E1.1 REQUISITOS TÉCNICOS Y PLATAFORMAS DEL PROYECTO

Versión 1.1

Información de la Documentación

Web del Proyecto	https://capsul-ia.es/
Fecha	30/06/2024
Nivel Diseminación	Público
Autor	Diego Remírez (GTD), Francisco J. Cazorla (BSC)
Colaboradores	Leticia Pascual (SML), Enrico Mezzetti (BSC)
Revisor	Lorea Belategi (IKL), Alfonso González (ZYL)
Palabras clave	Requisitos, Plataformas

Project MIG-20232044, subsidized by the CDTI – Center for Industrial Technological Development, supported by the Ministry of Science and Innovation, and approved in the call for the year 2023 of the procedure for granting aid for “Science and Innovation Missions” within the framework of the TransMisiones 2023 initiative, within the scope of the State Program to Catalyze Innovation and Business Leadership of the State Plan for Scientific and Technical Research and Innovation 2021-2023. Project PLEC2023-010240 funded by MICIU/AEI/10.13039/501100011033”

Registro de Modificaciones

Versión	Descripción del cambio
V0.1	Primer Borrador
V0.2	Primera versión
V1.0	Primera versión para revisar
V1.1	Se modifica el nivel de diseminación de confidential a público

Tabla de Contenidos

1. Introducción

Este documento se divide en dos partes principales. La primera, presenta el conjunto de requisitos que se han formalizado por los socios de CAPSUL-IA (Seción 2). Dichos requisitos tienen su origen en la memoria técnica del proyecto, y recogen el resultado de las elaboraciones hechas entre los socios durante los seis primeros meses del proyecto. La organización de éstos sigue un orden jerárquico partiendo del pliego de condiciones de la convocatoria. Desde el propósito del proyecto se establece una estrategia en forma de METAS, las cuales se materializan en la definición del ente central del proyecto, la CÁPSULA. De la especificación de la CÁPSULA se desprenden los distintos objetivos ligados a la investigación (PILARES) y desarrollo (DEMOSTRADORES) ligados a las actividades del proyecto. Todo esto queda recogido y explicado en detalle a lo largo de la Seción 2.

La segunda parte del documento resume el criterio seguido y el resultado de la investigación llevada a cabo entre los socios del proyecto para la selección de las plataformas hardware sobre las cuales se evaluarán las metodologías y tecnologías propuestas en el proyecto (Seción 3).

Debe tenerse en cuenta que durante el desarrollo del proyecto, dada la naturaleza de rápida evolución de la Inteligencia Artificial y el hardware creado para su óptima y versátil ejecución, el contenido y desarrollo de este documento está condicionado por el estado del arte en el momento de su creación, pudiendo darse la posibilidad de actualización o de redefinición de algunos de los requisitos aquí presentados, los cuales serán reflejados y justificados en los posteriores documentos a entregar.

2. Requisitos

La metodología seguida para asegurar la consecución de los requisitos del proyecto ha consistido en la definición de un esquema que permita su seguimiento (trazabilidad). Este esquema, de carácter jerárquico (ver **¡Error! No se encuentra el origen de la referencia.**), parte de los objetivos definidos en la convocatoria Transmisiones 2023. Cada nivel recoge y refina los requisitos del nivel anterior de forma que la consecución de los requisitos de un nivel determinado asegura la compleción del nivel anterior (superior).

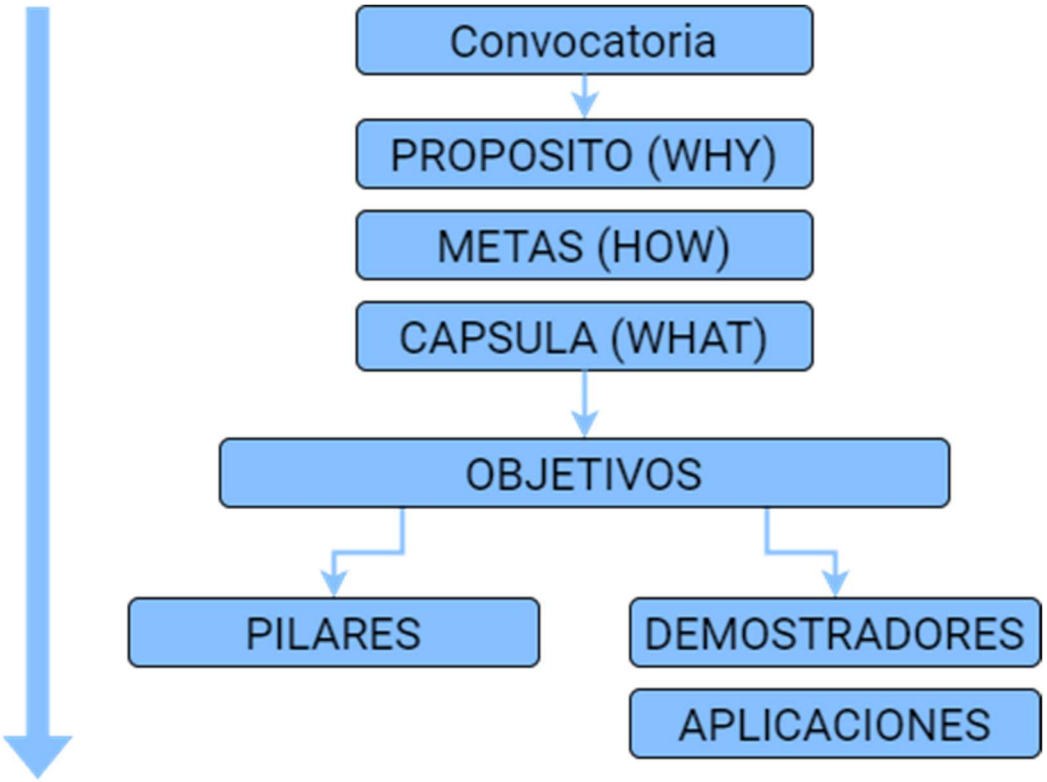


Figura 1 - Jerarquía de Requisitos de CAPSUL-IA.

2.1. Convocatoria

CAPSUL-IA se centra en contribuir al desarrollo de una industria avanzada y competitiva explotando las posibilidades de la Inteligencia Artificial (IA), específicamente técnicas de *Machine Learning* (ML). CAPSUL-IA contribuye a acercar las nuevas tecnologías mediante una propuesta que facilita el uso y su aplicación en los retos actuales de la industria 4.0. Los requisitos de la convocatoria se resumen en la Tabla *1*.

Referencia	Elemento	Origen
CONV-010	CAPSUL-IA debe beneficiar a la Industria 4.0 española.	Convocatoria
CONV-020	CAPSUL-IA debe aportar PROGRESO a la Industria 4.0 española.	Convocatoria
CONV-030	CAPSUL-IA debe aportar COMPETITIVIDAD a la Industria 4.0 española.	Convocatoria

CONV-040	CAPSUL-IA debe explotar la IA como medio para aportar PROGRESO y COMPETITIVIDAD a la Industria 4.0 española.	Convocatoria

Tabla 1 - Requisitos. Convocatoria.

2.2. CAPSUL-IA

El consorcio fue diseñado para responder con el proyecto CAPSUL-IA a la convocatoria. El proyecto cuenta con la agrupación CDTI (GTD SIR, ZYLK, PAL, SML) y la agrupación AEI (BSC e IKERLAN).

CAPSUL-IA es el proyecto de I+D (propósito, estrategia y concreción) del consorcio para responder a los requisitos de la convocatoria. Los requisitos a nivel de proyecto se especifican a continuación.

2.2.1. PROPÓSITO (WHY)

El propósito precisa la forma en que CAPSUL-IA enfoca los requisitos planteados en la convocatoria, estableciendo el objetivo último del proyecto.

Referencia	Elemento	Origen
WHY	Facilitar a la Industria 4.0 española el despliegue de todas las oportunidades que ofrece la IA, tanto en la creación de nuevos y mejores productos y servicios, como en la mejora y optimización de procesos y metodologías.	CONV-010 CONV-020 CONV-030 CONV-040

Tabla 2 - Requisitos de propósito.

2.2.2. ESTRATEGIA (HOW)

CAPSUL-IA identifica cuatro aspectos clave para la consecución del propósito del proyecto. Éstos conforman la estrategia en forma de cuatro METAS que recogen los elementos clave para la integración de la IA en la industria 4.0.

Referencia	Elemento	Origen
META-010	IA ASEQUIBLE: CAPSUL-IA debe habilitar un acceso asequible a la IA, mediante: • Soluciones prediseñadas • Soluciones integradas para una gran variedad de aplicaciones	WHY
META-020	IA ADAPTATIVA: CAPSUL-IA debe habilitar soluciones que se adapten a los requisitos del entorno de aplicación, permitiendo múltiples compromisos entre, como mínimo, las siguientes métricas: • Coste • Precisión • Velocidad [de inferencia] • Robustez	WHY
META-030	IA SEGURA: CAPSUL-IA debe habilitar soluciones íntegras y seguras, compatibles con su utilización en aplicaciones críticas. Se define aplicaciones críticas como aquellas cuyo mal funcionamiento podría resultar en la pérdida de vidas humanas, daños al medio ambiente y/o daños materiales graves.	WHY

META-040	IA CONFIABLE: CAPSUL-IA debe habilitar soluciones de IA explicables y/o auditables.	WHY

Tabla 3 - Requisitos de estrategia.

Nota: LA META-040 no fue definida en la propuesta. Todos los socios han aceptado su inclusión. Este requisito de meta captura específicamente los elementos de confiabilidad que antes se cubrían también en la META-030.

2.2.3. CONCRECIÓN (WHAT)

Para dar solución a las METAS establecidas, CAPSUL-IA plantea una propuesta técnica basada en paquetes de soluciones IA prediseñadas que resuelven arquetipos de problemas.

Una **CÁPSULA** de IA es un modelo de IA, junto con su tratamiento de los datos (de entrenamiento y para inferencia), que resuelve un cierto tipo de problema concreto, con restricciones específicas en términos de explicabilidad, seguridad funcional y de plataforma donde realizar la inferencia, y en el contexto de un marco operativo con una pila de software y un entorno potencialmente distribuido.

La instanciación de una CÁPSULA para resolver un problema específico se denomina **APLICACIÓN**.

Referencia	Elemento	Origen
CAPS-010	Una CÁPSULA debe ser suficientemente específica para resolver mediante un MODELO una tipología concreta de problemas.	META-010 META-020
CAPS-020	Una CÁPSULA debe ser suficientemente genérica para poder ser utilizada (instanciada) para distintas APLICACIONES.	META-010 META-020
CAPS-030	Una CÁPSULA debe implementar la gestión y procesamiento de los DATOS requeridos para el entrenamiento y la inferencia.	META-010 META-020
CAPS-040	Una CÁPSULA debe implementar una estrategia de selección de PLATAFORMA y preconfiguraciones para las PLATAFORMAS soportadas.	META-010 META-020
CAPS-050	Una CÁPSULA debe seguir las metodologías e incorporar los paquetes de SEGURIDAD FUNCIONAL cuando su tipología de problema pueda comportar riesgos.	META-030
CAPS-060	Una CÁPSULA debe incorporar mecanismos de EXPLICABILIDAD cuando estos sean relevantes a su tipología de problema y MODELO.	META-040
CAPS-070	Una CÁPSULA debe implementar mecanismos de OPERACIÓN y GOBERNANZA que faciliten la instanciación de la CÁPSULA en una APLICACIÓN, incluyendo el despliegue y puesta en marcha de ésta.	META-010 META-020

Tabla 4 - Requisitos de CÁPSULA.

2.3. Objetivos

CAPSUL-IA define una serie de objetivos medibles que definen la metodología de trabajo para lograr los requisitos de CÁPSULA. Estos objetivos y sus subobjetivos asociados están ligados a las actuaciones y tareas definidas en la memoria técnica del proyecto, heredando su organización temporal en el marco del proyecto.

Los OBJ1-6 se organizan entorno a los 6 PILARES de investigación que proporcionan los resultados necesarios para la consecución de las CÁPSULAS. El OBJ7 recoge estos

resultados y los integra en tres CÁPSULAS con TRL 3-4 llamadas **DEMOSTRADORES**. A su vez, estos DEMOSTRADORES se instancian en cuatro APLICACIONES con el propósito de validar la viabilidad de éstos.

Referencia	Elemento	Origen
OBJ1- MODELOS	Investigar MODELOS de IA satisfactorios con los que dotar a las CÁPSULAS. Los MODELOS deben dar respuesta a problemas industriales satisfaciendo múltiples requisitos simultáneamente en forma de métricas relacionadas con las METAS: ASEQUIBLE, ADAPTATIVA, SEGURA y CONFIABLE (coste, precisión, velocidad de inferencia, robustez, …).	CAPS-010 CAPS-020
OBJ2-DATOS	Investigar gestión, procesamiento y formato de datos. Reducir los DATOS requeridos para entrenar y para inferencia.	CAPS-030
OBJ3-PLATAFORMA	Investigar estrategias de selección y configuración de plataformas. Definir estrategias para la rápida selección de la PLATAFORMA de cómputo y su configuración que se adapte a las necesidades específicas del problema y del MODELO.	CAPS-040
OBJ4-SEGURIDAD	Investigar diseños de cápsulas de IA seguras. Dotar las cápsulas de IA con las características necesarias para poder ser integradas en sistemas críticos acorde a los estándares de seguridad funcional existentes (p.ej., IEC 61508) y emergentes (p.ej., IEC/ISO 5469).	CAPS-050
OBJ5-CONFIABILIDAD	Diseño de Mecanismos de Explicabilidad. Dotar las cápsulas de IA, especialmente las que implementen modelos de inferencia, para que puedan ofrecer una descripción eficaz y suficiente de cómo y por qué los MODELOS generan una inferencia.	CAPS-060
OBJ6-OPERACIÓN	Investigar estrategias de gobernanza del sistema. Diseño de herramientas y metodologías para agilizar la puesta en marcha y control del sistema sin impactar el cumplimiento de los requisitos de la CÁPSULA.	CAPS-070
OBJ7-APLICACIÓN	CAPSUL-IA debe realizar la prueba de concepto de la CÁPSULA mediante tres DEMOSTRADORES. • DEMO1: MODELOS de lenguaje conversacionales; • DEMO2: Optimización multicriterio explicativa; • DEMO3: Visión por computador. CAPSUL-IA debe obtener información que guíe la investigación a través de la evaluación de la eficacia y eficiencia de cuatro APLICACIONES instanciadas de los DEMOSTRADORES: • APLICACIÓN 1: DEMO 1 instanciada en la interacción con documentos legales en departamento legal. • APLICACIÓN 2: DEMO 2 instanciada en el soporte a la decisión explicativa (en centros de control). • APLICACIÓN 3: DEMO 3 instanciada en robots para manufactura y salud. • APLICACIÓN 4: DEMO 3 instanciada en el conteo de personas en tiendas para análisis de ocupación y mapas de calor.	CAPS-010 CAPS-020 CAPS-030 CAPS-040 CAPS-050 CAPS-060 CAPS-070

Tabla 5 - Objetivos de CAPSUL-IA.

2.4. Pilares (OBJ1-6)

Los Pilares son cada una de las seis (6) áreas de I+D del proyecto CAPSUL-IA que, de forma conjunta y cooperativa, permiten satisfacer los OBJETIVOS: esto es construir CÁPSULAS (WHAT) conforme a unas METAS (WHY) para gran variedad de dominios y aplicaciones (ADAPTATIVAS), así como de combinaciones de características y compromisos (ASEQUIBLES, SEGURAS, CONFIABLES).

2.4.1. MODELIZACIÓN (MOD) asociado a OBJ1

CAPSUL-IA investiga MODELOS que satisfagan requisitos específicos para satisfacer la especificación de CÁPSULA (conforme a distintas métricas). En particular, este PILAR debe obtener, al menos, los siguientes resultados:

- Tipo de modelo a usar;
- Parámetros de los modelos;
- Ensamblaje de los modelos.

Además, se investigarán, al menos, los siguientes aspectos:

- El manejo de DATOS de entrada;
- El diseño del MODELO;
- El entrenamiento del MODELO;
- La mitigación de la incertidumbre;
- El acompañamiento mediante garantías de uso y explicación.

Los requisitos del pilar de MODELIZACIÓN se listan en la Tabla *6*.

Referencia	Elemento	Origen
P01-MOD-010	CAPSUL-IA investiga la selección y adaptación de los MODELOS en función de la tarea específica a realizar.	OBJ1
P01-MOD-020	CAPSUL-IA investiga el establecimiento de conexiones con MODELOS existentes para promover la reutilización y la interoperabilidad entre distintas CÁPSULAS. En este sentido, CAPSUL-IA debe considerar la implementación de un sistema de gestión y configuración de modelos, control de versiones, y almacenamiento de datos.	OBJ1
P01-MOD-030	CAPSUL-IA debe adoptar un enfoque probabilístico de las fuentes de incertidumbre para modelar y cuantificar los tipos de errores de las tareas. En particular, CAPSUL-IA deberá abordar: • La incertidumbre epistémica del MODELO, esto es: la falta de conocimiento completo sobre el problema y sus características. • La incertidumbre aleatoria, esto es: la variabilidad inherente de los datos y su impacto en las inferencias y predicciones.	OBJ1
P01-MOD-040	CAPSUL-IA debe investigar técnicas de optimización de recursos y compresión de MODELOS para reducir la complejidad computacional y el tamaño de los propios MODELOS sin sacrificar ni su rendimiento ni su precisión.	OBJ1
P01-MOD-050	CAPSUL-IA debe investigar la evaluación, calibración y visualización de las inferencias y predicciones generadas por los MODELOS.	OBJ1

P01-MOD-060	CAPSUL-IA debe investigar la integración de los MODELOS con el software y hardware seleccionados.	OBJ1

Tabla 6 - Requisitos. Pilar de Modelización.

2.4.2. DATOS (DAT) asociado a OBJ2

CAPSUL-IA investiga el tratamiento de los datos de entrada previo a su utilización en el entrenamiento o en la inferencia con el objeto de, al menos:

- Filtrar datos irrelevantes o redundantes;
- Pretratar los datos para mejorar la efectividad del entrenamiento o la inferencia.

Los requisitos del pilar de DATOS se listan en la Tabla 7.

Referencia	Elemento	Origen
P02-DAT-010	CAPSUL-IA debe implementar una selección de las fuentes de datos, considerando su representatividad en relación con la tarea especificada.	OBJ2
P02-DAT-020	CAPSUL-IA debe implementar mecanismos para asegurar la calidad y coherencia de los datos antes de su uso en los MODELOS.	OBJ2
P02-DAT-030	CAPSUL-IA debe investigar y desarrollar métodos avanzados para la detección temprana de anomalías y ejemplos “adversarials” en los datos que puedan comprometer el rendimiento y la generalización de los MODELOS.	OBJ2
P02-DAT-040	CAPSUL-IA debe implementar políticas para garantizar la privacidad y la confidencialidad de los datos utilizados, conforme a los estándares y la regulación aplicables (legislación aplicable).	OBJ2
P02-DAT-050	CAPSUL-IA debe implementar políticas para garantizar la protección de los derechos y la equidad de las personas concernidas durante la recopilación, uso y almacenamiento de datos.	OBJ2
P02-DAT-060	CAPSUL-IA debe implementar soluciones de almacenamiento escalables y eficientes.	OBJ2
P02-DAT-070	CAPSUL-IA debe desarrollar metodologías y herramientas que faciliten la integración de fuentes y formatos heterogéneos, abordando la interoperabilidad, el mapeo de esquemas y la normalización de datos para una representación unificada y coherente.	OBJ2
P02-DAT-080	CAPSUL-IA debe implementar metodologías y métricas para evaluar la calidad de los datos utilizados en los MODELOS. Las métricas deben abordar, al menos, los siguientes aspectos: completitud, consistencia y precisión.	OBJ2
P02-DAT-090	CAPSUL-IA debe desarrollar técnicas de calibración de la calidad de los datos, incluyendo, al menos, la identificación y corrección de: Errores; Sesgos; y Desequilibrios.	OBJ2
P02-DAT-100	CAPSUL-IA debe investigar técnicas de reducción de DATOS y de transferencia de aprendizaje.	OBJ2

Tabla 7 - Requisitos. Pilar de Datos.

2.4.3. PLATAFORMA (PLAT) asociado a OBJ3

CAPSUL-IA investiga estrategias para seleccionar el nivel (bajo, medio, alto) de las plataformas objetivo y las configuraciones de estas. También estrategias de mapeo de componentes software a elementos de computación hardware. Los aspectos a tener en cuenta deben incluir una abstracción de las propiedades que permita valorar si una propuesta de plataforma, configuración, y mapeo cumple con los requisitos de una cápsula. Las métricas a tener en cuenta son las siguientes:

- Rendimiento con interés en la contención entre diferentes aplicaciones de IA y con otras aplicaciones que no son de IA;
- Consumo de energía; y
- Seguridad

La Tabla *8* recoge los requisitos identificados a nivel del pilar de PLATAFORMA.

Referencia	Elemento	Origen
P03-PLAT-010	CAPSUL-IA debe diseñar al menos una estrategia de exploración de opciones de configuración y optimización de varias métricas (multi-objective optimization).	OBJ3
P03-PLAT-020	CAPSUL-IA debe diseñar una estrategia de segregación de recursos que permita dar garantías para distintas métricas. E.g.: El tiempo requerido para hacer una inferencia cuando haya requisitos de tiempo real.	OBJ3
P03-PLAT-030	CAPSUL-IA identificará una metodología para mapear los distintos componentes de una cápsula de IA a los elementos de computación de la plataforma.	OBJ3
P03-PLAT-040	CAPSUL-IA identificará al menos una abstracción que permita razonar acerca de las propiedades de la plataforma y relacionarlas con los requisitos de la cápsula en los términos definidos por las métricas de interés de la cápsula.	OBJ3
P03-PLAT-050	Se definirán al menos dos pre-configuraciones de la plataforma (PCP) o arquetipos de configuración por plataforma.	OBJ3 P03-PLAT-040
P03-PLAT-060	Las tecnologías del pilar de plataforma se probarán en 3 plataforma, una de cada tipo definido: ALTO, MEDIO, y BAJO	OBJ3

Tabla 8 - Requisitos. Pilar de Plataforma.

2.4.4. SEGURIDAD FUNCIONAL (FUSA) asociado a OBJ4

CAPSUL-IA adecua los procesos de desarrollo de las CÁPSULAS (y su posterior funcionamiento) para:

- Cumplir con los requerimientos de la Seguridad Funcional; y
- Garantizar la integridad [de la función implementada].

CAPSUL-IA investiga la instanciación de diferentes paquetes de seguridad funcional que capturen los requerimientos de seguridad de una CÁPSULA, cuando la CÁPSULA se utilice en sistemas críticos; o en sistemas relacionados con la seguridad funcional. En esta línea la Tabla 9 lista los requisitos a nivel de Seguridad funcional (FUSA).

Referencia	Elemento	Origen
P04-FUSA-010	CAPSUL-IA debe implementar un proceso de desarrollo para IA (conforme al ciclo de vida de [desarrollo de] sistemas seguros). El objeto de un proceso de desarrollo controlado es la mitigación de los errores sistemáticos. Este proceso de desarrollo a implementar debe contemplar, al menos: • Mecanismos para analizar el impacto de las decisiones de diseño propias de la IA; • Procesos para documentar los argumentos de seguridad; y las evidencias en línea con los procesos de certificación. Para la especificación de dicho proceso se realizará un estudio previo de los estándares existentes y emergentes.	OBJ4 OBJ5
P04-FUSA-020	CAPSUL-IA debe investigar e implementar mecanismos para la correcta gestión de los errores durante la ejecución de las CÁPSULAS (p.ej. fallos e insuficiencias en modelos). Estos mecanismos deben ser, al menos: • La definición de las arquitecturas de seguridad; • Mecanismos de diagnóstico; • Mecanismos de tolerancia; y • Mecanismos de reacción.	OBJ4
P04-FUSA-030	CAPSUL-IA debe validar los mecanismos de diagnóstico y elementos de supervisión.	OBJ4
P04-FUSA-040	CAPSUL-IA debe definir una arquitectura modular que permita desacoplar la CÁPSULA de IA y el integrador del sistema de seguridad.	OBJ4
P04-FUSA-050	CAPSUL-IA debe definir las restricciones y consideraciones a tener en cuenta en la definición de la arquitectura modular para la inferencia e integración segura.	OBJ4
P04-FUSA-060	CAPSUL-IA debe desarrollar mecanismos que faciliten el despliegue mediante paquetes de seguridad funcional que acompañen a las CÁPSULAS IA.	OBJ4

Tabla 9 - Requisitos. Pilar de FUSA.

2.4.5. EXPLICABILIDAD (EXP) asociado a OBJ5

La explicabilidad o interpretabilidad en IA se refiere a la capacidad de entender y comunicar cómo y por qué un modelo de IA toma decisiones específicas. Este concepto es crucial, especialmente en sistemas de IA complejos como los modelos de aprendizaje profundo, que suelen funcionar como "cajas negras", es decir, producen resultados sin que se pueda observar directamente el proceso que los genera.

En aplicaciones industriales, la explicabilidad es esencial para que los ingenieros y operadores puedan entender y confiar en las decisiones tomadas por el sistema de IA. Además, la explicabilidad es clave para cumplir con normativas y estándares de seguridad, ya que proporciona una vía para auditar y validar el comportamiento del sistema, asegurando que funcione dentro de los límites aceptables y no cause fallos inesperados.

CAPSUL-IA investiga los métodos y modelos de IA más convenientes para facilitar la interpretabilidad de una inferencia a un operador. Además, se investiga también la capacidad de auditar el modelo IA de una CÁPSULA mediante la trazabilidad de los datos y la reconstrucción de la inferencia. Los requisitos de explicabilidad se recogen en la Tabla *10*.

Referencia	Elemento	Origen
P05-EXP-010	CAPSUL-IA debe investigar métodos y técnicas que permitan comprender y explicar cómo los MODELOS toman decisiones. Esto debe implicar, al menos: • Entender los procesos internos del MODELO. • Entender cómo se relacionan las decisiones con los datos de entrada.	OBJ5
P05-EXP-020	CAPSUL-IA debe investigar MODELOS confiables orientados a la toma de decisión en entornos industriales que permitan generar la explicación de la inferencia.	OBJ5
P05-EXP-030	CAPSUL-IA debe investigar métodos para comunicar y visualizar de forma efectiva la explicación de la inferencia.	OBJ5
P05-EXP-040	CAPSUL-IA debe investigar métodos de validación y verificación de la explicación de la inferencia, definiendo métricas para ambos casos.	OBJ5
P05-EXP-050	CAPSUL-IA debe implementar un registro completo y transparente de los datos utilizados para entrenar y operar los MODELOS. Este registro debe permitir auditar y comprender el comportamiento del MODELO.	OBJ5

Tabla 10 - Requisitos. Pilar de Modelización.

2.4.6. OPERACIÓN (OPE) asociado a OBJ6

CAPSUL-IA investiga técnicas y metodologías que aporten robustez y gobernabilidad a la CÁPSULA para proveer, al menos:

- Alta disponibilidad;
- Redundancia; y
- Seguridad.

CAPSUL-IA investiga mecanismos y procedimientos para aprovisionar e implementar toda la arquitectura de software sobre cada plataforma; y a partir de los “artefactos” que proveen de Sistema Operativo y librerías necesarias (dependencias, paquetes, etc.), a la plataforma hardware. El PILAR de OPERACIÓN utiliza el resultado de la investigación del resto de socios (datos, código, aplicaciones, y artefactos) para construir una infraestructura que dote de robustez y tolerancia a fallos las aplicaciones y sistemas que surgen de la compilación y ejecución de estos artefactos. Para ello se utilizan técnica de infraestructura como código (IaC), virtualización, orquestación, entrega continua y MLOps (DevOps para sistemas de IA). Ver Tabla *11*.

Nota: Los artefactos son el resultado de la investigación del resto de PILARES y el punto de partida de la investigación en el PILAR de OPERACIÓN.

Referencia	Elemento	Origen
P06-OPE-010	CAPSUL-IA debe investigar herramientas de Infraestructura como Código (IaC) que permitan aprovisionar y configurar nodos de forma autónoma y automática para escalar el número de máquinas (ya sean físicas o virtuales) en función de los recursos y universalizar el aprovisionamiento de recursos de software sobre las plataformas hardware.	OBJ6
P06-OPE-020	CAPSUL-IA debe investigar herramientas de gestión de configuración, implementación de aplicaciones y organización de paquetes de software.	OBJ6
P06-OPE-030	CAPSUL-IA debe testear la instalación y el despliegue de CÁPSULAS.	OBJ6
P06-OPE-040	CAPSUL-IA debe investigar mecanismos que aseguren la robustez, fiabilidad y alta disponibilidad de las CÁPSULAS.	OBJ6
P06-OPE-050	CAPSUL-IA debe investigar métodos de despliegue e integración continua que permitan actualizar los recursos software con "zero-downtime" (orientado a entornos y aplicaciones críticas).	OBJ6
P06-OPE-060	CAPSUL-IA investigará métodos de interacción con usuarios.	OBJ6

Tabla 11 - Requisitos. Pilar de Operación.

2.5. DEMOSTRADORES Y APLICACIONES (OBJ7)

CAPSUL-IA concibe las cápsulas a desarrollar para ser instanciadas por aplicaciones concretas. En esta sección se identifican las cápsulas concretas a investigar y desarrollar. Estas cápsulas, en forma de DEMOSTRADORES en TRL3 o TRL4, servirán como prueba de concepto y se instanciarán para APLICACIONES específicas como medio para detectar limitaciones y errores que nos permitan retroalimentar los resultados de las investigaciones para alcanzar mejores soluciones.

2.5.1. DEMOSTRADORES

CAPSUL-IA plantea desarrollar pruebas conceptuales como medio para validar la relevancia y viabilidad, incluso de cara a una futura explotación industrial, de los conceptos investigados, para lo cual también se materializarán los DEMOSTARDORES.

En la Tabla *12* se muestran los requisitos a nivel de demostrador específico identificados, estos derivan de la funcionalidad especificada en OBJ7 y de los propios requisitos definidos para las cápsulas.

Referencia	Elemento	Origen
DEMO1	**Prueba de concepto de una CÁPSULA de MODELOS de lenguaje conversacionales**	**OBJ7**

DEMO1-010	DEMO1 debe utilizar un MODELO de Lenguaje de Aprendizaje Automático (LLM).	CAPS-010
DEMO1-020	DEMO1 debe proveer herramientas para poder hacer fine-tuning del modelo en caso de que sea posible.	CAPS-020
DEMO1-030	DEMO1 debe permitir utilizar una base de datos de embeddings con representaciones vectoriales de los documentos relacionada para una búsqueda y recuperación eficiente.	CAPS-030
DEMO1-040	DEMO1 debe tener mecanismos para permitir interactuar con documentación confidencial de forma privada y sin comprometer su confidencialidad (por ejemplo: sin subirla a la nube o sin recurrir a aplicaciones de terceros), en los casos de que los modelos utilizados lo permitan.	CAPS-030
DEMO1-050	DEMO1 debe estar adaptada a su utilización sobre una plataforma adecuada, incorporando preconfiguraciones para su uso.	CAPS-040
DEMO1-060	DEMO 1 debe utilizar prompt engineering, adaptados a cada aplicación, para mejorar la calidad y relevancia de las respuestas generadas por su MODELO.	CAPS-070
DEMO1-070	DEMO1 debe proveer de una interfaz gráfica que cumpla las necesidades básicas para poder interactuar con la cápsula.	CAPS-070
DEMO2	**Prueba de concepto de una CÁPSULA de optimización multicriterio explicativa**	**OBJ7**
DEMO2-010	DEMO2 debe servir como ayuda en la toma de decisiones de los operarios en centros de control industriales basándose en optimización multicriterio explicativa.	OBJ7
DEMO2-020	DEMO2 debe implementar un MODELO que permita la acción de control en tiempo real y que permita al operario del centro de control entender el estado de la planta.	CAPS-010
DEMO2-030	DEMO2 debe permitir el reentrenamiento completo para los MODELOS más simples y la adecuación o *fine-tuning* para los MODELOS más complejos.	CAPS-020
DEMO2-040	DEMO2 debe implementar una interfaz de comunicación con una base de DATOS industrial que le permita adquirir y preprocesar los DATOS necesarios para el reentrenamiento y la operación.	CAPS-030
DEMO2-050	DEMO2 debe estar adaptada a su utilización sobre una plataforma adecuada, incorporando preconfiguraciones para su uso.	CAPS-040
DEMO2-060	DEMO2 debe implementar el ciclo de vida de seguridad funcional, posibilitando su uso en entornos de operación industrial. Este ciclo de vida debe incluir metodologías y procedimientos a cumplir en las APLICACIONES instanciadas de esta CÁPSULA.	CAPS-050
DEMO2-061	DEMO2 debe implementar indicadores de confianza en la inferencia del control en tiempo real.	CAPS-050

DEMO2-062	DEMO2 debe mantener completa trazabilidad de su operación, posibilitando su reconstrucción para posteriores análisis.	CAPS-050
DEMO2-070	DEMO2 debe implementar un HMI (Human Machine Interface) que permita al operario del centro de control interactuar con el MODELO.	CAPS-060
DEMO2-071	DEMO2 debe implementar un HMI que permita comprender las inferencias del MODELO y el estado del proceso, utilizando elementos gráficos para facilitar la interpretación de los datos.	CAPS-060
DEMO2-072	DEMO2 debe implementar mecanismos que permitan al operario actuar sobre el MODELO a través del HMI.	CAPS-060
DEMO3	**Prueba de concepto de una CÁPSULA de visión por computador**	**OBJ7**
DEMO3-010	DEMO3 debe extraer características significativas de las imágenes y permitir la detección y clasificación precisa de objetos mediante la utilización de arquitecturas de redes neuronales profundas.	CAPS-010
DEMO3-011	En problemas de clasificación, la salida debe consistir en la anotación de la clase correcta de objeto correspondiente.	CAPS-010
DEMO3-012	En problemas de detección, la salida debe consistir en la anotación de la clase de los objetos presentes y su clase asociada, así como sus coordenadas.	CAPS-010
DEMO3-020	DEMO3 debe poder ser entrenable de forma eficiente y escalable.	CAPS-020
DEMO3-030	DEMO3 debe poder ser afinado (finetuning) para mejorar la capacidad de clasificación y detección en tareas específicas.	CAPS-020
DEMO3-040	DEMO3 debe poder reaprovechar modelos pre-entrenados en grandes conjuntos de datos (Transfer Learning) para acelerar el entrenamiento y mejorar el rendimiento.	CAPS-030
DEMO3-050	DEMO3 debe poder ser alimentado mediante un conjunto amplio y diverso de datos etiquetados que contengan imágenes (o videos) representando los objetos a clasificar o detectar. Se proporcionarán etiquetas para las diferentes categorías de objetos a clasificar. El conjunto de datos etiquetados será preprocesado para asegurar su calidad y uniformidad (se utilizarán formatos comunes: JPEG, PNG, con una resolución uniforme recomendada). El preprocesado incluirá, entre otros: el redimensionado, la normalización de colores y la eliminación de ruido.	CAPS-030
DEMO3-060	DEMO3 debe ser lo suficientemente flexible y configurable para poder ser instanciada y desplegada en diferentes PLATAFORMAS.	CAPS-040
DEMO3-070	DEMO3 debe utilizar los datos para ajustar parámetros, hiperparámetros y evaluar el rendimiento del MODELO. Los datos disponibles se dividirán en tres conjuntos: entrenamiento, validación y prueba.	CAPS-030
DEMO3-080	Se aplicarán técnicas para garantizar la seguridad del sistema.	CAPS-050

DEMO3-090	DEMO3 debe estar preparada para aplicar mecanismos de explicabilidad dependiendo de las necesidades de las aplicaciones en las que se instancie el demostrador.	CAPS-060
DEMO3-100	DEMO3 debe proveer de una interfaz gráfica que cumpla las necesidades básicas para poder interactuar con el modelo.	CAPS-070

Tabla 12 - Requisitos. Demostradores.

2.5.2. APLICACIONES

Las aplicaciones permiten evaluar la eficacia y eficiencia de las soluciones investigadas en los distintos pilares, actuando como pruebas de concepto. Se han seleccionado 4 aplicaciones que permiten seleccionar todas las cápsulas del proyecto:

- **Aplicación1** (DEMO 1): Interacción con documentos legales en departamento legal
- **Aplicación2** (DEMO 2): Soporte a la decisión explicativo (en centros de control)
- **Aplicación3** (DEMO 3): Robots en Manufactura y Salud
- **Aplicación4** (DEMO 3): Conteo de personas en tiendas para análisis de ocupación y mapas de calor

Los requisitos por aplicación se presentan en la Tabla *13*.

Referencia	Elemento	Origen
APP1	**Interacción con documentos legales en departamento legal, como APLICACIÓN instanciada de DEMO 1.**	**OBJ7**
APP1-010	La APLICACIÓN 1 debe instanciar DEMO1 con el fin de probar la CÁPSULA en la interacción ágil y precisa con documentos legales confidenciales de un departamento legal o bufete de abogados.	DEMO1-010 DEMO1-070
APP1-020	La APLICACIÓN 1 debe permitir realizar consultas sobre cláusulas similares, buscar jurisprudencia relevante y recuperar información clave para casos actuales y pasados.	DEMO1-010
APP1-030	La APLICACIÓN 1 debe hacer fine-tuning del MODELO con el fin de que este se adapte mejor a las consultas.	DEMO1-020
APP1-040	La APLICACIÓN 1 debe implementar una base de datos de embeddings con las representaciones vectoriales de los documentos legales empleados en la aplicación.	DEMO1-030
APP1-050	La APLICACIÓN 1 debe ser desplegada sobre la plataforma MEDIO con el fin de que los datos se mantengan en local y se pueda garantizar la privacidad de estos. Además, se investigará el uso de la plataforma ALTO.	DEMO1-040 DEMO1-050

APP1-040	La APLICACIÓN 1 debe proveer una interfaz adecuada que permita la correcta interacción del usuario final con el modelo. En esta interfaz se deberá poder: • Interactuar con el modelo mediante mensajes de texto. • Recibir las respuestas del modelo y ser claramente diferenciadas de los mensajes propios del usuario. • Gestión documental: se podrá consultar, eliminar o añadir documentos que forman parte de la base de datos.	DEMO1-060 DEMO1-070
APP2	**Instanciar DEMO2 en una APLICACIÓN de sistema de soporte a la decisión en la sala de control de un sistema distribuido.**	**OBJ7**
APP2-010	La APLICACIÓN 2 debe instanciar DEMO2 con el fin de probar la CÁPSULA en la toma de decisión en salas de control de sistemas distribuidos.	OBJ7 DEMO2-010
APP2-020	La APLICACIÓN 2 debe optimizar en tiempo real las acciones de control sobre un sistema distribuido de generación eléctrica. Permitiendo al operario conocer el estado del sistema y el proceso de inferencia de las acciones de control.	DEMO2-020
APP2-030	La APLICACIÓN 2 debe comunicarse con una base de datos de prueba que simule la operación real de la planta.	DEMO2-040
APP2-040	La APLICACIÓN 2 debe ser desplegada sobre la plataforma ALTO con el fin de que el bucle de control sea local y no dependa de conexiones a la red.	DEMO2-050
APP2-050	La APLICACIÓN 2 debe seguir el ciclo de vida de seguridad funcional con el fin de evitar pérdidas económicas derivadas de fallos en las acciones de control.	DEMO2-060
APP2-051	La APLICACIÓN 2 debe contar con indicadores de confianza en la optimización basados en la calidad de los datos	DEMO2-061
APP2-052	La APLICACIÓN 2 debe mantener la trazabilidad de sus datos y posibilitar la reconstrucción completa del proceso de control.	DEMO2-062
APP2-060	La APLICACIÓN 2 debe contar con un HMI grafico que permita al operador visualizar el estado de la planta, diferentes métricas relativas a la generación eléctrica y el estado de la optimización de las acciones de control.	DEMO2-071
APP2-061	La APLICACIÓN 2 de permitir a través del HMI configurar el proceso de optimización de las acciones de control.	DEMO2-072
APP3	**Visión por computador, como APLICACIÓN instanciada de DEMO 3.**	**OBJ7**
APP3-010	La APLICACIÓN 3 debe instanciar DEMO3 para probar la CÁPSULA en aplicaciones de robótica que impliquen la detección de objetos y conteo de personas en un entorno de trabajo/fabricación.	DEMO3-010

APP3-020	La APLICACIÓN 3 debe detectar y localizar (bounding boxes) en tiempo real un conjunto de clases de objetos y personas objetivo a partir de un flujo de vídeo producido por una cámara robótica.	DEMO3-011 DEMO3-012 DEMO3-050
APP3-030	La APLICACIÓN 3 debe permitir añadir fácilmente nuevas clases de objetos o personas al sistema proporcionando imágenes anotadas de muestra a la CÁPSULA mediante un proceso racionalizado.	DEMO3-020 DEMO3-030 DEMO3-040 DEMO3-050
APP3-040	La APLICACIÓN 3 debe ser desplegada sobre al menos la plataforma MEDIO.	DEMO3-060
APP3-050	La APLICACIÓN 3 puede exponer en cualquier momento la lista de parámetros e hiperparámetros utilizados por la CÁPSULA, para garantizar la trazabilidad y la reproducibilidad.	DEMO3-070
APP3-060	Dado que la APLICACIÓN 3 podría tratar datos personales (incluyendo, por ejemplo, imágenes de personas físicas), la APLICACIÓN 3 debe ofrecer garantías sólidas sobre la privacidad y seguridad de los datos utilizados, así como de las personas que puedan aparecer en ellos.	DEMO3-080
APP3-070	La APLICACIÓN 3 debe permitir visualizar las imágenes anotadas en tiempo real	
APP3-080	La APLICACIÓN 3 debe proveer una interfaz adecuada que permita la correcta interacción del usuario final con el modelo.	DEMO3-100
APP4	**Conteo de Personas en Tiendas para Análisis de Ocupación y Mapas de Calor, como APLICACIÓN instanciada de DEMO 3.**	**OBJ7**
APP4-010	La APLICACIÓN 4 debe instanciar DEMO3 con el fin de probar la CÁPSULA en el conteo de personas en una tienda utilizando imágenes previamente grabadas.	DEMO3-010
APP4-020	La APLICACIÓN 4 debe detectar, contabilizar y obtener de estadísticas de ocupación y patrones de tráfico.	DEMO3-012
APP4-030	La APLICACIÓN 4 debe emplear MODELOS entrenados exprofeso para el problema o MODELOS pre-entrenados en grandes conjuntos de imágenes con el fin de evitar el sobreajuste del MODELO a fuentes de imágenes concretas.	DEMO2-040
APP4-040	La APLICACIÓN 4 debe afinar el MODELO mediante *fine-tuning* para mejorar el rendimiento sobre las fuentes de imágenes empleadas en la APLICACION.	DEMO3-030
APP4-050	Se investigará la posibilidad de preparación del aplicativo para funcionar con imágenes/video en tiempo real.	DEMO3-050
APP4-060	La APLICACIÓN 4 debe ser desplegada sobre al menos la plataforma MEDIO.	DEMO3-060

APP4-070	La APLICACIÓN 4 debe contar con mecanismos de explicabilidad que ayuden al usuario a entender el proceso de inferencia mediante mapas de calor.	DEMO3-090
APP4-080	La APLICACIÓN 4 debe proveer una interfaz adecuada que permita la correcta interacción del usuario final con el modelo.	DEMO3-100

Tabla 13 - Requisitos. Aplicación.

3. Selección de plataformas

A continuación, se recoge el trabajo realizado para la selección de las plataformas y las pilas de software correspondientes sobre las que se ejecutará el proyecto.

Las cápsulas a desarrollar en el proyecto se conciben para funcionar en plataformas con diferentes relaciones entre rendimiento y precio. En la propuesta del proyecto se identificaron tres arquetipos de plataforma representativas del estado del arte.

- **BAJO**. Plataformas diseñadas principalmente para el *edge*, donde el bajo consumo de energía y coste son dos factores determinantes. Esto va emparejado con un conjunto más limitado en número y variedad de aceleradores.
- **MEDIO**. Plataformas que balancean coste, consumo y métricas similares.
- **ALTO**. Plataformas con las mejores innovaciones de cada fabricante y que tienen como objetivo un máximo rendimiento bajo unas restricciones de consumo o temperatura (fijadas por lo que realmente puede soportar la plataforma).

El propósito de estas plataformas es ejecutar la inferencia de los MODELOS de las CÁPSULAS atendiendo a los requisitos de las APLICACIONES específicas.

Existen diferentes tipologías de plataformas, que generalmente balancean entre flexibilidad y eficiencia en la inferencia:

Application-Specific Integrated Circuits (ASICs)

Los ASICs están diseñados para una tarea específica y pueden ofrecer un rendimiento y eficiencia energética sobresalientes. Por el contrario, al ser tan especializados su rango de aplicaciones suele ser limitado.

Unidades de Procesamiento Tensorial (TPUs)

Desarrolladas por Google, las TPUs están diseñadas específicamente para cargas de trabajo de aprendizaje profundo. Ofrecen alto rendimiento y eficiencia energética para la inferencia y el entrenamiento de redes neuronales, integrándose con TensorFlow. Sin embargo, son muy poco flexibles en los modelos a ejecutar.

Unidades de Procesamiento Gráfico (GPUs)

Las GPUs, como las ofrecidas por NVIDIA y AMD, han sido pioneras en el campo de la inteligencia artificial. Su arquitectura paralela las hace ideales para operaciones de matrices y vectores, componentes clave de los cálculos de redes neuronales. Ejemplos destacados incluyen las series NVIDIA Tesla y A100.

Field Programmable Gate Arrays (FPGAs)

Los FPGAs, como los de Xilinx e Intel, ofrecen una flexibilidad excepcional al permitir la personalización de la arquitectura del hardware para tareas específicas de IA. Esta flexibilidad puede resultar en alta eficiencia y rendimiento para aplicaciones específicas.

Sistemas en un Chip (SoCs)

Los SoCs combinan múltiples componentes en un solo chip, incluyendo CPU, GPU y otros aceleradores. Estos sistemas, como los de la línea Jetson de NVIDIA o las Ultrascale de Xilinx, están diseñados para aplicaciones locales, donde la inferencia necesita realizarse cerca del lugar donde se generan los datos.

3.1. Plataformas hardware

Durante la reunión de lanzamiento del proyecto, y las semanas siguientes, los socios han revisado los requisitos de los casos de estudio en términos de rendimiento y soporte de librerías. Durante la primera reunión, todos los socios del proyecto coincidieron en la conveniencia de mantener el número de arquetipos de plataformas en 3. Esto cubre la necesidad del proyecto de mostrar sus beneficios en una plataforma para varios segmentos del mercado a la vez que se adecua esfuerzo a invertir al tamaño del proyecto.

En base a una investigación de mercado, las plataformas propuestas fueron:

PLATAFORMA	CAPACIDAD	RENDIMIENTO	SOPORTE SOFTWARE	SEGURIDAD
NVIDIA Orin Nano	BAJA	ALTO	ALTO	MEDIA
Jetson Orin NX 16GB	MEDIA-BAJA	ALTO	ALTO	MEDIA
Jetson Orin AGX 32/64GB	ALTA	ALTO	ALTO	MEDIA-ALTA
Xilinx VERSAL Core	MUY ALTA	MUY ALTO	MEDIO	MEDIA-ALTA
Xilinx Versal Edge	MEDIA-ALTA	MUY ALTO	MEDIO	MEDIA-ALTA
Xilinx Zynq UltraScale+	MEDIA	MUY ALTO	BAJO	MEDIA-ALTA
Safety Platform (ST, Aurix, ...)	BAJA	MEDIO	BAJO	ALTA

Tabla 14 - Comparación de plataformas hardware

Con el fin de facilitar el proceso de selección, se pidió a los socios que expresaran sus requisitos iniciales en la pila de hardware (HW)/software (SW). A nivel de hardware (plataforma), los resultados de esta actividad se resumen en la Tabla 15**¡Error! No se encuentra el origen de la referencia.**.

¿Qué plataforma?	GTD	PAL	SML	ZYL	BSC	IKR
NVIDIA Orin Nano			X	X	X	X
Jetson Orin NX 16GB				X		
Jetson Orin AGX 32GB	-	X		X		
Jetson Orin AGX 64GB	X		X			X
Xilinx VERSAL Core	X				X	~
Xilinx Versal Edge						
Xilinx Zynq UltraScale+						
Safety Platform (ST, Aurix, ...)						~

Tabla 15 - Requisitos de plataforma HW expresados por los socios de CAPSUL-IA.

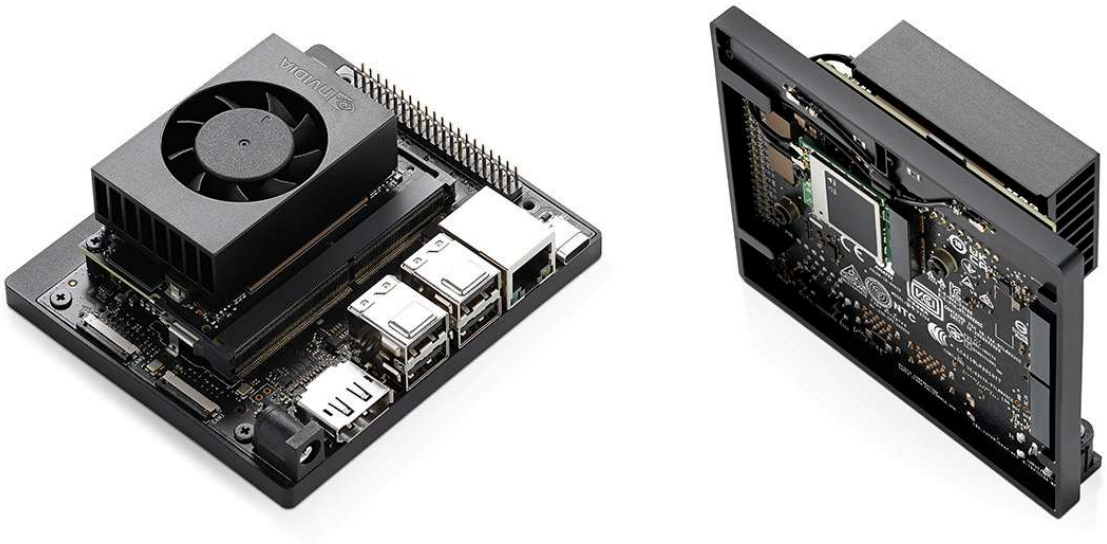
A partir del análisis de esos requisitos,

- **BAJO** La serie NVIDIA Jetson Orin Nano [1] ha sido seleccionada por su excelente eficiencia consumo-rendimiento, clave para entornos Edge.
- **MEDIO**. Los socios han seleccionado la serie Jetson Orin AGX 32GB o 64GB [1] (este último es el preferido debido al mayor tamaño de la memoria) por ser la plataforma que mejor balanceaba coste, consumo y rendimiento en este segmento.
- **ALTO**. Los socios seleccionaron una plataforma que proporciona un gran rendimiento y tiene CPUs, GPUs, y DLA (Deep Learning Accelerators). En particular ha sido la AMD Xilinx Versal [2].

El coste ha sido tenido en cuenta de forma que el precio de todas las plataformas esté alrededor de los diez mil euros.

3.1.1. PLATAFORMAS JETSON

La familia Jetson de NVIDIA está basada en dispositivos GPU Ampere adaptados al procesado en *edge*. Junto con doce núcleos CPU ARM-78AE en la versión Orin AGX y seis en la Orin Nano, estas plataformas ofrecen una gran capacidad de computación paralela (275 y 40 TOPS), ideal para algoritmos de IA con un eficiente consumo eléctrico (60 y 15 W).



Además de su potente hardware, los productos Jetson de NVIDIA mantienen un alto nivel de integración con el software empleado en el desarrollo y entrenamiento de modelos de IA. Esto incluye el uso de herramientas y bibliotecas optimizadas, como TensorRT para la inferencia de redes neuronales, NVIDIA DeepStream para análisis de vídeo en tiempo real, y cuDNN para el procesamiento eficiente de redes neuronales profundas. Para la gestión de contenedores y el despliegue de aplicaciones, NVIDIA ofrece soporte para Docker y Kubernetes, facilitando la implementación y escalabilidad de soluciones de IA.

El stack de software de Jetson se complementa con el SDK de JetPack, que proporciona un conjunto completo de herramientas y bibliotecas para desarrollar aplicaciones de IA y computación acelerada. JetPack incluye el sistema operativo basado en Ubuntu, controladores, bibliotecas de CUDA, y herramientas de desarrollo como Nsight Systems y Nsight Graphics. Este entorno de desarrollo integral permite a los desarrolladores construir, depurar y optimizar aplicaciones con alto rendimiento.

A pesar de esta integración avanzada, una mayor personalización y optimización de las soluciones de IA puede requerir una curva de aprendizaje pronunciada debido al software propietario de NVIDIA. Sin embargo, la empresa proporciona una amplia documentación, tutoriales y una comunidad activa para ayudar a los desarrolladores a superar estos desafíos.

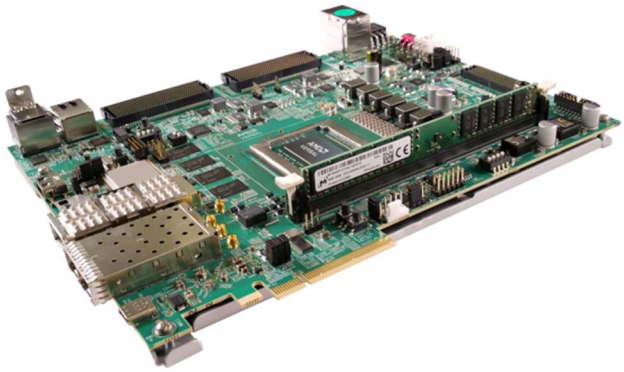
El precio de estos dispositivos puede ser relativamente elevado frente a plataformas de capacidad similar. Sin embargo, las plataformas de desarrollo Jetson son extremadamente versátiles y empiezan a contar con certificaciones que las acercan a su uso industrial en operaciones críticas. La robustez y fiabilidad de las plataformas Jetson las convierten en

una opción atractiva para aplicaciones en sectores como la robótica, vehículos autónomos, ciudades inteligentes y la automatización industrial.

3.1.2. PLATAFORMA VERSAL CORE

La familia Versal de Xilinx está basada en la arquitectura adaptable ACAP (Adaptive Compute Acceleration Platform) que combina núcleos de CPU, GPU y FPGA en un único SoC. Los dispositivos Versal incluyen núcleos de CPU ARM Cortex-A72 duales y hasta 400 núcleos de inteligencia artificial (AI Engines), proporcionando una capacidad de computación versátil y de alto rendimiento. Esta arquitectura heterogénea permite alcanzar un rendimiento excepcional en tareas de computación intensiva y procesamiento de datos.

Los productos Versal de Xilinx ofrecen una integración con el software empleado en el desarrollo y despliegue de aplicaciones de IA y computación acelerada, aunque se conocen ciertas limitaciones en este aspecto debido al hardware específico.



Xilinx proporciona Vitis AI, un entorno de desarrollo unificado que facilita la creación y optimización de aplicaciones de IA. Este se complementa con el Vitis Unified Software Platform, que proporciona un entorno completo para desarrollar aplicaciones en hardware Xilinx.

La posibilidad de usar este software de alto nivel facilita el desarrollo en las Versal, sin embargo, a menudo puede ser necesario hacer implementaciones de bajo nivel accediendo directamente a los recursos de la FPGA, lo cual supone una gran barrera de entrada para desarrolladores no familirizados con la descripción hardware (VHDL/Verilog).

El precio de los dispositivos Versal puede ser elevado en comparación con otras plataformas de capacidad similar. No obstante, los dispositivos FPGA cuentan con un largo historial de uso en aplicaciones críticas no solo en el sector de la industria sino en el espacial, sanitario y militar.

3.2. Pila de Software

Una vez que todos los socios han acordado las plataformas objetivos comunes para las tres categorías (ALTO, MEDIO, BAJO), es importante identificar elementos relevantes en las pilas de software (SO, bibliotecas, compiladores, etc.). Esto último tiene como objetivo evitar una desalineación en la forma en que el software debe compilarse y ejecutarse en la placa objetivo: dicha desalineación, de hecho, podría impedir el necesario intercambio de artefactos de software entre los socios y potencialmente invalidando los supuestos sobre la explicabilidad y el rendimiento. Esto es particularmente cierto en plataformas de vanguardia como las consideradas en CAPSUL-IA, donde la pila de software generalmente está adaptada a la plataforma y, en consecuencia, a menudo es inestable, lo que a su vez implica actualizaciones frecuentes. La presencia de una gran cantidad de elementos en la pila de software y las frecuentes actualizaciones y parches aumentan la probabilidad de que se produzcan incompatibilidades de software (a menudo inesperadas e indocumentadas).

No obstante, se ha hecho un esfuerzo por identificar preliminarmente los principales elementos que se espera que se incluyan en la pila de software para todos los socios y casos de estudio. A continuación, se reportan los principales elementos por socio. Además de la identificación de herramientas y bibliotecas en la pila adoptada, la configuración final de la plataforma depende de la versión específica de cada elemento, ya que esto puede afectar en gran medida las funcionalidades y la compatibilidad proporcionadas. Por estos motivos, siempre que ha sido posible, identificamos una versión tentativa, basada en el soporte declarado de la plataforma. Sin embargo, dicha información se consolidará solo después de que las placas se hayan adquirido y configurado por completo, capturando así las actualizaciones/parches de software necesarios y posibles incompatibilidades. En ese momento, también podremos fijar la configuración de trabajo básica por objetivo y por socio.

En la Tabla 16**¡Error! No se encuentra el origen de la referencia.** se recogen los requisitos identificados por cada socio sobre los diferentes elementos de la pila de software.

¿Qué SW tool/component?	GTD	PAL	SML	ZYL	BSC	IKR
Operating System. Linux						
Linux (Ubuntu/Debian)	X	X	X	X	X	X
Ubuntu 22.04		X	X			
ROS 2 (Humble)		X				
Software packages						
Container / Docker	X	X	X	X	X	X
Tensorflow		X		X	X	X
Pytorch	X	X	X	X	X	X
Other AI Libraries (specify)						
ONNX (to share ML models within robot)	X	~	X			
Darknet					X	X
Caffe						
Other demo specific						
Programming languages supported						
Java				X		
Python	X	X	X	X	X	X
C/C++	X	X	X		X	X
CUDA	X				X	X
Other						
PostgreSQL	X	X	X	X		
Timescale	X					
MariaDB			X			
NoSQL (Cassandra, MongoDB)			X			
SQLite		X	X			
Chroma			X			
VHDL	X					

Tabla 16 - Requisitos sobre la pila de software expresado por los socios de CAPSUL-IA.

El objetivo principal de la selección de la pila de software ha sido identificar las bibliotecas de soporte y la cadena de herramientas que se usarán como parte del entorno de desarrollo de software CAPSUL-IA.

Como se mencionó anteriormente, las versiones de software se especificarán en detalle una vez que las plataformas estén configuradas correctamente.

3.3. Estado

En el momento de la escritura de este entregable todos los socios han adquirido, o están en proceso de adquirir las placas BAJO y MEDIO. Para la placa ALTO se consideró el esperar a que se comercializara la versión 2.0 de la plataforma Versal. Sin embargo, tras preguntar a los contactos en AMD los socios han sido informados que dicha placa no se espera que se esté comercializando hasta finales de 2025. Es por ello, que la decisión ha sido seguir con la versión actual que está siendo adquirida ya por algunos socios.

4. Definiciones, Acrónimos y Abreviaciones

CÁPSULA	En CAPSUL-IA, solución IA preconfigurada constituida por MODELO, configuración de PLATAFORMA, gestión de los datos y soporte a la seguridad funcional y a la explicabilidad (para las soluciones relevantes).
DEMOSTRADOR	En CAPSUL-IA, prueba de concepto de una CÁPSULA con TRL 3-4
DLA	*Deep Learning Accelerator*
FuSa	*Functional Safety*
HMI	*Human Machine Interface*
IA	Inteligencia Artificial incluyendo Machine Learning (ML)
ML	*Machine Learning*
MODELO	En CAPSUL-IA, algoritmo basado en IA con hiperparámetros definidos
PLATAFORMA	En CAPSUL-IA, conjunto de software y hardware que permite la ejecución de los MODELOS
TRL	Technology Readiness Level

5. Referencias

AMD. (s.f.). *Functional Safety*. Obtenido de https://www.xilinx.com/products/technology/functional-safety.html

AMD. (s.f.). *Versal Serie AI Core*. Obtenido de https://www.amd.com/es/products/adaptive-socs-and-fpgas/versal/ai-core-series.html

Infineon. (s.f.). *32-bit AURIX TriCore Microcontroller*. Obtenido de https://www.infineon.com/cms/en/product/microcontroller/32-bit-tricore-microcontroller/

Infineon. (s.f.). *Microcontroller Safety Products PRO-SIL*. Obtenido de https://www.infineon.com/cms/en/product/microcontroller/microcontroller-safety-products-pro-sil-iso26262/

NVIDIA. (s.f.). *Jetson Ecosystem*. Obtenido de https://developer.nvidia.com/embedded/ecosystem

NVIDIA. (s.f.). *Jetson Safety Extension*. Obtenido de https://developer.nvidia.com/embedded/jetson/jetson-safety

NVIDIA. (s.f.). *NVIDIA Jetson Orin*. Obtenido de https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-orin/

XILINX. (s.f.). *Versal: AI Engine & Programming Environment*. Obtenido de https://www.xilinx.com/publications/events/developer-forum/2018-frankfurt/versal-ai-engine-and-programming-environment.pdf