

Hacia una Arquitectura Híbrida Extensible para Sistemas IA-Ready: Diseño y Resultados Preliminares

Facundo L. Chayle

Universidad Abierta Interamericana. Facultad de Tecnología Informática. Centro de Altos Estudios en Tecnología Informática. Buenos Aires. Argentina.

flchayle@outlook.com

***Abstract.** La incorporación creciente de inteligencia artificial (IA) en sistemas de software ha puesto en evidencia limitaciones estructurales en las arquitecturas tradicionales, particularmente en términos de acoplamiento y capacidad de evolución. A pesar del avance en modelos y técnicas de IA, la literatura reciente muestra una ausencia de arquitecturas genéricas diseñadas para integrar estas capacidades como componentes de primer nivel. Este trabajo presenta una investigación doctoral orientada al diseño de una arquitectura híbrida extensible para sistemas IA-ready, cuyo objetivo es desacoplar la lógica de negocio de los componentes de inteligencia artificial y facilitar su integración en entornos heterogéneos. Como resultado preliminar, se desarrolló un scaffolding funcional que implementa los principales componentes de la arquitectura propuesta, permitiendo validar de manera inicial su viabilidad conceptual y técnica.*

1. Introducción

La inteligencia artificial (IA) se ha convertido en un componente central de numerosos sistemas de software, impulsando avances en dominios como la atención al cliente, la salud, la educación, las finanzas y los servicios digitales. Modelos de aprendizaje automático, técnicas de procesamiento de lenguaje natural y sistemas de recomendación forman hoy parte del núcleo funcional de muchas aplicaciones (Ruiz & Velásquez, 2023; Nicolescu & Tudorache, 2022). Sin embargo, esta adopción acelerada ha puesto de manifiesto desafíos estructurales que trascienden el diseño de modelos y algoritmos, vinculándose directamente con la arquitectura del software que los soporta.

En la práctica, la integración de capacidades de IA suele realizarse mediante enfoques ad hoc o altamente acoplados a la lógica de negocio, lo que dificulta la evolución de los sistemas y la adopción de nuevos paradigmas de ejecución como el procesamiento distribuido o el edge computing (Veiga et al., 2023). Diversos enfoques arquitectónicos han sido propuestos para abordar parcialmente estos desafíos, incluyendo arquitecturas basadas en microservicios, pipelines de inferencia, plataformas de MLOps y despliegues serverless. Sin embargo, la literatura evidencia que muchas de estas soluciones continúan siendo específicas de dominio o carecen de mecanismos adaptativos que consideren el contexto operativo del sistema (Nguyen et al., 2023; Cadavid et al., 2022).

En este contexto, surge la necesidad de repensar el diseño arquitectónico de las aplicaciones que incorporan inteligencia artificial, proponiendo estructuras que permitan tratar a la IA como una capacidad transversal, desacoplada y extensible, capaz de operar

en distintos entornos de ejecución y adaptarse a restricciones como latencia, disponibilidad o privacidad (Huincal et al., 2018).

La investigación doctoral que se presenta en este trabajo se enmarca en este desafío y tiene como objetivo explorar y validar una arquitectura híbrida extensible orientada a sistemas IA-ready. El trabajo se apoya en un análisis sistemático del estado del arte, así como en el diseño e implementación de prototipos que permitan evaluar la viabilidad técnica de la propuesta.

2. Preguntas y objetivos de la Investigación

2.1. Preguntas de Investigación

La pregunta principal que guía esta investigación es la siguiente:

P1. ¿Cómo diseñar una arquitectura de software genérica y extensible que permita integrar capacidades de inteligencia artificial de forma desacoplada de la lógica de negocio?

A partir de esta pregunta central, se desprenden las siguientes preguntas de investigación específicas:

P2. ¿Cómo puede una arquitectura de software soportar la ejecución híbrida de componentes de inteligencia artificial en entornos cloud y edge, manteniendo la modularidad y la evolución independiente de sus componentes?

P3. ¿Cómo incorporar estrategias de orquestación dinámica que permitan seleccionar y coordinar componentes de inteligencia artificial en función del contexto operativo del sistema?

P4. ¿Cómo integrar principios de seguridad y validación de datos orientados a flujos de inteligencia artificial dentro del diseño arquitectónico de un sistema de software desacoplado?

En conjunto, estas preguntas buscan analizar no solo la estructura de la arquitectura propuesta, sino también los mecanismos que habilitan su adaptación a distintos contextos, así como su evolución frente a las demandas cambiantes de los sistemas.

Estos contextos incluyen sistemas distribuidos que incorporan capacidades de inteligencia artificial en dominios como plataformas de servicios digitales, sistemas de recomendación, aplicaciones educativas inteligentes y plataformas de procesamiento de transacciones, donde la integración de componentes de IA plantea desafíos arquitectónicos significativos.

2.2. Objetivos de la Investigación

En correspondencia con las preguntas planteadas, el objetivo general de esta investigación doctoral es:

Diseñar y evaluar una arquitectura híbrida extensible orientada a sistemas IA-ready que permita la integración efectiva de capacidades de inteligencia artificial en aplicaciones de software.

Para alcanzar este objetivo general, se establecen los siguientes objetivos específicos:

- Analizar sistemáticamente el estado del arte en arquitecturas de software con integración de inteligencia artificial.
- Definir una arquitectura de referencia que incorpore principios de modularidad y ejecución híbrida cloud–edge.
- Diseñar e implementar un prototipo funcional que materialice la arquitectura propuesta.
- Explorar mecanismos de orquestación dinámica que permitan adaptar la ejecución de componentes de IA.
- Identificar desafíos pendientes y líneas de trabajo futuras orientadas a la evaluación empírica y evolución de la arquitectura.

3. Estado del Arte y Conocimiento Actual

La integración de inteligencia artificial en sistemas de software ha sido abordada desde múltiples enfoques arquitectónicos en la literatura reciente, impulsada por la necesidad de escalar soluciones inteligentes en entornos complejos y heterogéneos (Ruiz & Velásquez, 2023; Nicolescu & Tudorache, 2022). Diversos trabajos han explorado estrategias para integrar modelos de aprendizaje automático y análisis de datos en aplicaciones distribuidas.

Uno de los enfoques predominantes se basa en arquitecturas de microservicios, donde las capacidades de IA se encapsulan en servicios independientes que interactúan con la lógica de negocio a través de interfaces bien definidas. Este paradigma favorece la modularidad y la escalabilidad y ha sido ampliamente adoptado junto con prácticas de MLOps para automatizar el ciclo de vida de los modelos (Nguyen et al., 2023; Cadavid et al., 2022). Sin embargo, muchas de estas soluciones se encuentran orientadas a entornos de nube y asumen flujos de inferencia relativamente estáticos.

Otro conjunto de propuestas se basa en arquitecturas serverless y pipelines de inferencia para reducir la complejidad operativa. Si bien estos enfoques resultan efectivos en escenarios centrados en eventos o cargas variables, presentan limitaciones relacionadas con latencia, gestión del estado e integración con la lógica del dominio (Veiga et al., 2023).

En los últimos años, el edge computing ha cobrado relevancia como alternativa para reducir latencias y procesar datos cerca de su origen. Diversos trabajos han explorado la ejecución de modelos de IA en dispositivos periféricos, especialmente en dominios como IoT y sistemas ciberfísicos (Huincalef et al., 2018; Veiga et al., 2023). No obstante, estas propuestas suelen presentarse como extensiones ad hoc de arquitecturas existentes, sin definir mecanismos estandarizados de coordinación entre el procesamiento en el borde y la nube.

Asimismo, existen contribuciones en gobernanza de modelos, interpretabilidad y seguridad de sistemas basados en IA, aunque suelen abordarse como capas externas a la arquitectura principal, dificultando su adopción sistemática (Nazir et al., 2024).

El análisis sistemático realizado evidencia que la mayoría de las soluciones existentes tienden a ser específicas de dominio, presentan altos niveles de acoplamiento entre la IA

y la lógica de negocio y carecen de una visión holística que contemple ejecución híbrida y orquestación dinámica (Chayle et al., 2024).

En este escenario, se identifica un vacío en torno a arquitecturas diseñadas explícitamente para sistemas IA-ready que permitan evolucionar capacidades inteligentes sin comprometer la mantenibilidad ni la independencia tecnológica. Este vacío constituye el punto de partida de la investigación doctoral presentada en este trabajo.

4. Metodología de Investigación

La investigación presentada en este trabajo adopta un enfoque exploratorio y de diseño, orientado al desarrollo y evaluación progresiva de una arquitectura de software para sistemas con inteligencia artificial. Este enfoque combina técnicas de investigación en ingeniería de software con estrategias de diseño arquitectónico.

El trabajo se estructura en tres etapas principales. En primer lugar, se llevó a cabo un Mapeo Sistemático de la Literatura con el objetivo de identificar enfoques existentes para la integración de IA en arquitecturas de software, detectar tendencias predominantes y reconocer brechas de investigación relevantes. Esta etapa permitió delimitar el problema, definir las preguntas de investigación y establecer los requisitos arquitectónicos iniciales.

En segundo lugar, se desarrolló una propuesta arquitectónica conceptual basada en los hallazgos del análisis bibliográfico y en principios de modularidad, separación de responsabilidades y ejecución híbrida. Durante esta etapa se definieron los componentes principales del sistema, sus responsabilidades y los mecanismos de interacción necesarios para coordinar el procesamiento inteligente dentro de la arquitectura.

Finalmente, se implementó un scaffolding funcional que materializa los conceptos arquitectónicos propuestos y permite validar la coherencia del diseño, analizar la interacción entre componentes y evaluar la factibilidad técnica de la arquitectura.

La evaluación de la arquitectura se concibe como un proceso progresivo que combina validación conceptual, análisis cualitativo y experimentación técnica, permitiendo refinar la propuesta y sentar las bases para futuras evaluaciones empíricas.

Este enfoque metodológico se encuentra directamente vinculado con las preguntas de investigación planteadas. El mapeo sistemático de la literatura permitió identificar los desafíos arquitectónicos asociados a la integración de inteligencia artificial (P1), mientras que el diseño de la arquitectura explora mecanismos de ejecución híbrida y orquestación dinámica (P2 y P3). Finalmente, la implementación del scaffolding funcional proporciona una validación inicial de la viabilidad técnica de la arquitectura propuesta y permite explorar mecanismos de seguridad y validación de datos en los flujos de inteligencia artificial (P4).

5. Propuesta Arquitectónica

La propuesta desarrollada en esta investigación consiste en el diseño de una arquitectura híbrida extensible orientada a sistemas IA-ready, concebida para integrar capacidades de inteligencia artificial como un componente transversal del sistema, desacoplado de la lógica de negocio y adaptable a distintos contextos de ejecución. El objetivo central es

permitir la incorporación progresiva de funcionalidades inteligentes sin comprometer la mantenibilidad, la escalabilidad ni la independencia tecnológica de las aplicaciones.

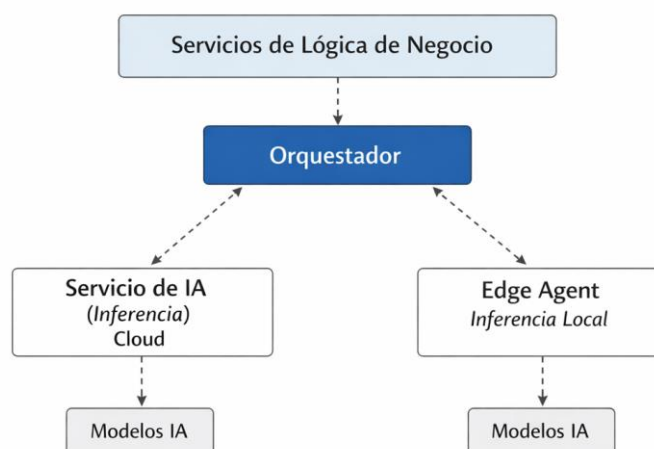


Figura 1. Vista conceptual de la arquitectura híbrida extensible para sistemas IA-ready

La arquitectura se apoya en tres principios fundamentales: separación de responsabilidades, extensibilidad arquitectónica y ejecución híbrida. A partir de estos principios, se define un conjunto de componentes delimitados, cuyas interacciones permiten coordinar el procesamiento inteligente de manera flexible y contextual.

El primer componente es un servicio de orquestación encargado de centralizar la lógica de negocio del sistema y coordinar las interacciones con los servicios de inteligencia artificial. Este servicio actúa como punto de entrada para las solicitudes del dominio e incorpora mecanismos de decisión que determinan cuándo y cómo delegar el procesamiento a los componentes inteligentes, permitiendo mantener desacoplada la lógica de negocio de los detalles de implementación de la IA.

El segundo componente corresponde a un servicio especializado de inteligencia artificial, implementado como un microservicio independiente que encapsula la lógica de inferencia junto con tareas de preprocesamiento, validación de entradas y postprocesamiento de resultados. Este servicio es agnóstico al dominio de aplicación, lo que facilita su reutilización y la sustitución o actualización de modelos sin impacto sobre el resto del sistema.

Como extensión opcional, la arquitectura contempla la incorporación de un agente de ejecución en el borde (edge agent), diseñado para operar en entornos con restricciones de latencia, conectividad o privacidad. Este agente permite ejecutar inferencias localmente cuando el contexto operativo lo requiere, complementando la ejecución en la nube y habilitando escenarios híbridos.

La comunicación entre componentes se diseña de forma flexible, permitiendo tanto interacciones síncronas mediante interfaces REST como mecanismos asíncronos basados en eventos o colas de mensajes. Esta decisión facilita la adaptación del sistema a distintos estilos de interacción y habilita su evolución hacia arquitecturas reactivas.

Un aspecto central de la propuesta es la incorporación de mecanismos de orquestación dinámica que permiten seleccionar el motor de inferencia más adecuado en función del contexto operativo. Factores como latencia, disponibilidad del entorno edge o carga del

sistema pueden ser considerados para decidir si una solicitud debe procesarse localmente o delegarse a servicios remotos, optimizando el uso de recursos sin aumentar la complejidad de la lógica de negocio.

Desde la perspectiva de seguridad y gobernanza, la arquitectura integra principios de Zero Trust AI mediante validaciones explícitas sobre los datos de entrada y salida de los componentes inteligentes, permitiendo controlar los flujos de información y reducir riesgos asociados a datos inconsistentes o maliciosos.

En conjunto, la arquitectura propuesta constituye un marco flexible para la integración de inteligencia artificial en sistemas de software, habilitando la evolución progresiva de las capacidades inteligentes y proporcionando una base sólida para futuras extensiones.

6. Resultados Obtenidos hasta la Fecha

Como parte del desarrollo de la investigación doctoral, se han alcanzado una serie de resultados parciales que permiten validar la viabilidad conceptual y técnica de la arquitectura propuesta. Estos avances no constituyen aún una evaluación empírica exhaustiva, pero aportan evidencia sobre la factibilidad del enfoque.

El principal resultado obtenido hasta el momento es la implementación de un *scaffolding* funcional que materializa los componentes centrales de la arquitectura híbrida extensible. Dicho *scaffolding* se compone de un servicio de orquestación y un servicio de inteligencia artificial claramente desacoplados, desarrollados utilizando tecnologías y lenguajes distintos. Esta decisión permitió evaluar, en una etapa temprana, el grado de independencia tecnológica alcanzado por la arquitectura y su capacidad para operar en entornos heterogéneos.

El servicio de inteligencia artificial implementado encapsula la lógica de inferencia junto con tareas de validación, preprocesamiento y postprocesamiento de datos, funcionando como un componente reutilizable y ajeno a la lógica específica del dominio. Por su parte, el servicio de orquestación concentra la lógica de negocio y actúa como mediador entre las solicitudes del sistema y los componentes inteligentes, incorporando mecanismos iniciales de decisión que permiten delegar el procesamiento según el contexto operativo.

Adicionalmente, se desarrolló un agente de ejecución en el borde como componente opcional, lo que permitió explorar escenarios híbridos en los que la inferencia puede realizarse localmente o de forma remota. Si bien este agente se encuentra en una etapa inicial, su integración temprana facilitó la validación del enfoque híbrido y evidenció la capacidad de la arquitectura para incorporar nuevos componentes sin modificaciones sustanciales en su estructura base.

A partir de la implementación del *scaffolding*, se obtuvieron una serie de aprendizajes preliminares relevantes. En primer lugar, se confirmó que la separación entre lógica de negocio e inteligencia artificial simplifica la evolución independiente de ambos componentes. En segundo lugar, se observó que la incorporación de mecanismos de orquestación dinámica resulta clave para evitar la propagación de decisiones contextuales hacia el resto del sistema, manteniendo una arquitectura más limpia y mantenible.

Estos resultados iniciales sugieren que la arquitectura propuesta es técnicamente viable y proporciona una base para explorar escenarios más complejos, aunque persisten desafíos relacionados con la definición de criterios de orquestación y métricas claras de

evaluación. Estos aspectos constituyen líneas de trabajo activas dentro de la investigación y serán abordados en las siguientes etapas del proyecto.

7. Aportes Diferenciales de la Propuesta

La principal contribución de la arquitectura propuesta no reside en la introducción de nuevos modelos o técnicas de inteligencia artificial, sino en el reposicionamiento de la IA como una capacidad arquitectónica de primer nivel dentro de los sistemas de software. En contraste con enfoques tradicionales que integran la inteligencia artificial de manera ad hoc o dependiente del dominio, la propuesta plantea una estructura que permite tratarla como un componente transversal, desacoplado y gobernable.

Un primer aporte diferencial se vincula con la explicitación del desacoplamiento entre la lógica de negocio y los componentes de inteligencia artificial como decisión arquitectónica central. Este enfoque facilita la evolución independiente de ambos aspectos y habilita escenarios de reutilización difíciles de alcanzar en arquitecturas fuertemente acopladas.

En segundo lugar, la investigación aporta una visión integrada de ejecución híbrida cloud-edge, concebida desde el diseño arquitectónico y no como una extensión posterior. Esta perspectiva permite analizar de manera sistemática las implicancias del contexto operativo en la ejecución de componentes de IA, sentando las bases para arquitecturas más adaptativas y conscientes de restricciones reales como latencia, conectividad o privacidad.

Finalmente, la propuesta introduce la orquestación dinámica como un problema arquitectónico explícito, en lugar de tratarla como una preocupación de implementación. Este cambio de enfoque habilita nuevas líneas de investigación orientadas a la toma de decisiones contextuales, la gobernanza de flujos de IA y la evolución autónoma de los sistemas inteligentes.

En conjunto, estos aportes posicionan a la propuesta como un marco conceptual reutilizable para el diseño de sistemas *IA-ready*, y establecen una base sólida para la exploración futura de mecanismos de evaluación, control y adaptación en arquitecturas de software con inteligencia artificial integrada.

8. Conclusiones y Trabajos Futuros

La creciente incorporación de inteligencia artificial en sistemas de software ha evidenciado limitaciones en arquitecturas tradicionales caracterizadas por alto acoplamiento y escasa adaptabilidad. En este contexto, la investigación presentada aborda un problema arquitectónico orientado a definir estructuras que permitan integrar la IA como una capacidad de primer nivel dentro de las aplicaciones.

A partir del análisis sistemático del estado del arte, se identificó un vacío en torno a arquitecturas genéricas y extensibles diseñadas explícitamente para sistemas *IA-ready*. En respuesta a esta brecha, el trabajo propone una arquitectura híbrida extensible que prioriza el desacoplamiento entre la lógica de negocio y los componentes de inteligencia artificial, incorporando principios de ejecución híbrida cloud-edge, orquestación dinámica y gobernanza de los flujos de IA.

Los resultados obtenidos hasta el momento, materializados en un scaffolding funcional y en la implementación de componentes arquitectónicos clave, permiten validar la viabilidad conceptual y técnica del enfoque propuesto. Si bien estos resultados no constituyen aún una evaluación empírica exhaustiva, aportan evidencia sobre la factibilidad de la arquitectura y su potencial para abordar los desafíos identificados.

Como líneas futuras de investigación, se prevé avanzar en la evaluación empírica sistemática de la arquitectura mediante experimentos controlados que permitan analizar su impacto en métricas como latencia, flexibilidad, escalabilidad y sobrecarga asociada a los mecanismos de orquestación dinámica. Asimismo, se plantea profundizar en la definición de criterios de decisión para la selección dinámica de componentes de IA, incorporando métricas de desempeño histórico y políticas configurables.

Otra línea de trabajo consiste en validar la arquitectura en escenarios reales o representativos de distintos dominios para analizar su comportamiento bajo cargas variables y su facilidad de adopción. Desde la perspectiva de seguridad y gobernanza, se prevé avanzar en la formalización y evaluación de los mecanismos de Zero Trust AI, incluyendo estrategias avanzadas de validación de datos y detección de comportamientos anómalos.

Finalmente, se plantea como línea exploratoria la incorporación de mecanismos de adaptación y aprendizaje autónomo que permitan a la arquitectura evolucionar en función del desempeño observado y de los cambios en el entorno operativo. En conjunto, estas líneas delimitan la continuidad de la investigación doctoral y refuerzan el carácter evolutivo de la arquitectura propuesta.

En particular, la evaluación considerará atributos de calidad relevantes para arquitecturas de software, tales como escalabilidad, extensibilidad, mantenibilidad y latencia, alineados con modelos de calidad como el estándar ISO/IEC 25010.

10. Referencias

- Cadavid, H., Andreu, G., & González, M. (2022). *Arquitecturas de software para la integración de inteligencia artificial en sistemas distribuidos*. Journal of Software Engineering Research and Development, 10(1), 1–18.
- Chayle, F. (2024). *Arquitecturas para Aplicaciones con Inteligencia Artificial: Un Mapeo Sistemático de la Literatura*. Trabajo académico, Universidad Abierta Interamericana.
- Huincalef, J., Maldonado, J. C., & Muñoz, R. (2018). *Scalable architectures for intelligent systems in distributed environments*. IEEE Latin America Transactions, 16(3), 850–858.
- Nguyen, T., Tran, D., & Le, H. (2023). *Architectural approaches for integrating artificial intelligence into software systems*. Journal of Systems and Software, 198, 111612.
- Nicolescu, R., & Tudorache, M. (2022). *Human-centered artificial intelligence: A framework for intelligent systems*. Springer.
- Nazir, S., Khan, M. A., & Hussain, S. (2024). *Security and governance challenges in intelligent software systems*. Future Generation Computer Systems, 150, 45–58.

- Ruiz, M., & Velásquez, J. D. (2023). *Artificial intelligence as a disruptive technology: Implications for software systems*. IEEE Software, 40(2), 89–96.
- Veiga, M., López, M., & García, P. (2023). *Challenges of integrating artificial intelligence in IoT and edge computing architectures*. Sensors, 23(4), 2105.