

AI4ES

DATATHON 2023

MIEMBROS:



ENTIDADES FINANCIADORAS



→ Agenda

1. Introducción y presentación convocatoria
2. Retos Datathon 2023
 - Reto A - Industrial: Procesamiento de imágenes
 - Reto B - Social: Analítica de datos
3. Información práctica
4. Preguntas



MIEMBROS:




ENTIDADES FINANCIADORAS



The background is an abstract, textured composition dominated by deep blue and purple hues. It features a bright, glowing light source on the right side, which creates a strong lens flare and illuminates the surrounding areas. The texture is grainy and layered, with various shades of blue, purple, and hints of yellow and green scattered throughout, giving it a cosmic or ethereal feel.

Introducción

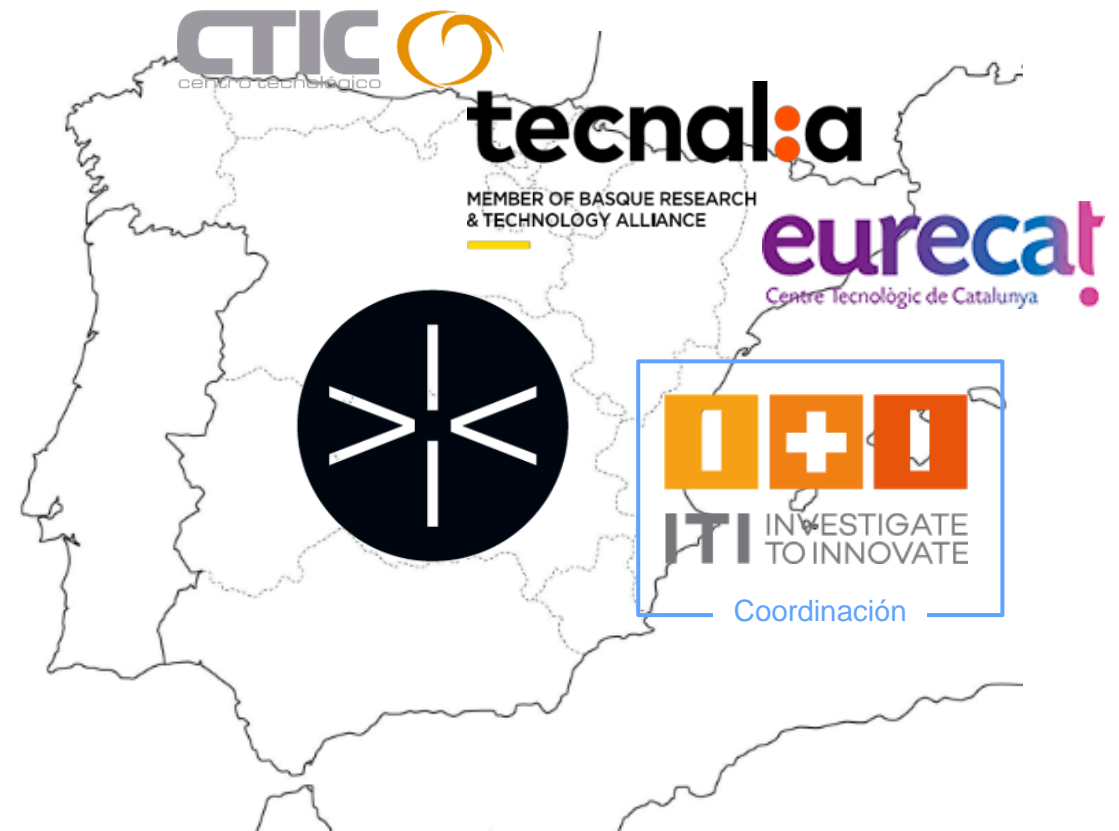
ΛI4E S



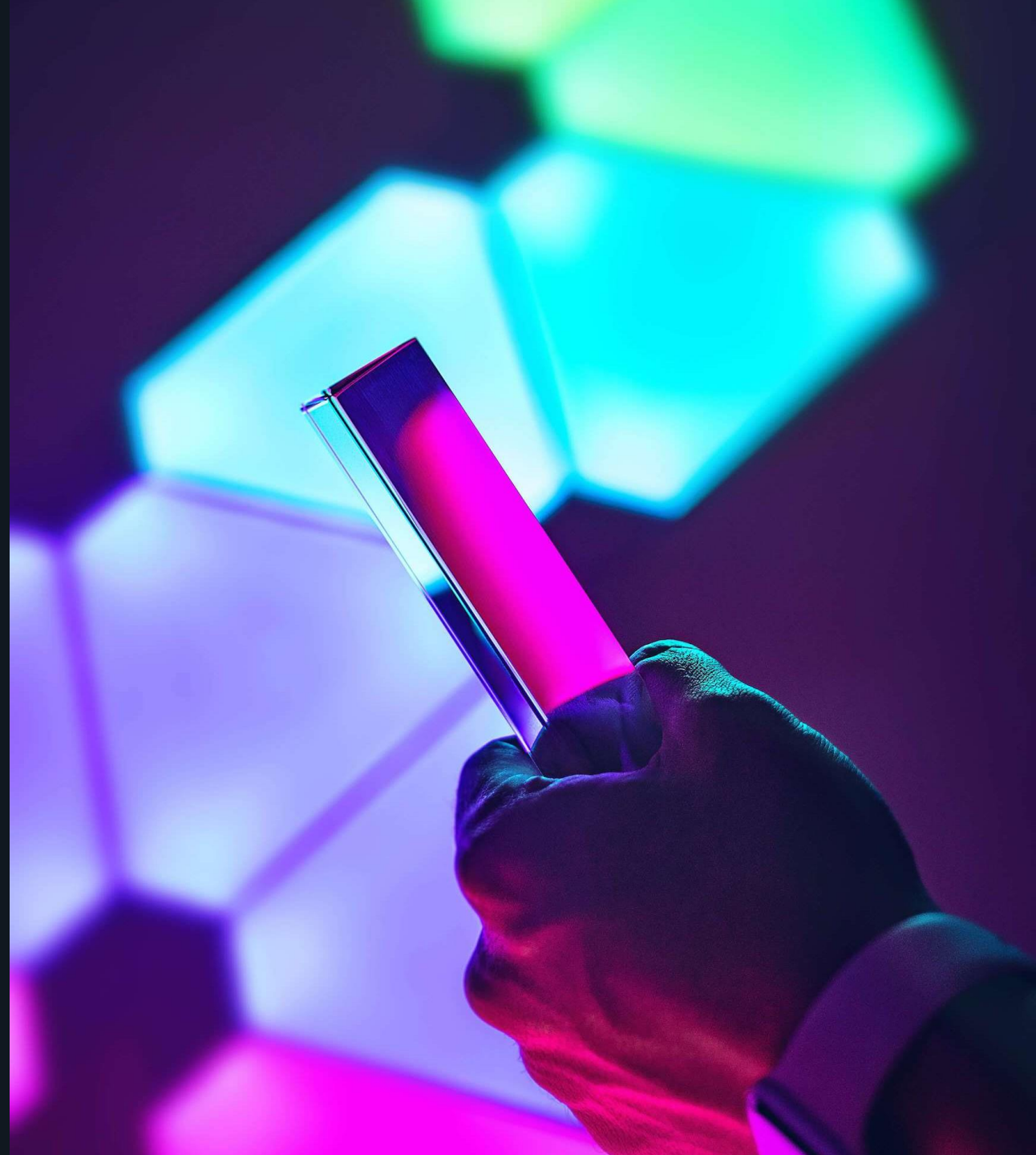
AI4ES es una red de excelencia
en tecnologías habilitadoras
basadas en el dato

Quiénes somos

- **4 Centros Tecnológicos I+D**
- **Capacidades y experiencia** en tecnologías Big Data, IA (ML y DL) e infraestructuras computación para manejo datos
- **+280 profesionales** (investigadores e ingenieros) en estas tecnologías
- Amplio **posicionamiento en foros y plataformas** tecnológicas internacionales
- **+90 proyectos europeos I+D 2018-2020**
- **+127 empresas españolas** colaboradoras en proyectos I+D europeos



El objetivo de AI4ES es
potenciar la economía
basada en Datos e
Inteligencia Artificial a
nivel nacional



¿Por qué AI4ES?

Credits: The Indian Express

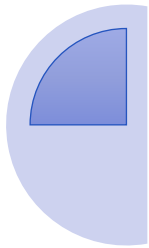


Las tecnologías digitales están cambiando drásticamente la forma en que diseñamos, producimos, comercializamos y consumimos todo tipo de bienes y servicios.

La especialización en estas tecnologías genera efectos significativos y positivos en el rendimiento económico de las regiones, como pueden ser:

- Impacto sustancial en la **creación de empleo altamente cualificado**
- Relevancia sistémica en todas las fases de desarrollo de producto, entre **diversas cadenas de valor industrial**
- Capacidad para **mejorar la salud, la seguridad, el desarrollo sostenible y la conectividad** entre los sistemas y los individuos
- Habilitador de múltiples **aplicaciones industriales en múltiples sectores**, contribuyendo a la generación de excelencia global, nuevo conocimiento y nuevas formas de participación

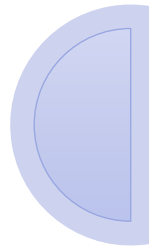
Nuestra visión



Visión

Ser **referente español** en I+D+I y transferencia en Tecnologías Habilitadoras Digitales relativas al procesamiento y análisis inteligente de datos

Potenciar la economía basada en datos e inteligencia artificial a nivel nacional



Misión

Agregar, de forma armonizada y coordinada, las actividades, capacidades, e infraestructuras de I+D+I en tecnologías de aprendizaje automático y extracción de valor de los datos

Cerrar la brecha entre la investigación y la industria, **facilitando la transferencia y buscando el impacto**, principalmente en las PYMES, proporcionando soluciones y servicios innovadores que requieren de **acceso seguro y confiable a datos, procesamiento eficiente, análisis avanzado, aprendizaje automático e inteligencia artificial**



Retos Tecnológicos

→ La actividad de la red se estructura en 4 objetivos científico-tecnológicos, que constituyen los 4 pilares de la red



• Modelos de aprendizaje para datos heterogéneos (multimodales, de diversas fuentes), **degradados o de baja calidad** (datos faltantes, insuficientes, no significativos, erróneos, con anomalías, etc.)

OT 1: Data Analytics

• Modelos de aprendizaje continuo y sistemas que **soporten su ciclo de vida**, que permitan ir incorporando nuevo conocimiento de forma incremental de forma automática o con una mínima supervisión

OT 2: IA Development



• Nuevos paradigmas de computación especializados para **Inteligencia Artificial**, que permitan aumentar la eficiencia del cómputo de algoritmos de aprendizaje automático, explorando conexiones con GAIA-X.

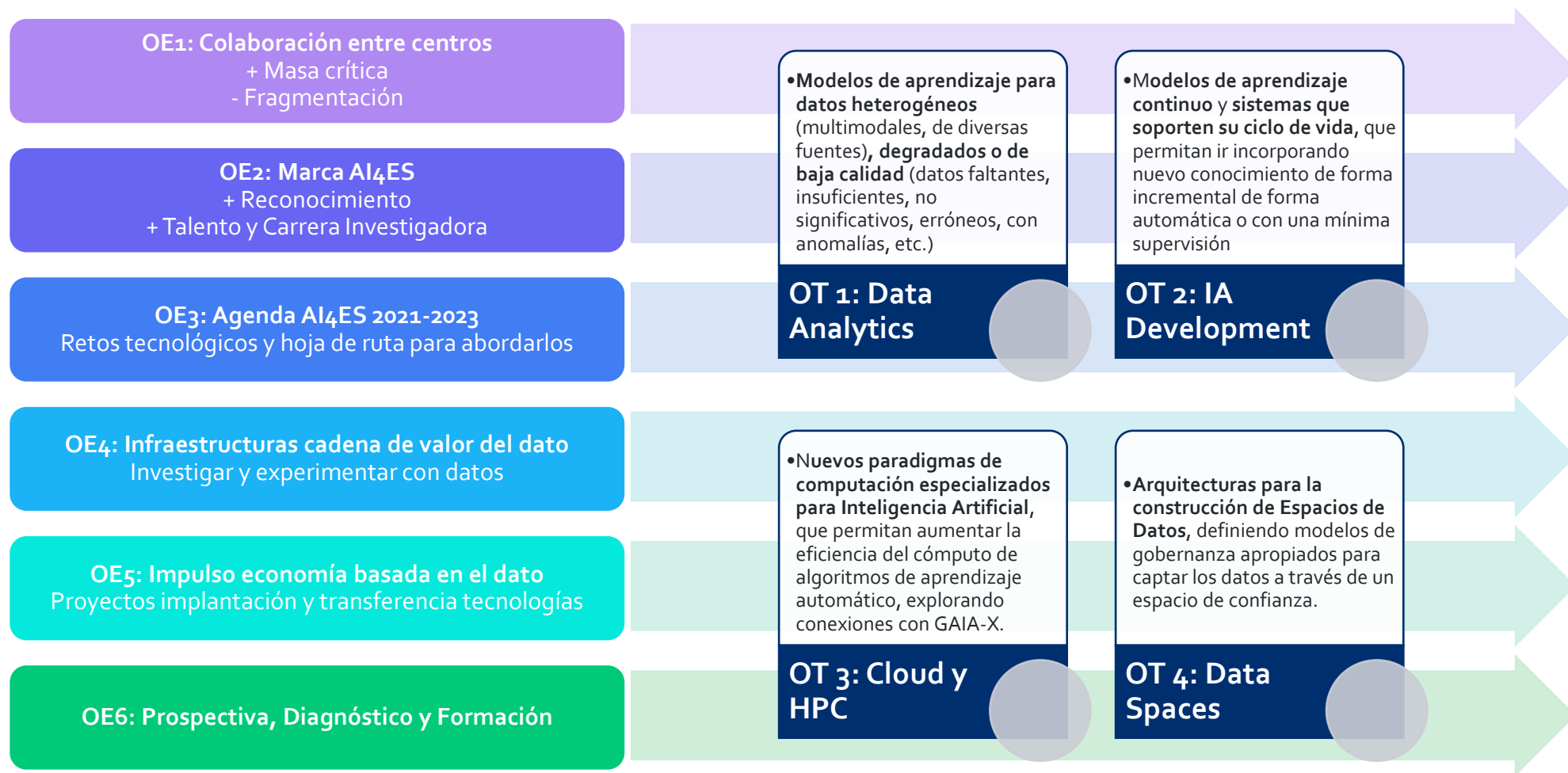
OT 3: Cloud y HPC

• Arquitecturas para la **construcción de Espacios de Datos**, definiendo modelos de gobernanza apropiados para captar los datos a través de un espacio de confianza.

OT 4: Data Spaces



Retos Tecnológicos y Estratégicos





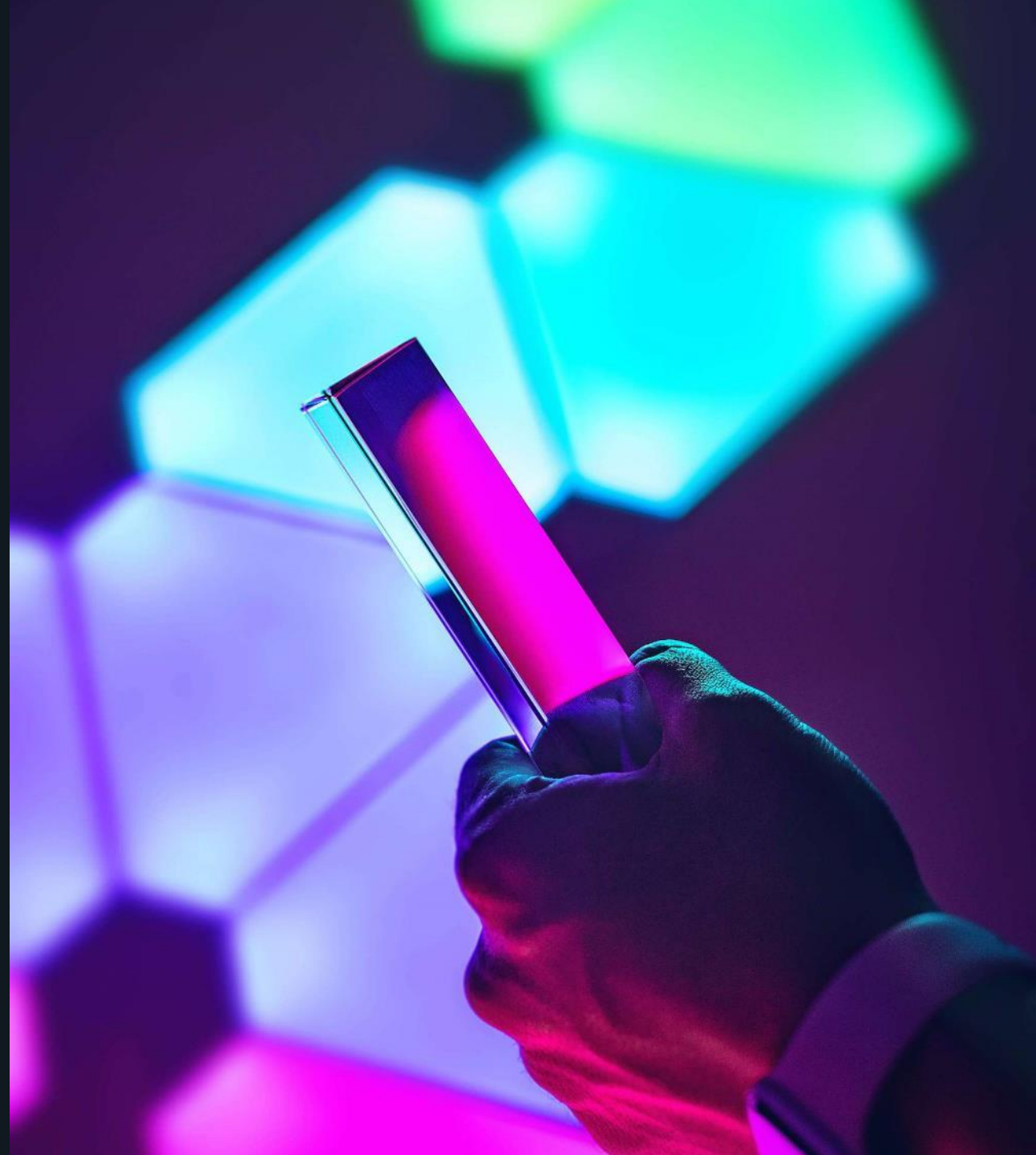
AI4E

DATATHON 2023

Objetivos Datathon

- **Atracción y reconocimiento del talento** en tecnologías AI4ES
- Estimular acciones de formación, capacitación, e I+D, para el avance de las tecnologías AI4ES a nivel nacional
- Estimular la economía basada en el dato a través del lanzamiento de retos en tecnologías AI4ES

Buscamos a los mejores
estudiantes y
profesionales en el uso de
tecnologías IA para
procesamiento de datos





Características

- 3 categorías de participación, según grado de experiencia
 - A. Futuras promesas
(< 3 años experiencia)
 - B. Especialistas
(+3 años experiencia)
 - C. Técnicos e Investigadores AI4ES

Premios

- 3 premios para cada reto, según categoría participación
 - Futuras promesas: 1.000€
 - Especialistas: 2.000€
 - Miembros AI4ES: 1.000€ en especie

Retos 2023

	Reto A	Reto B
Proponente AI4ES	ITI	CTIC
Tipo reto técnico	Procesamiento imágenes	Analítica datos
Ámbito	Industrial	Social
Sector	Manufactura	Energía

The background is an abstract, textured surface with a color palette dominated by deep blues, purples, and hints of green and yellow. The texture appears granular and uneven, with darker, more saturated areas interspersed with lighter, more vibrant patches. The overall effect is reminiscent of a microscopic view of a mineral or a close-up of a painted surface.

Retos Datathon 2023

Reto A (Industrial):
Procesamiento de imágenes

Reto A (Industrial): Procesamiento de imágenes

Detección de defectos superficiales en madera



1. Introducción

- Obtener piezas anómalas en entornos industriales es complejo. Debemos trabajar con piezas de control
- Modelar la normalidad de piezas con texturas con alta variabilidad es complejo
- ZeroGravity3D, sistema de inspección industrial desarrollado por ITI, captura imágenes de piezas inspeccionadas en caída libre ([enlace a web](#))

2. Objetivos

- Detectar anomalías en las superficies de piezas con texturas de madera
 - Porosidades
 - Manchas
 - Roturas
 - Nudos

Reto A (Industrial): Procesamiento de imágenes

Detección de defectos superficiales en madera

3. Dataset

- Base de datos con 802 imágenes de train y 2398 de test (637MB) ([Kaggle](#))
- Imágenes de test etiquetadas a nivel de píxel con la clase y si es una región incierta
- Métricas a nivel de píxel y a nivel de imagen



Reto A (Industrial): Procesamiento de imágenes

Detección de defectos superficiales en madera

4. Aspectos a tener en cuenta

- Librerías y algoritmos útiles: librería PyOD, algoritmo EfficientAD, Deep Learning + KNN...
- Opción: trabajar con ventanas en vez de con la imagen completa
- Dentro de la base de datos hay un notebook introductorio
- Se provee el código para cargar la BD y sacar las métricas
- Las predicciones tienen que tener la misma resolución que la imagen de la que provienen
- Se deben entregar también ficheros con las predicciones del modelo continuas (**no binarias**), un fichero con el mismo nombre que la imagen que representa

Reto A (Industrial): Procesamiento de imágenes

Detección de defectos superficiales en madera



4. Personas de contacto

→ Omar del Tejo Catala: odeltejo@iti.es

The background is an abstract, textured surface with a color palette dominated by deep blues, purples, and hints of green and yellow. The texture appears granular and uneven, with darker, more saturated areas interspersed with lighter, more vibrant patches. The overall effect is reminiscent of a microscopic view of a mineral or a close-up of a painted surface.

Retos Datathon 2023

Reto B (Social):
Analítica de datos

1. Introducción

- Se plantea como reto la predicción del consumo energético de los dos edificios de oficinas de CTIC, localizadas en Asturias (Gijón y Peón).
- Dadas las características propias de un contexto laboral con horario de oficina, sus patrones de consumo son relativamente estables a lo largo del tiempo.

2. Objetivos

- Desarrollar sistemas predictivos del consumo energético para ambos edificios de oficinas.
- Generar predicciones para dos meses completos a partir de los datos proporcionados.
- Almacenar dichas predicciones en CSVs con el mismo formato que los datos de entrada.

Predicción de consumo energético en edificios

3. Dataset

- Datos de consumo energético de los dos edificios de oficinas de CTIC, localizadas en Asturias (Gijón y Peón).
- Se proporcionarán datos de 22 meses consecutivos para ambas instalaciones, con datos recogidos horariamente, y con el consumo en kWh asociado a cada instante.

```
"ts", "Consumo_kWh"  
2021-06-26 01:00:00, 0.429  
2021-06-26 02:00:00, 0.43  
2021-06-26 03:00:00, 0.427  
2021-06-26 04:00:00, 0.434  
2021-06-26 05:00:00, 0.434  
2021-06-26 06:00:00, 0.433  
2021-06-26 07:00:00, 0.424  
2021-06-26 08:00:00, 0.407  
2021-06-26 09:00:00, 0.402  
2021-06-26 10:00:00, 0.411  
2021-06-26 11:00:00, 0.407  
2021-06-26 12:00:00, 0.401
```


Predicción de consumo energético en edificios

4. Aspectos a tener en cuenta

- Los modelos predictivos podrán realizarse con los tipos de algoritmos que se consideren pertinentes, no hay limitaciones. Además, podrá utilizarse cualquier lenguaje de programación.
- Cada participante deberá proporcionar dos ficheros CSV donde se recojan las que ellos consideren como las mejores predicciones horarias de los modelos generados para los siguientes dos meses a partir del fin de los datos de entrenamiento proporcionados (desde "2023-04-27 01:00:00" hasta "2023-06-26 00:00:00"). Dicho fichero a de seguir el mismo formato que los CSV de entrada, bajo los nombres *gijon.csv* y *peon.csv*.

Reto B (Social): Analítica de datos

Predicción de consumo energético en edificios



4. Personas de contacto

→ Pelayo Quirós: pelayo.quiros@fundacionctic.org

→ Chus García: chus.garcia@fundacionctic.org

The background is an abstract, textured composition of various shades of blue and purple. It has a grainy, almost crystalline appearance, with darker, more saturated areas interspersed with lighter, more vibrant blue patches. A prominent, bright, diagonal streak of light blue and white light cuts across the lower half of the image, adding a sense of movement and depth. The overall effect is ethereal and dynamic.

Información práctica

Recursos

→ Canales de comunicación

→ Lista correo para **difusión general**: inscritos-datathon-2023@ai4es.es

→ Comunicaciones generales desde la organización a todos los inscritos

→ **Email de contacto general** (dudas, incidencias): datathon@ai4es.es

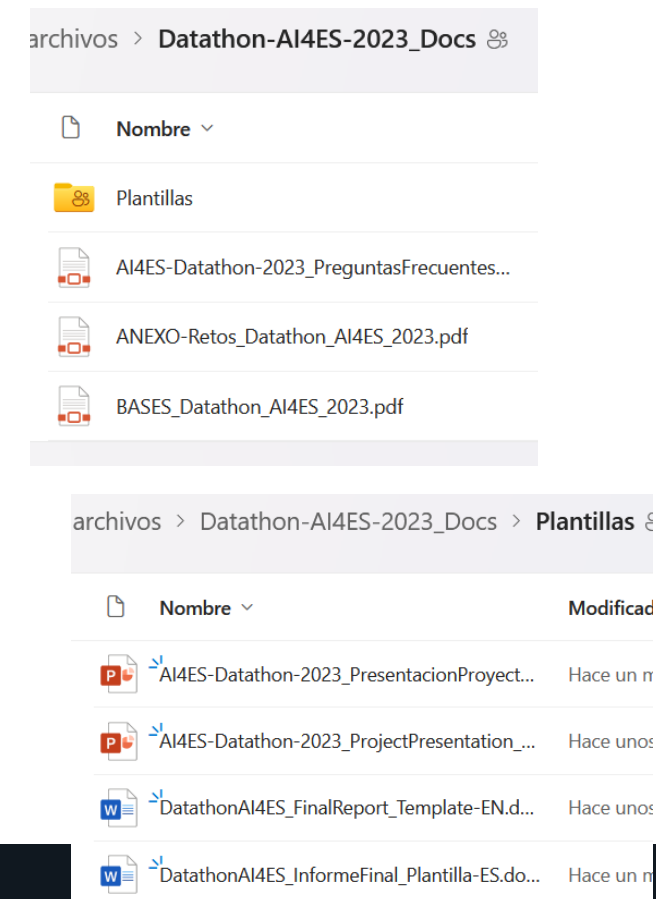
→ Incluir el reto al que se refiere la consulta en el asunto del correo, para agilizar la gestión

→ Repositorios:

→ **Repositorio común** ([ENLACE](#)): Bases y Anexo, presentación informativa, FAQs, y plantillas documentos

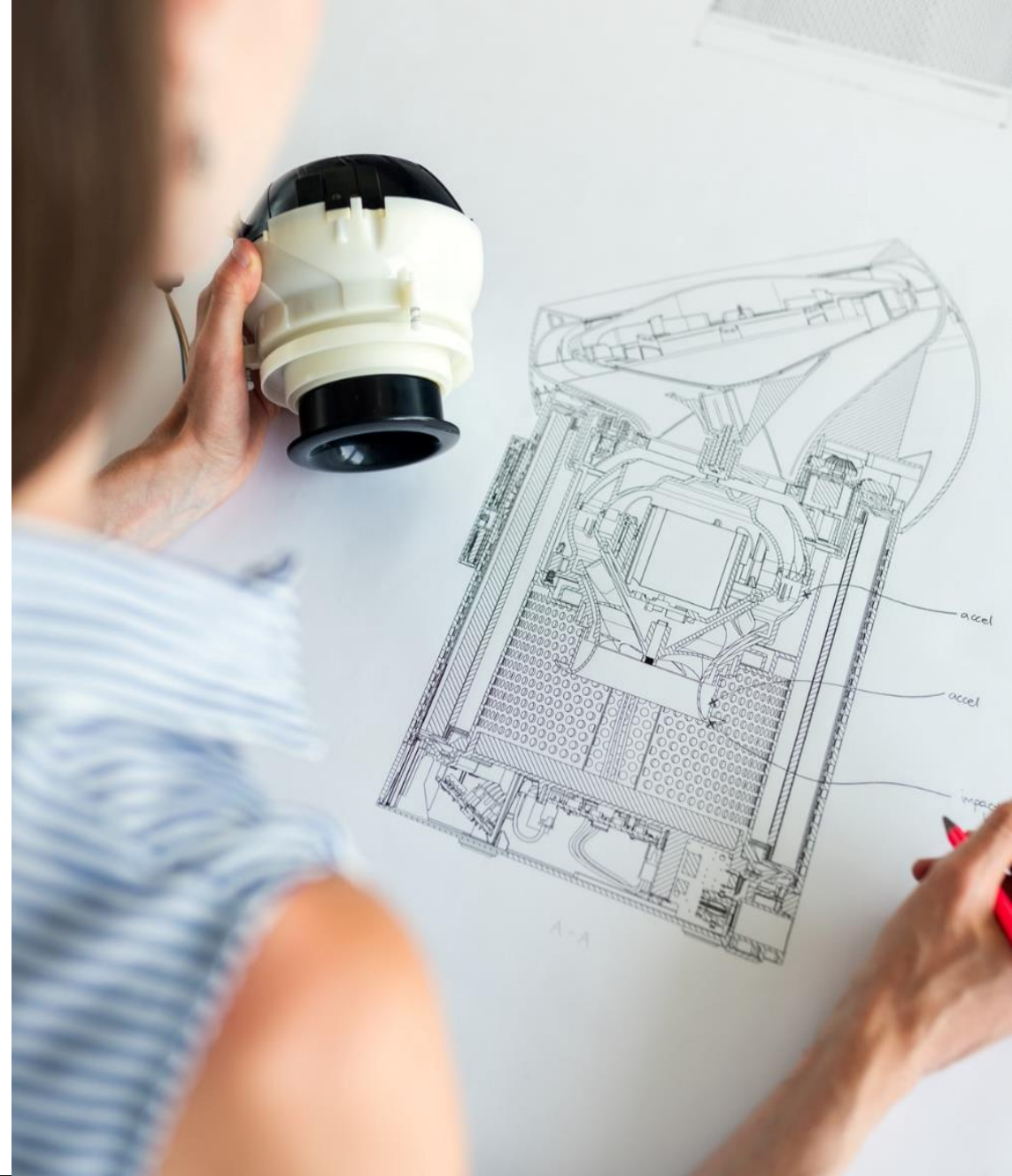
→ **Espacio de trabajo** (por invitación): Cada reto cuenta con un grupo Microsoft Teams dedicado, donde se proporciona acceso a los datasets del reto y documentación específica del reto

→ Si problemas con el acceso, contactar con Datathon@ai4es.es, indicando el reto en el asunto



Presentación del proyecto

- Informe resumen del proyecto realizado (5 páginas máximo)
 - Enfoque adoptado
 - Características del desarrollo
 - Principales resultados y conclusiones
- Presentación oral ante Jurado
 - 15 minutos: 10 exposición + 5 preguntas
 - En remoto
- Disponibles plantillas:
 - Informe resumen ([ES](#) | [EN](#))
 - Presentación oral ([ES](#) | [EN](#))



Plantilla Informe final

AI4ES

DATATHON 2023

Informe final Datathon AI4ES

Categoría: Futuras promesas / Especialistas / Miembros AI4ES (DEJAR ÚNICAMENTE LA CATEGORÍA QUE APLIQUE)

Participante:

Nombre y apellidos del participante / Nombre equipo y de sus integrantes

Email contacto:

Lugar, 23/11/2023

AI4ES Datathon 2023 - Informe final Datathon AI4ES

1

Contenido

Instrucciones 2

1 Enfoque 3

1.1 Reto abordado y principales desafíos 3

1.2 Gestión de datos y pre-procesamiento 3

1.3 Estrategia de particionado de datos 3

2 Desarrollo 3

2.1 Descripción del algoritmo 3

2.2 Entrenamiento 3

3 Resultados 4

4 Conclusiones 4

AI4ES

AI4ES Datathon 2023 - Informe final Datathon AI4ES

3

1 Enfoque

1.1 Reto abordado y principales desafíos

Descripción del reto abordado, desde la perspectiva de los problemas y dificultades que plantean los datos, facilitados para su análisis y que hacen necesario el uso de algoritmos para su procesamiento.

1.2 Gestión de datos y pre-procesamiento

Describir las operaciones realizadas sobre los datos proporcionados antes de proceder al entrenamiento del modelo (carga de datos, normalización y/o otras computaciones o transformaciones). Si se han incorporado datos adicionales a los facilitados, indicarlo.

1.3 Estrategia de particionado de datos

Describir la estrategia usada para particionar el/los dataset(s) indicando si ha sido realizado de forma aleatoria sobre el dataset completo o dirigido por algún parámetro. Indicar también los ratios de particionado (y/o ratios) usados.

2 Desarrollo

2.1 Descripción del algoritmo

Describir el algoritmo

2.2 Entrenamiento

Describir las estrategias de entrenamiento, tipos de parámetros utilizados, etc.

AI4ES

AI4ES Datathon 2023 - Informe final Datathon AI4ES

4

3 Resultados

Proporcionar los resultados obtenidos tras el proceso de desarrollo, y un breve análisis de dichos resultados.

4 Conclusiones

Resumir los principales aspectos del trabajo realizado, destacando aquellos que sean más destacables y relevantes. Indicar qué características del trabajo realizado son más innovadoras o creativas, desde la perspectiva y experiencia del participante.

Realizar una valoración, justificada, de en qué medida el enfoque seguido tendría los siguientes atributos: Escalabilidad (escalado ante mayores volúmenes de datos, robustez ante cambios de escenario o contexto); Facilidad de implantación (en el sentido de despliegue e integración); Generalización (la solución podría aplicarse a otros problemas o es específica).

AI4ES

Plantilla presentación oral



0. Criterios de evaluación

- Impacto e importancia del reto solucionado (comprensión y resolución)
- Uso de los datos (gestión, pre-procesamiento, particionado)
- Grado de acabado conseguido (consistencia y robustez resultados)
- Creatividad e innovación en el enfoque seguido
- Claridad de la presentación oral
- Criterios técnicos específicos adicionales valorables: grado de escalabilidad, facilidad de implantación, generalización

1. Enfoque

- Breve presentación del problema abordado (reto) y las principales dificultades
- Presentación del uso de los datos: gestión y pre-procesamiento realizado para preparar los datos, y estrategia de partición realizada

2. Desarrollo

- Describir el algoritmo
- Describir las estrategias de entrenamiento

3. Resultados

- Presentar los resultados obtenidos y un breve análisis

4. Conclusiones

- Presentar una valoración general del proyecto realizado, destacando los aspectos más relevantes o que podrían ser diferenciales, innovadores, o creativos.
- Valorar atributos como escalabilidad, facilidad de implantación y generalización.

Evaluación

→ Criterios evaluación

- Impacto e importancia del reto solucionado (comprensión y resolución)
- Uso de los datos (gestión, pre-procesamiento, particionado)
- Grado de acabado conseguido (consistencia y robustez resultados)
- Creatividad e innovación en el enfoque seguido
- Claridad de la presentación oral
- Criterios técnicos específicos adicionales valorables: grado de escalabilidad, facilidad de implantación, generalización

→ Jurado Datathon

- Se constituirá un Jurado por cada reto, especializado en procesamiento de imágenes y analítica datos
- Cada Jurado constará de 4 miembros, uno por cada centro de la red (ITI, CTIC, EURECAT, TECNALIA), siendo técnicos especialistas en la temática.

Fechas importantes

- 10-Noviembre-2023 23.59: Fin inscripciones
- **23-Noviembre-2023 23.59**: Fin Datathon (límite envío proyecto)
- 27 al 29 Noviembre: Presentación de proyectos ante el Jurado
- 30 Noviembre 2023: Fallo del Jurado y comunicación resultados
 - AI4ES Summit 2023: <https://www.ai4es.com/ai4es-summit-2023/>
(evento híbrido, inscripción gratuita abierta al público)

The background is an abstract, textured composition of various shades of blue and purple. It has a grainy, almost crystalline appearance, with darker, more saturated areas interspersed with lighter, more vibrant patches. A bright, circular light source is visible in the upper right quadrant, casting a strong, radial glow across the scene. This light creates a sense of depth and highlights the intricate textures of the background. The overall effect is ethereal and dynamic, with the light source acting as a focal point that draws the viewer's eye.

Preguntas

Preguntas

- Turno abierto para preguntas
- Las preguntas y respuestas frecuentes se recopilarán e incluirán en documento FAQ
- Preguntas offline: datathon@ai4es.es



AI4ES

RED DE EXCELENCIA EN
TECNOLOGÍAS HABILITADORAS
BASADAS EN EL DATO

Gracias

<https://ai4es.com> →

MIEMBROS:



ENTIDADES FINANCIADORAS:

