inteligencia Artificial

Instituto de Física de Cantabria

Buenos datos, la base para una IA fiable



Qué es EOSC y por qué importa para IA y datos

- EOSC (European Open Science Cloud) es la iniciativa europea para crear un entorno federado donde investigadores puedan publicar, encontrar y reutilizar datos, herramientas y servicios de forma abierta y fiable, a través de una “web de datos y servicios FAIR”.
- Es una federación de nodos (EOSC Federation) que conecta infraestructuras nacionales y temáticas, catálogos de recursos y servicios de computación, con autenticación federada y reglas comunes.

“AI for FAIR data” y “FAIR data for AI”, e introduce el concepto de **AI-ready FAIR research data** como una prioridad central.

EOSC Nodes

- Entidades (nacionales, temáticas o institucionales) que publican datos y servicios FAIR dentro del ecosistema EOSC.
- Funcionan como puentes entre las infraestructuras europeas, nacionales y de investigación.
- Adoptan reglas comunes: metadatos, PIDs, accesibilidad, interoperabilidad y trazabilidad.

Lo que viene: próximos proyectos y convocatorias

- IA y ML para mejorar la FAIRificación y la curación de datos cerca de la fuente.
- Espacios de datos sectoriales
- Armonización de metadatos y esquemas federados.
- Pasar de datos “FAIR” a datos “FAIR y automáticamente explotables por IA”.
- Capacidades para entrenar modelos pan-europeos (foundation models científicos, gemelos digitales sectoriales, etc.).

Instituto de Física de Cantabria

CSIC Consejo Superior de Investigaciones Científicas

UC Universidad de Cantabria

Jornadas sobre Información de Biodiversidad y Administraciones Ambientales

19/11/2025

Los modelos de IA son tan buenos (o malos) como los datos que los entrenan

Fernando Aguilar Gómez

aguilarf@ifca.unican.es

Advanced Computing and e-Science Group

https://advancedcomputing.ifca.es