

Memoria General del Proyecto AI.IDEA

Hackathon Soluciones GenAI para la Biodiversidad
Programa Nacional de Algoritmos Verdes (PNAV)

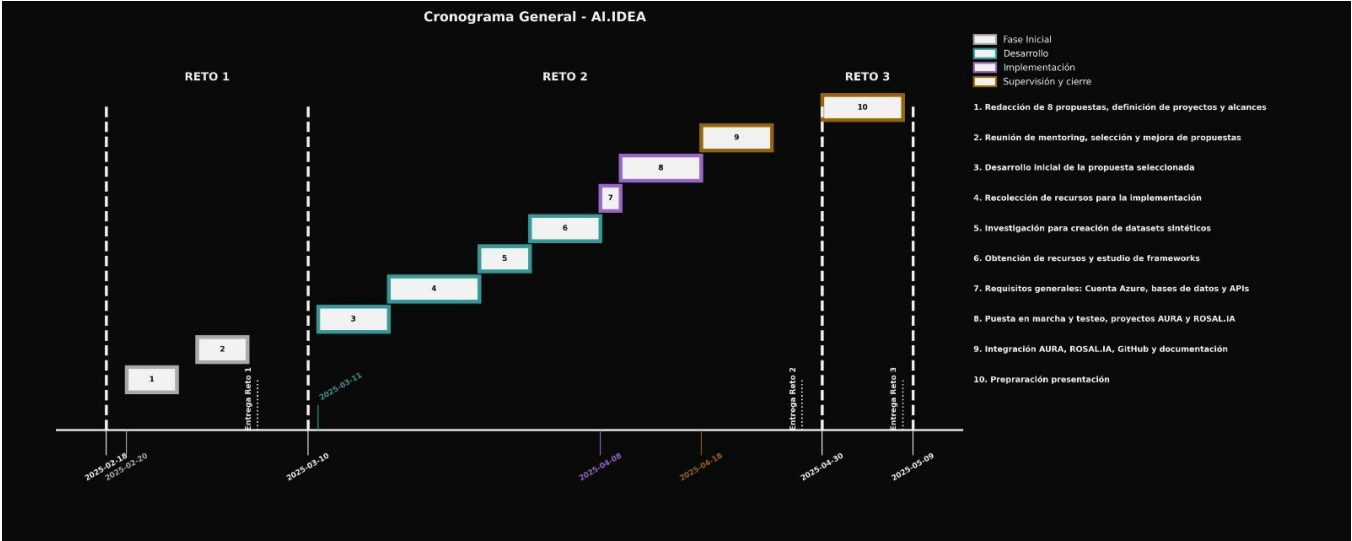


Tabla de contenido

Memoria General del Proyecto AI.IDEA	1
1. Resumen Ejecutivo	1
1.1 Ficha técnica del proyecto AI.IDEA	2
2. Contexto y Justificación	3
3. Equipo de Trabajo	3
4. Desarrollo por Fases	4
Reto 1: Generación y Selección Inicial de Propuestas (18/02/2025 – 10/03/2025)	4
Reto 2: Desarrollo y Consolidación de Propuestas (10/03/2025 – 28/05/2025)	5
Reto 3: Presentación final y cierre (05/06/2025 – 25/06/2025)	7
5. Referencias	8
6. Material complementario	8
6.1 Conjunto de ideas preliminares	8

1. Resumen Ejecutivo

El proyecto **AI.IDEA** nace en el marco del **Hackathon “Soluciones GenAI para la Biodiversidad”**, una iniciativa impulsada por el [Programa Nacional de Algoritmos Verdes \(PNAV\)](#), con el objetivo de fomentar el desarrollo de soluciones tecnológicas basadas en inteligencia artificial generativa aplicadas a la sostenibilidad ambiental y la conservación de la biodiversidad.

Este hackathon nacional está estructurado en tres retos consecutivos, que abarcan desde la ideación y validación inicial hasta la consolidación técnica y la presentación final ante un jurado especializado. El equipo desarrolló inicialmente ocho ideas innovadoras, de las cuales tres fueron seleccionadas para avanzar hacia su desarrollo técnico.

Durante la segunda fase del hackathon, se consolidaron dos soluciones principales bajo el paraguas del proyecto AI.IDEA: **AURA** (Análisis de Uso de Recursos de Acontecimientos), una herramienta para medir y compensar la huella de carbono de eventos, y **ROSALIA** (Repository Of Scientific Articles on Listed Species), un repositorio inteligente de literatura científica sobre especies catalogadas.

Ambas líneas se desarrollaron de forma paralela, incorporando criterios de eficiencia energética, viabilidad técnica y aplicabilidad en escenarios reales.

Esta memoria recoge el proceso completo del proyecto, desde su gestación hasta su presentación final, con especial énfasis en la gestión colaborativa, el uso de herramientas tecnológicas como Azure y CodeCarbon, y el aprendizaje continuo en diferentes ámbitos: técnicos, estratégicos y organizativos.

Reto 1: Generación y Selección Inicial de Propuestas (18/02/2025 – 10/03/2025)	4
Reto 2: Desarrollo y Consolidación de Propuestas (10/03/2025 – 28/05/2025)	5
Reto 3: Presentación final y cierre (05/06/2025 – 25/06/2025)	7

AI.IDEA representa un ejemplo concreto de cómo la inteligencia artificial puede utilizarse para generar soluciones efectivas, escalables y alineadas con los objetivos de conservación ambiental.

1.1 Ficha técnica del proyecto AI.IDEA

Nombre del proyecto	AI.IDEA (Artificial Intelligence for Digital Environmental Applications)
Líneas de desarrollo	AURA (Análisis de Uso de Recursos de Acontecimientos)
	ROSALIA (Repository Of Scientific Articles on Listed Species)
Duración del proyecto	18 de febrero – 25 de junio de 2025
Fases del hackathon	Reto 1: Definición de Casos de Uso
	Reto 2: Desarrollo Caso de Uso
	Reto 3: Desarrollo del Business Case
Equipo participante	Fernando Hurtado Bocanegra, Sergio Díez López, Adrián Ramos González y Miguel Loredó Valdés
Mentoría asignada	Mónica Puebla Estrada

Tecnologías utilizadas	Microsoft Azure, Python, CodeCarbon, APIs de biodiversidad..., herramientas de IA generativa
Organización convocante	Programa Nacional de Algoritmos Verdes (PNAV)
Enlace oficial	algoritmosverdes.gob.es/hackathon
Repositorio GitHub	https://github.com/AEDI-IA/Ai.dea?tab=readme-ov-file#aiidea

2. Contexto y Justificación

La pérdida de biodiversidad y el cambio climático demandan soluciones tecnológicas innovadoras para una gestión ambiental sostenible. En este contexto, el Programa Nacional de Algoritmos Verdes (PNAV) lanzó el hackathon “Soluciones GenAI para la Biodiversidad”, que impulsa proyectos que combinan inteligencia artificial generativa y criterios ambientales.

El proyecto AI.IDEA surge para mejorar la gestión del conocimiento ambiental y ofrecer herramientas analíticas prácticas. El equipo abordó retos desde enfoques tecnológicos, científicos y organizativos, enfocándose en la digitalización de informes, la medición de huella de carbono y el acceso a información científica actualizada.

Sus dos propuestas principales aportan soluciones complementarias: AURA, identifica y mide la huella, promoviendo la compensación mediante inversión en áreas afectadas —ecosistemas, fauna y flora, gestión forestal, espacios protegidos y suelos— para mitigar impactos ambientales; ROSALIA fortalece el conocimiento científico, facilitando su acceso abierto y estructurado para apoyar la toma de decisiones informada.

El hackathon se centró en patrimonio natural y biodiversidad, fomentando soluciones en ecosistemas, fauna y flora, gestión forestal, espacios protegidos y suelos. Aunque se plantearon casos de uso para orientar a los participantes, se permitió la libertad para desarrollar ideas propias dentro de esta temática.

3. Equipo de Trabajo

El equipo AI.IDEA estuvo conformado por integrantes de distintas especialidades y niveles de experiencia, unidos por un objetivo común y un compromiso compartido. La heterogeneidad del grupo facilitó una aproximación más holística a los retos planteados, enriqueciendo tanto la formulación de las ideas como su implementación práctica. La complementariedad entre perfiles técnicos, estratégicos, creativos y de comunicación dio lugar a soluciones sólidas, realistas y con potencial de escalabilidad. El trabajo en equipo se estructuró sobre la base de la colaboración continua y la toma de decisiones compartida.

Foto	Nombre	Experiencia	LinkedIn
------	--------	-------------	----------

	Fernando Hurtado Bocanegra	Doctor en Ecología y científico de datos, veo en este hackathon una oportunidad para aprender y poner a prueba mis conocimientos, con la intención de aportar valor añadido al evento.	https://www.linkedin.com/in/fernando-hurtado-bocanegra/
	Sergio Díez López	Estudiante Computer Science and Artificial Intelligence, con experiencia previa en Hackathones y creación de una start-up.	https://www.linkedin.com/in/sergio-d%C3%A9z-l%C3%B3pez-907257273/
	Adrián Ramos González	Graduado en Desarrollo de Aplicaciones Multiplataforma y Máster en Diseño 3D, debutando por primera vez en un hackathón.	https://www.linkedin.com/in/adrian-ramos-237801196/
	Miguel Loredó Valdés	Estudiante del grado en Transporte y Logística, con experiencia en hackathones previos y ecosistemas de emprendimiento.	https://www.linkedin.com/in/miguel-loredo-vald%C3%A9s-9811b6251/

4. Desarrollo por Fases

Reto 1: Generación y Selección Inicial de Propuestas (18/02/2025 – 10/03/2025)

El proyecto **AI.IDEA** tuvo su origen en el marco del Hackathon “Soluciones GenAI para la Biodiversidad”, impulsado por el Programa Nacional de Algoritmos Verdes (PNAV). Desde el inicio, el equipo se integró activamente en la dinámica del desafío, participando en la sesión de apertura del Reto 1 el 18 de febrero de 2025. Este primer encuentro marcó el comienzo del proceso de ideación y de la constitución del equipo.

El día 21 de febrero se formalizó oficialmente la conformación del equipo, con la validación de sus integrantes por parte de la organización. A partir de ese momento se inició una etapa intensa de generación de ideas y organización operativa. Se establecieron canales de comunicación estratégicos, incluyendo un grupo oficial en Telegram para la interacción con la organización, y un grupo interno de WhatsApp para coordinar acciones cotidianas de manera ágil. Paralelamente, el equipo se registró en la plataforma Microsoft for Startups Founders, lo que permitió acceder a recursos técnicos como Azure, esenciales para el desarrollo posterior.

Durante las semanas siguientes, se llevó a cabo un proceso colaborativo de ideación, nutrido por sesiones informativas y mentorías proporcionadas por el ecosistema del hackathon. La participación en actividades como la sesión virtual con el Instituto de Estudios en Biodiversidad (IEBNB) el 20 de febrero y las sesiones de mentoría del 27 de febrero resultó fundamental para afinar los objetivos y comprender la viabilidad de las propuestas. El equipo desarrolló un total de ocho ideas preliminares, todas documentadas en un formato común que facilitó su evaluación (véanse en [material complementario](#)). Estas propuestas incluían herramientas para digitalizar informes ambientales, calcular la huella de carbono en eventos, y facilitar el acceso a literatura científica actualizada sobre especies catalogadas.

La mentora asignada al equipo, Mónica Puebla Estrada, jugó un papel clave en la consolidación de estas ideas. Su asesoramiento permitió mejorar los enfoques técnicos y estratégicos de cada propuesta. El 25 de febrero se enviaron los primeros borradores a la mentora, se eligió formalmente el nombre del equipo y se activó un espacio colaborativo en Google Drive para centralizar recursos. De forma paralela, se comenzaron a entregar formularios de seguimiento según los requerimientos del hackathon.

El 5 de marzo se entregaron las propuestas definitivas para el Reto 1. Tres de ellas fueron seleccionadas por la organización para pasar a la siguiente fase: una propuesta de digitalización de informes de impacto ambiental, un software para calcular la huella de carbono de eventos, y un repositorio de artículos científicos actualizados sobre biodiversidad. Estas iniciativas darían forma al proyecto AI.IDEA a través de dos soluciones principales: **AURA** (Análisis de Uso de Recursos de Acontecimientos) y **ROSALIA** (Repository Of Scientific Articles on Listed Species). El 10 de marzo, se anunció oficialmente que AI.IDEA avanzaba a la siguiente etapa del hackathon.

Reto 2: Desarrollo y Consolidación de Propuestas (10/03/2025 – 28/05/2025)

Con el inicio del Reto 2, el equipo se reunió para deliberar cuál de las tres propuestas seleccionadas debía ser desarrollada en profundidad. Para orientar esta decisión, se definieron criterios específicos —eficiencia energética, escalabilidad técnica y viabilidad comercial— en sintonía con los objetivos del hackathon, especialmente los contemplados en los Retos 2 y 3. A través de una tabla comparativa de evaluación (ver imagen), se asignaron puntuaciones a cada propuesta de acuerdo con estos criterios. **AURA**, centrada en la medición de la huella de carbono en eventos, obtuvo la calificación más alta (22 puntos), seguida muy de cerca por **ROSALIA**, enfocada en la digitalización y consulta de literatura científica sobre biodiversidad (21 puntos). Por su parte, **AI.DEA**, a pesar de sus fortalezas técnicas, se situó en tercer lugar con 18 puntos. Aunque se descartó el desarrollo de la propuesta AI.DEA, el equipo decidió conservar sus raíces como identidad general del proyecto, en reconocimiento a su origen conceptual.

PROPUESTA	AI.DEA	AURA	ROSAL.IA
Conjunto de datos generado:	2	2	2
Complejidad técnica de la solución:	3	3	3
Calidad de la solución:	2	3	3
Eficiencia energética:	1	3	3
Claridad y precisión en la descripción del problema:	2	3	3
Estructura y detalle en las tareas realizadas:	3	3	3
Viabilidad comercial y de mercado:	2	3	1
Presentación y comunicación:	3	3	3
TOTAL:	18	22	21

La escasa diferencia entre AURA y ROSAL.IA, sumada a su clara complementariedad técnica y al entusiasmo del equipo por ambas líneas de trabajo, motivó la decisión de desarrollarlas de forma paralela durante el Reto 2.

Así se dividió el proyecto en dos ramas: **Aura_Project** y **Rosal.ia_Project**, cada una con su propio entorno documental y organización técnica. Esta decisión permitió al equipo abordar los desafíos desde una perspectiva más amplia y potenciar las capacidades colectivas mediante una gestión dual del desarrollo.

Durante las semanas siguientes se consolidaron los pilares técnicos y estratégicos de ambas líneas. El equipo centró su atención en tres ejes clave: la identificación de casos de uso reales como referencias prácticas, la definición general del diseño conceptual de las soluciones, y la exploración de fuentes de datos relevantes y verificables. Para asegurar una comprensión común entre los miembros del equipo, se compartieron recursos educativos en formato de “píldoras de conocimiento” y se fijaron los objetivos de la siguiente fase.

El equipo se puso en contacto con organizadores de eventos reales para obtener información sobre la medición y compensación de emisiones de carbono. Un ejemplo relevante fue *La Semana de Montaña*, un evento de cinco días celebrado en Gijón, Asturias, con 1.951 asistentes en total y una media de 390 por sesión. Durante el evento, se midió la huella de carbono (alcance 3) y se compensó con el objetivo de lograr huella 0. El cálculo fue realizado por la empresa Taxus Medioambiente, utilizando un sencillo cuestionario al que los asistentes pudieron acceder mediante un código QR, lo que representó un desafío en términos de obtener datos fiables. Además, se proporcionaron todas las facturas y datos de compras y servicios relacionados al evento, mientras que los consumos de energía fueron facilitados por el Ayuntamiento de Gijón. La compensación se gestionó a través del programa medioambiental del Ayuntamiento, que incluirá la plantación de árboles en terrenos municipales.

El 24 de marzo, se mantuvo una nueva reunión con la mentora, quien ofreció retroalimentación clave para afinar el enfoque técnico y conceptual. A partir de ahí, se

estableció una hoja de ruta más precisa, fijando como hito la estructuración final del proyecto para el 31 de marzo.

Durante los primeros días de abril, se definieron las tareas pendientes, se resolvieron dudas técnicas sobre Azure y se avanzó en el ajuste fino de los modelos. El 11 de abril, se enviaron instrucciones específicas sobre el uso de recursos provistos por Microsoft y se planificaron los últimos pasos de desarrollo. Se marcó como objetivo completar todas las entregas entre el 19 y el 22 de abril.

El 22 de abril, la organización anunció el aplazamiento del evento presencial final del hackathon y extendió la fecha de entrega del Reto 2 hasta el 28 de mayo. Esta extensión fue bienvenida por el equipo, que utilizó el tiempo adicional para refinar funcionalidades y estabilizar componentes técnicos esenciales. Fue durante esta etapa que se produjeron algunos ajustes internos: el 29 de abril, uno de los integrantes, Adrián, notificó su desconexión temporal del proyecto, comprometiéndose a reincorporarse en el Reto 3. Al día siguiente se realizó una nueva planificación de horarios para distribuir equitativamente la carga de trabajo entre los miembros restantes.

El 1 de mayo se integraron nuevas funcionalidades al sistema, como la optimización del *fetcher* de la API, la inclusión de la librería CodeCarbon para medir el impacto ambiental computacional y la solicitud de una API Key que permitiría acceder a un dataset extenso de más de 65.000 especies. El 5 de mayo, se consolidó la base de datos necesaria para una generación optimizada y, tres días después, se tomó la decisión de volver a utilizar exclusivamente el espacio colaborativo de Google Drive, a fin de mantener el código preliminar fuera de plataformas públicas como GitHub. El 8 de mayo se reafirmó esta decisión, estableciendo Google Drive como entorno principal de trabajo para evitar la exposición temprana del proyecto.

El 26 de mayo, se debatieron alternativas para la generación de *datasets* y modelos, evaluando ventajas y desventajas de mantener los scripts en archivos separados frente a una estructura unificada. Se identificaron cuellos de botella provocados por los límites de tiempo en ciertas APIs y se acordó registrar los tiempos de ejecución, así como las características de las máquinas virtuales utilizadas. Además, se realizó una revisión final de los diagramas de AURA y Rosalía. Persisten incompatibilidades en Azure al ejecutar AURA, debido a conflictos entre librerías internas y las del entorno del modelo, pese a los intentos de solución mediante URLs externas. Por su parte, Rosalía funciona en local, pero muestra limitaciones de rendimiento ante grandes volúmenes de datos; se planteó entonces el uso de máquinas virtuales con IPs distintas para gestionar peticiones en paralelo. La falta de respuesta a la consulta técnica enviada el 8 de mayo sigue complicando la resolución de estos problemas.

El 27 de mayo, se celebró una reunión técnica centrada en la ejecución del código en Azure. Se abordaron estrategias para mitigar *timeouts*, mejorar la gestión de errores, optimizar el uso de máquinas virtuales en Rosalía y organizar la integración del sistema con Streamlit.

Se creó un código para la extracción y suma del CO₂ proveniente del CodeCarbon en cada log, calculando más de 250 logs y obteniendo un resultado total acumulado de 0.412776389512317 kg CO₂eq de emisiones.

El Reto 2 cerró formalmente el 28 de mayo de 2025 a las 11:00 h, tras más de dos meses de trabajo técnico intensivo. Las propuestas fueron evaluadas por el jurado, y el 5 de junio se

comunicaron los resultados, anunciando los equipos seleccionados para avanzar al siguiente nivel.

Reto 3: Presentación final y cierre (05/06/2025 – 25/06/2025)

Esta fase contempla la preparación y exposición de la versión final del proyecto, con especial énfasis en su viabilidad desde el punto de vista de negocio. El 6 de junio a las 12:00 h, se realizó el kick-off virtual del Reto 3 a través de MS Teams, donde se proporcionaron las directrices para esta última etapa.


La entrega inicial de la presentación en formato PowerPoint estaba prevista para el 8 de junio, pero posteriormente se amplió el plazo al 15 de junio. El evento final presencial se celebrará oficialmente el 25 de junio en Madrid, donde los equipos finalistas presentarán sus soluciones ante el jurado final. *La memoria se ira actualizando paulatinamente durante el transcurso del hackathon.*


5. Referencias


Apis, librerías, listado de fuentes bibliográficas, recursos técnicos, documentos oficiales del Programa Nacional de Algoritmos Verdes y enlaces web consultados para fundamentar el desarrollo del proyecto, así como bibliografía científica sobre biodiversidad y huella de carbono.


Recursos	Descripción	Link
IEPNB	Listados oficiales y normativa sobre especies protegidas en territorio español.	https://iepnb.gob.es/api/catalogo/v_lista_patronespecie_normas
Crossref	Proporciona metadatos estructurados de publicaciones académicas y científicas. Permite acceder a información bibliográfica como títulos, autores, resúmenes y DOIs de artículos, facilitando su integración en sistemas de recuperación de literatura.	https://api.crossref.org/works
Semantic Scholar	Motor de búsqueda académica basado en inteligencia artificial, que ofrece acceso a artículos científicos, resúmenes y métricas de impacto. Su API permite realizar búsquedas temáticas avanzadas y extraer información clave para tareas de minería de texto o recuperación aumentada.	https://api.semanticscholar.org/graph/v1/paper/search?query=
Department for Business, Energy &	Proporciona los factores oficiales de conversión de distancia a emisiones de gases de efecto invernadero, utilizados como referencia en el cálculo de la huella de carbono.	www.gov.uk/government/collections/government-conversion-factors-


Industrial Strategy (UK)		for-company-reporting
Countries-States-Cities-Database (dr5hn):	Base de datos global que recopila información estructurada de países, estados y ciudades, incluyendo coordenadas geográficas, códigos ISO y zonas horarias.	github.com/dr5hn/countries-states-cities-database
EDJNet / European_routes:	Repositorio con datos de rutas aéreas y ferroviarias reales en Europa, útil para simular trayectos coherentes en modelos de movilidad.	github.com/EDJNet/european_routes
Instituto Nacional de Estadística	Fuente oficial del padrón municipal español, que incluye datos demográficos actualizados a nivel de municipio.	www.ine.es/dynt3/inbase/es/index.htm?padre=525
Airline Passenger Satisfaction	Dataset que recoge información sobre la experiencia de pasajeros aéreos, incluyendo su clase de viaje, útil para inferencias sobre distribución modal.	www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction
City Mayors	Listado de las 500 mayores ciudades europeas por población, empleado para mejorar el realismo geográfico en simulaciones de eventos internacionales.	www.citymayors.com/features/euro_cities.html
World Population Review	Portal con estimaciones actualizadas de población urbana a nivel global, utilizado para construir muestras sintéticas representativas.	worldpopulationreview.com/cities
Agribalyse	Es la base de datos de análisis de ciclo de vida (ACV) más completa de Europa en alimentación, compatible con España (similar dieta y producción) para más de 2.500 productos.	https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.57745/XTENSJ


Compartido conmigo > Hackathon 


✓ 
















Tipo 

Personas 

Modificado 

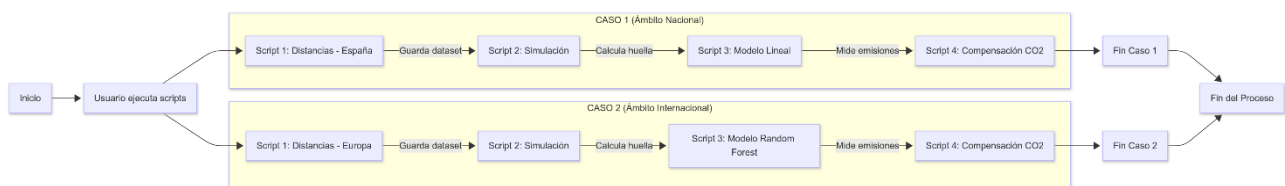
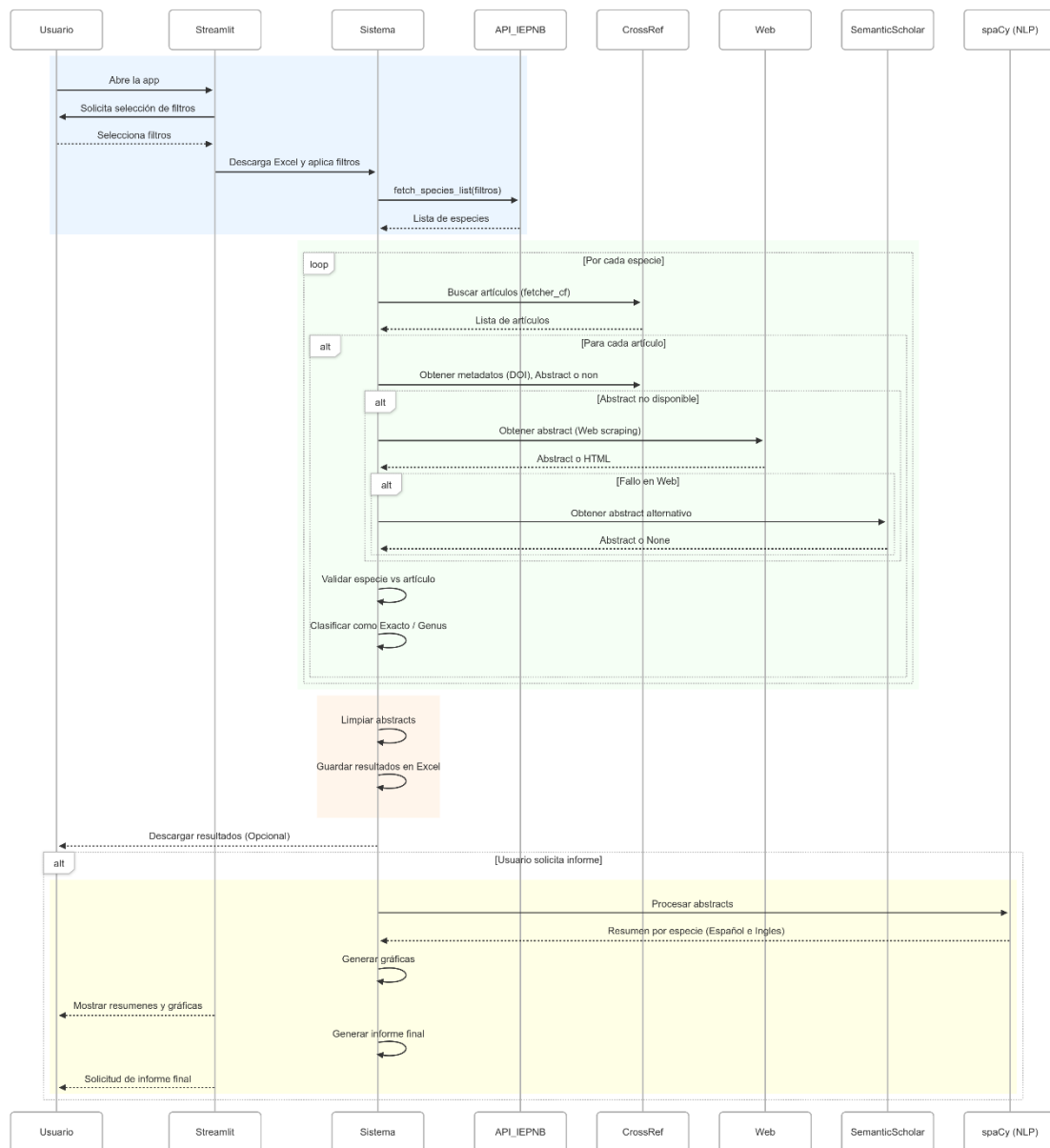
Fuente 

Nombre 	Propietario	Última modificación 	Tamaño de a	
 Documentación Reto 2	 yo	11 may 2025 yo	—	
 Primera etapa	 yo	11 may 2025 yo	—	
 SalamandraFoodFunction	 yo	15 may 2025 yo	—	

Entorno preliminar de Google Drive

6. Material complementario

6.1 Diagramas



6.2 Conjunto de ideas preliminares

En esta sección se recogen las ocho propuestas iniciales desarrolladas por el equipo durante la fase de ideación, a mediados de febrero. Estas propuestas abordan distintas problemáticas relacionadas con la digitalización, el análisis y el seguimiento del impacto ambiental y la biodiversidad. Su documentación en un formato común permitió una evaluación estructurada y sentó las bases para la selección y el desarrollo posterior de las soluciones más viables.

Propuesta 1	Digitalización de Informes de impacto ambiental y seguimiento.
Propuesta 2	Digitalización de Informes de impacto ambiental Red eléctrica, métricas y posible seguimiento.
Propuesta 3	Digitalización de Informes de Parques Nacionales para mejorar calidad del dato.
Propuesta 4	Software de Cálculo de Huella de Carbono para Eventos.
Propuesta 5	Generación de informes de impacto ambiental a partir de estándares estructurados.
Propuesta 6	Repositorio de información actualizado de artículos científicos sobre las especies que están catalogadas.
Propuesta 7	Creación de un índice de afección mixto de impacto ambiental.
Propuesta 8	Identificación de daños en pliegos de herbarios.

Propuesta 1: Digitalización de Informes de impacto ambiental y seguimiento

IA capaz de leer y procesar informes de Evaluaciones de Impacto Ambiental (EIA), Declaraciones de Impacto Ambiental e Informes de seguimiento, para transformarlos en bases de datos estructuradas. Estos informes, que contienen información detallada sobre proyectos de infraestructura, ubicación, biodiversidad afectada y medidas de mitigación, suelen estar en formato de texto extenso y poco estructurado. La IA extraería y organizaría estos datos para facilitar su análisis y comparación con bases oficiales, permitiendo detectar discrepancias, evaluar tendencias y mejorar la gestión ambiental. La idea es también sacar información de los muestreos realizados en esos informes y contrastarlos con las listas catalogadas del IEPNB.

Posibles fuentes de datos:

<https://sede.miteco.gob.es/portal/site/seMITECO/navServicioContenido#>

Enfoque: técnicas de NLP basadas en reglas con modelo LLM (usando Salamandra 7B si fuese posible)

Propuesta 2: Digitalización de Informes de impacto ambiental Red eléctrica, métricas y posible seguimiento

la capaz de leer y procesar informes ambientales de la Red Eléctrica y órganos asociados. La red de transporte de energía eléctrica en España cuenta con más de 45.000 km de líneas de alta tensión, dispone de datos geoespaciales, como cartografía de áreas sensibles y mapas de riesgo de colisión para aves, análisis de flora y fauna de alrededores, con un seguimiento periódico de las instalaciones que permite actualizar la información ambiental. Apoyándonos en la propuesta de corredores verdes que propusieron hace unos años (<https://www.ree.es/es/sala-de-prensa/actualidad/nota-de-prensa/2022/09/corredores-biodiversidad-red-electrica-buena-practica-medioambiental-Europa>). Para hacer un análisis de la fauna y flora y sus tendencias. Las ventajas que posee, es la centralización de los informes (Red eléctrica) y formato más homogéneo que quizás la propuesta 1. Esta empresa invierte decenas de millones al año en el ámbito de medioambiente, por tanto, un modelo que extrae

datos y revela métricas ambientales, podría ser un producto de alto valor para ellos. Siguiendo con el enfoque de su gran capacidad de recursos económicos, podríamos hacer una prueba piloto en un tramo de la red incorporando sensores IoT, para obtener datos de flora y fauna en tiempo real.

Posibles fuentes de datos: Red eléctrica y órganos colaboradores como CSIC.

[https://ieeb.fundacion-](https://ieeb.fundacion-biodiversidad.es/sites/default/files/aaff_informe_fichas_ree_digital.pdf)

[biodiversidad.es/sites/default/files/aaff_informe_fichas_ree_digital.pdf](https://ieeb.fundacion-biodiversidad.es/sites/default/files/aaff_informe_fichas_ree_digital.pdf)

Propuesta 3: Digitalización de Informes de Parques Nacionales para mejorar calidad del dato

Recuperar y organizar informes obsoletos para mejorar la calidad del dato y permitir el análisis de tendencias. Dado que estos documentos tienen estructuras no estandarizadas, se desarrollaría un clasificador que los categorice según criterios como año, ubicación y hábitat. Inicialmente, se entrenaría un SLM en un ámbito específico, como un parque parques nacionales, demostrando así su viabilidad.

Posibles fuentes de datos: Database de informes de parques nacionales.

Enfoque: Similar a la propuesta 1, si bien podría orientarse a generar una base de datos usando un clasificador.

Propuesta 4: Software de Cálculo de Huella de Carbono para Eventos

Desarrollar un software para calcular la huella de carbono en eventos, abordando las emisiones de alcance 3 (viajes, dietas, etc.), que suelen tener alta incertidumbre por falta de datos. Mediante un cuestionario para asistentes y anfitriones, un modelo de ML (como Random Forest) estimaría las emisiones de los viajes, integrándolas en el cálculo total del evento. Además, el software generaría reportes automáticos conforme a la normativa y usaría un LLM para recomendar estrategias de reducción de emisiones.

Su uso e implementación podría convertirse en un estándar e incluso ayudar en la gestión y regularización de la huella de carbono de organismos públicos y privados y en la gestión medioambiental de eventos

Posibles fuentes: Se pueden generar con IA a modo de ejemplo, no necesitamos datos reales Ejemplos: Universal Data, Mostly AI, módulo CTGAN....).

Enfoque: Se pueden plantear desde apps de checklist para la gestión de información de cuestionarios (o incluso forms de google), librerías tipo Folium para ayudar en los cálculos de trayectos junto a modelos RF y toda la integración en dashboards (ej. Plotly Dash) para la visualización.

Propuesta 5: Generación de informes de impacto ambiental a partir de estándares estructurados

La Evaluación de Impacto Ambiental (EIA) es un proceso regulado en España por la Ley 21/2013 para identificar y mitigar los efectos ambientales de proyectos antes de su ejecución. Grandes proyectos eléctricos, como parques eólicos o centrales fotovoltaicas, deben

someterse a este procedimiento, generando una gran cantidad de documentación pública. La propuesta es desarrollar una solución de IA que estructure y analice esta información, para crear un producto que beneficie a promotores y comunidades locales al facilitar la creación de futuros informes mediante conocimientos previos. (Es muy parecida a la idea de Adrian, pero en vez de extraer datos para la administración o democratizarlos para la población, sería un modelo productizable para ofrecer a constructoras y promotoras inicialmente de parques eólicos. También se podría hacer un modelo híbrido, de conocimiento para la población y de pago para la herramienta de creación de informes ambientales).

Enfoque: es el giro siguiente a la propuesta 1 (misma fuente de información), se puede realizar en paralelo con otro modelo de LLM, para sacar una estructura coherente. Se puede implementar a aplicaciones similares a la propuesta 4 de recogida de información para generar informes como producto

Propuesta 6: Repositorio de información actualizado de artículos científicos sobre las especies que están catalogadas.

Es bastante ambicioso. Usando las listas de especies catalogadas como base de datos, hacer webscraping de Scholar Google, y sacar los links de las publicaciones asociadas que citen textualmente a la especie y almacenarlo en la base de datos como campo de "publicaciones asociadas" a modo de lista para cada especie. La idea es usar también las sinonimias.

Aterrizar esto no es sencillo, pero tiene chicha y es aunar el conocimiento científico disponible para cada especie.

Enfoque: Implementar web scrapping en la búsqueda de la api de scholar google. Tiene los inconvenientes de usar una api para scrapping, pero también se puede ejecutar periódicamente para actualizar la información si hacemos deployment de las llamadas a la api en una nube (ejemplo, Azure)

Propuesta 7: Creación de un índice de afección mixto de impacto ambiental.

Aunar efectos negativos provenientes de incendios y erosión de suelos en un índice mixto como índice de afección (requeriríamos algo de apoyo por parte de la organización, pudiendo consultar a expertos en ambas áreas para la generación del índice). Aquí tiraríamos tanto de datos de los mapas de incendios como de erosión a nivel histórico. Deberíamos de recoger información geolocalizada y crear nueva a partir de ella. Se pueden utilizar modelos y herramientas de diferentes tipos y plantear también modelos predictivos. A favor, puede ayudar a entender de una mejor manera el impacto en conjunto que está ocurriendo sobre los ecosistemas, ya que la info de los inventarios de especies, que se hacen irregularmente, es una información mucho más sesgada.

Enfoque: Feature engineering (creación de variables) que midan el impacto negativo usan distintas fuentes de datos. Se pueden incorporar detección de anomalías (anomaly detection) y clasificadores

Propuesta 8: Identificación de daños en pliegos de herbarios.

Aprovechando la digitalización de los herbarios, realizar entrenamiento de modelos que ayuden a la identificación de daños en los pliegos a través de imágenes. Me explico, es generar un modelo que, analizando las imágenes digitalizadas de los herbarios, asistan en ver la degradación de esas especies conservadas en el herbario. Esto podría ayudar a ver cuando se necesitan pliegos adicionales que sustituyan a los existentes de esas especies en los herbarios, evitando la pérdida de especímenes que puedan ser clave el día de mañana. Está última habría que hablarla también con la organización. Sería genial ya que es un problema de conservación a nivel mundial que podría implementarse de forma sencilla en aquellos herbarios digitalizados.

Enfoque: Podríamos usar VAEs o similares que simplifiquen la información para alimentar clasificadores (rangos/clases de daños). Necesitaríamos colaboración con los herbarios para el correcto etiquetado de daños y usar modelos de aprendizaje supervisado.