Informe de progreso

MLEDGE - Aprendizaje automático en la nube y en el borde

(Aprendizaje automático en la nube y en el borde)

Diciembre de 2024

Información sobre el entregable

Nombre del documento:

Informe de progreso

Versión actual: 1.0

Proyecto: MLEDGE: aprendizaje automático en la nube y en el borde

Paquete de trabajo: P0 - Gestión de proyectos

Tarea: A0.1: Gestión de proyectos

Entregable: E0.2 – M24 – Informe de progreso

Autores: Javad Dogani (IMDEA)

Revisores: Nikolaos Laoutaris (IMDEA)

Historial de versiones

Versión	Fecha	Resumen de modificaciones
Versión 1.0	31-12-2024	Versión inicial del documento

Índice

1. Introducción

Este informe de situación proporciona una actualización completa de los distintos paquetes de trabajo y actividades realizadas en el proyecto MLEDGE durante los últimos 24 meses. Este informe, que funciona como una herramienta operativa fundamental, facilita la coordinación entre los socios del proyecto MLEDGE y garantiza la coordinación de esfuerzos en todas las actividades del proyecto. Destaca las contribuciones al proyecto a través de reuniones y eventos de colaboración a los que han asistido los miembros del proyecto. Este informe no solo hace un seguimiento de los avances, sino que también sienta las bases para el avance continuo del proyecto MLEDGE, garantizando que todos los socios estén bien informados y alineados estratégicamente para las actividades futuras. Este informe está dirigido principalmente a los puntos de contacto administrativos del proyecto UNIOCO, proporcionándoles información detallada y actualizaciones.

El documento está organizado en varias secciones clave:

- **Sección 2** : En esta sección se ofrece una descripción detallada de las actividades realizadas en cada paquete de trabajo, desde el mes 18, siguiente a la presentación del informe anterior, hasta el mes 24.
- **Sección 3** : En esta sección se describen los próximos pasos y estrategias planificadas para la siguiente fase del proyecto, con el objetivo de aprovechar el impulso alcanzado y abordar cualquier desafío emergente.

2. Informe de actividad

2.1. Objetivo del proyecto

El objetivo del proyecto MLEDGE es impulsar la implementación de FL como una capa intersectorial independiente pero optimizada sobre CloudEdge , utilizando aplicaciones y datos del mundo real para demostrar que esta sinergia puede generar grandes beneficios para todos. Esto permitirá un ecosistema próspero de servicios de borde de FL seguros y eficientes capaces de facilitar el uso de datos personales y B2B confidenciales para entrenar modelos de ML de consumidores, al tiempo que se protege la privacidad de los datos y sus propietarios.

Para allanar el camino para la adopción de FL en el borde para una cantidad cada vez mayor de aplicaciones que utilizan modelos de ML, MLEDGE está trabajando en el desarrollo de técnicas, bibliotecas y componentes que permitan una implementación más rápida de estos servicios. La Figura 1 resume la arquitectura de MLEDGE y los componentes básicos del proyecto.

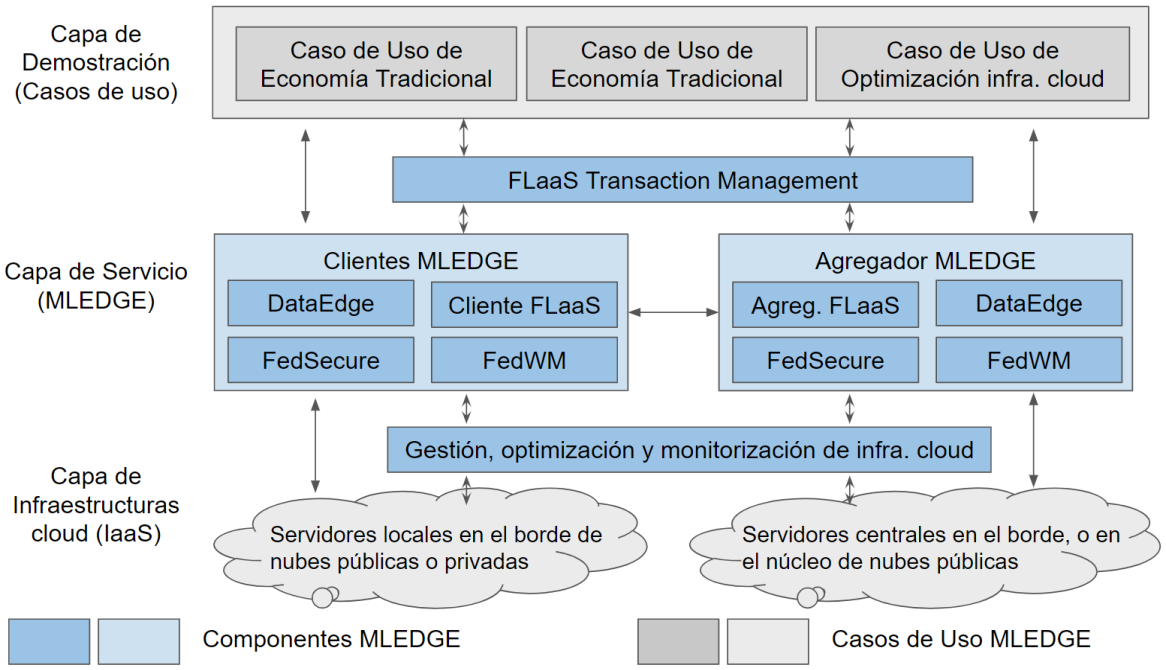

Figura 1. Diagrama de bloques de MLEDGE

2.2. Estructura de ejecución

. La Figura 2 ofrece un esquema de los paquetes de trabajo del proyecto y las relaciones entre ellos. El proyecto está estructurado en 7 paquetes de actividades (P0-P6). P0 cubre la gestión del proyecto y P1 tiene como objetivo definir los requisitos de los casos de uso y diseñar la arquitectura del proyecto. Además, se han planificado 4 paquetes de trabajo técnico (P2-P5) y un paquete final (P6) tiene como objetivo demostrar pruebas de concepto

y difundir, explotar y comunicar los resultados del proyecto.

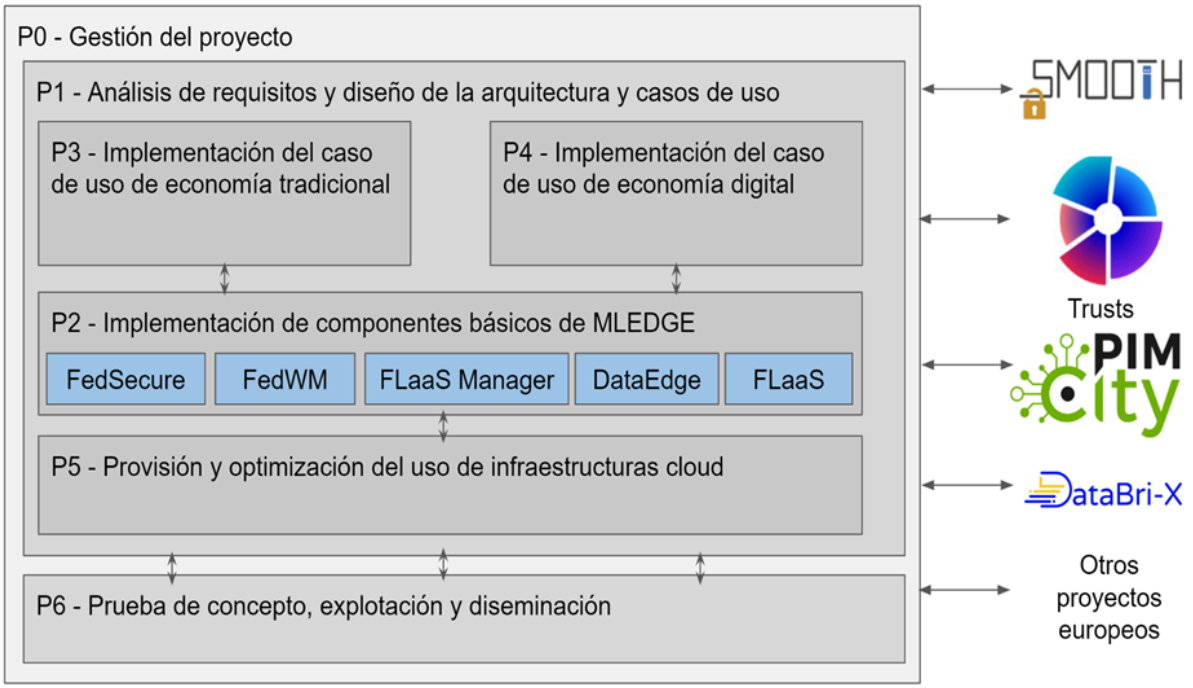

Figura 2. Paquetes de trabajo de MLEDGE

A continuación se ofrece un resumen del progreso y las actividades por paquete de trabajo.

2.3. Actividad por paquete de trabajo

2.3.1. P0 - Gestión de proyectos

Como parte de este paquete de trabajo, se han llevado a cabo las siguientes actividades:

- Configurar y mantener la infraestructura tecnológica, incluido el sitio web y las listas de correo.
- Organizar reuniones periódicas de seguimiento.
- Gestionar la coordinación estratégica a través de la revisión periódica de la visión del proyecto, análisis y solución de problemas en la implementación del plan de trabajo.
- Supervisar el proceso de licitación de los componentes técnicos del proyecto.
- Preparar una presentación oficial del proyecto y presentarla a los socios adjudicatarios de los diferentes componentes técnicos.
- Coordinar los diferentes componentes técnicos
- Controlar la calidad de la ejecución del proyecto y de los entregables.
- Revisar y perfeccionar las prácticas de gestión de proyectos para incorporar nuevos conocimientos y comentarios de la ejecución del proyecto.
- Supervisión de la colaboración entre investigadores y socios para la retroalimentación sobre los requisitos de datos
- Coordinadora de la participación de investigadores y socios en la sesión de encuentro de UNICO I+D
- Reunión estratégica para implementar los resultados de la investigación en las plataformas asociadas
- Participación en reuniones de revisión de los entregables del proyecto planificados para el mes 24.

- Supervisión de equipos de investigación
- Preparar este informe de actividades

2.3.2. P1 - Análisis de requerimientos y diseño de arquitectura y casos de uso

Como parte de este paquete de trabajo, se han llevado a cabo las siguientes actividades:

- Supervisar desde el punto de vista técnico y administrativo el proceso de subcontratación, el cual se realizó en dos partes debido a que la primera licitación fue declarada desierta.
- Realizar reuniones de revisión periódicas para evaluar el progreso y la integración de los diseños arquitectónicos, facilitando los ajustes necesarios para satisfacer las cambiantes demandas del proyecto.
- Coordinación inicial de los diferentes socios adjudicatarios del proyecto
- Informe a la dirección del proyecto

2.3.3. P2 - Implementación de componentes básicos de MLEDGE

Como parte de este paquete de trabajo, se han llevado a cabo las siguientes actividades:

- Actividades de investigación propias en los campos correspondientes
- Desarrollos vinculados a los diferentes componentes
- Documentación de la actividad investigadora
- Elaboración de artículos científicos – ver detalles en el sitio web del proyecto, particularmente en este enlace .
- Preparación de material para charlas y difusión
- Selección de trabajos científicos a incorporar en los paquetes de trabajo técnico P3-P5 (ver detalles en los entregables correspondientes)

2.3.4. P3 - Implementación del caso de uso de la economía tradicional

Tras su adjudicación el 3/2/2024 al consorcio formado por InmaRepro y Acuratio , y su posterior formalización el 8/04/2024, la actividad del paquete de trabajo se ha centrado en la definición de requisitos y diseño de la arquitectura de los casos de uso. Para ello se han llevado a cabo las siguientes actividades:

- Establecer una rutina para monitorear el progreso respecto del plan del paquete de trabajo, incluido el uso de indicadores de desempeño e informes de progreso para identificar cualquier desviación de los objetivos planificados.
- Participación en reuniones de seguimiento con el equipo del proyecto IMDEA Networks
- Verificar los aspectos técnicos en reuniones de seguimiento para garantizar que se cumplan todos los requisitos técnicos en los diferentes aspectos del proyecto.

- Organizar reuniones con investigadores para recopilar retroalimentación sobre la implementación de la investigación dentro de sus plataformas, utilizando esta retroalimentación para refinar y optimizar futuras actividades del proyecto.
- Preparación del entregable E3.2 – Implementación preliminar de los componentes del caso de uso de economía tradicional.
- Realizar comprobaciones exhaustivas para garantizar la coherencia técnica y la integridad de la implementación preliminar de los componentes para el caso de uso, como se describe en el entregable E3.2.
- Revisión de los comentarios sobre el entregable realizados por el equipo del proyecto

2.3.5. P4 - Implementación del caso de uso de la economía digital

Tras su adjudicación el 3/2/2024 al consorcio formado por Orange España y Acuratio , y su posterior formalización el 8/04/2024, la actividad del paquete de trabajo se ha centrado en la definición de requisitos y diseño de la arquitectura de los casos de uso. Para ello se han llevado a cabo las siguientes actividades:

- Establecer una rutina para monitorear el progreso respecto del plan del paquete de trabajo, incluido el uso de indicadores de desempeño e informes de progreso para identificar cualquier desviación de los objetivos planificados.
- Participación en reuniones de seguimiento con el equipo del proyecto IMDEA Networks
- Verificar los aspectos técnicos en reuniones de seguimiento para garantizar que se cumplan todos los requisitos técnicos en los diferentes aspectos del proyecto.
- Organizar reuniones con investigadores para recopilar retroalimentación sobre el suministro de datos para la investigación y utilizar esta retroalimentación para refinar y optimizar futuras actividades del proyecto.
- Preparación del entregable E4.2 – Implementación preliminar de los componentes del caso de uso de la economía digital.
- Realizar comprobaciones exhaustivas para garantizar la coherencia técnica y la integridad de la implementación preliminar de los componentes para el caso de uso, como se describe en el entregable E4.2.
- Revisión de los comentarios sobre el entregable realizados por el equipo del proyecto

2.3.6. P5 - Provisión y optimización de infraestructuras en la nube

Tras su adjudicación el 3/2/2024 al consorcio formado por Acuratio SL, y su posterior formalización el 8/04/2024, la actividad del paquete de trabajo se ha centrado en la definición de requisitos y diseño de la arquitectura de los casos de uso. Para ello se han llevado a cabo las siguientes actividades:

- Establecer una rutina para monitorear el progreso respecto del plan del paquete de trabajo, incluido el uso de indicadores de desempeño e informes de progreso para identificar cualquier desviación de los objetivos planificados.
- Participación en reuniones de seguimiento con el equipo del proyecto IMDEA Networks
- Verificar los aspectos técnicos en reuniones de seguimiento para garantizar que se cumplan todos los requisitos técnicos en los diferentes aspectos del proyecto.
- Organizar reuniones con investigadores para recopilar retroalimentación sobre la implementación de la investigación dentro de sus plataformas, utilizando esta retroalimentación para refinar y optimizar futuras actividades del proyecto.
- Preparación del entregable E5.2 – Implementación preliminar de componentes de infraestructura en la nube.
- Realizar comprobaciones exhaustivas para garantizar la coherencia técnica y la integridad de la implementación preliminar de los componentes para el caso de uso, como se describe en el entregable E5.2.
- Revisión de los comentarios sobre el entregable realizados por el equipo del proyecto

2.3.7. P6 - Prueba de concepto, explotación y difusión

Como parte de este bloque de trabajo se han realizado las siguientes actividades:

- Creación de un canal de noticias en el sitio web del proyecto (https://mledge.networks.imdea.org/en/news/)
- Publicación periódica de noticias en la página de LinkedIn del proyecto (https://www.linkedin.com/company/mledge-project/) y en el sitio web del proyecto.
- Elaboración de material y difusión del trabajo en página web y redes sociales
- Presentaciones y charlas para difundir el trabajo del proyecto.

A continuación se presenta un resumen de las publicaciones científicas resultantes de los desarrollos del proyecto hasta la fecha:

- Santiago Andres Azcoitia, Costas Iordanou, and Nikolaos Laoutaris. (2023) Understanding the Price of Data in Commercial Data Marketplaces . IEEE International Conference on Data Engineering ICDE. Los Angeles, California, USA. April 2023.

- Tianyue Chu, Alvaro Garcia-Recuero, Costas Iordanou, Georgios Smaragdakis, and Nikolaos Laoutaris Securing Federated Sensitive Topic Classification against Poisoning Attacks . Network and Distributed System Security (NDSS) Symposium. 2023.

- Devriş Işler, Elisa Cabana, Álvaro Garcia-Recuero, Gieorgia Koutrika, Nikolaos Laoutaris, FreqyWM: Frequency Watermarking for the New Data Economy, accepted for publication in IEEE International Conference on Data Engineering ICDE'24.

- Tianyue,Chu, Nikolaos Laoutaris. "FedQV: Leveraging Quadratic Voting in Federated Learning." *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2024.

- Tianyue Chu, Mengwei Yang, Nikolaos Laoutaris, and Athina Markopoulou. "Information-Theoretical Bounds on Privacy Leakage in Pruned Federated Learning." In *ISIT 2024 Workshop on Information-Theoretic Methods for Trustworthy Machine Learning*. 2024.

Además se han realizado las siguientes publicaciones y charlas:

- Santiago Andres Azcoitia. Presentación del paper Understanding the Price of Data in Commercial Data Marketplaces . at the IEEE International Conference on Data Engineering ICDE. Los Angeles, California, USA. April 2023. (Video link).

- Tianyue Chu. Presentación del paper Securing Federated Sensitive Topic Classification against Poisoning Attacks . Network and Distributed System Security (NDSS) Symposium. 2023.

- Santiago Andres Azkoitia and Alba Ribera Martinez Data Marketplaces and the Data Governance Act: A Business Model Perspective . September 2023. Kluwer Competition Law Blog.

- Santiago Andrés Azcoitia and Alba Ribera Martínez Data Marketplaces and the Data Governance Act: A Business Model Perspective . November 2023. PLAMADISO (Platforms, Markets, and the Digital Society) talk at the Weizenbaum Institute for the Networked Society.

- Santiago Andrés Azcotiia, Charla: Towards a Human-centric data economy. Higher Technical School of Telecommunications Engineers, Polytechnic University of Madrid.

- Tianyue Chu. Presentación del paper FedQV: Leveraging Quadratic Voting in Federated Learning . 2024 ACM SIGMETRICS/IFIP PERFORMANCE Joint the ACM on Measurement and Analysis of Computing Systems, 2024.

- Tianyue Chu. Presentación del paper, Information-Theoretical Bounds on Privacy Leakage in Pruned Federated Learning, ISIT 2024 Workshop on Information-Theoretic Methods for Trustworthy Machine Learning, 2024.

- Tianyue Chu. Seminario: FedQV: Leveraging Quadratic Voting in Federated Learning. IMDEA Network Institute, Madrid.

- Javad Dogani. Presentación de Actividades de Investigación del Proyecto MLEDGE, sesión Meet-Up UNICO I+D, septiembre 2024.

- Naicheng Li. Seminario: Privacy-preserving Chunk Scheduling in a BitTorrent Implementation of Federated Learning. IMDEA Network Institute, Madrid.

3. Próximos pasos

En la primera mitad de 2025, está previsto que comiencen las actividades de implementación del diseño con las empresas a las que se les han asignado los paquetes de trabajo 3, 4 y 5. Al mismo tiempo, las actividades de investigación seguirán avanzando de acuerdo con el plan de trabajo, y los hallazgos se integrarán progresivamente en los casos de uso durante las colaboraciones con estas empresas. Además, estamos en proceso de completar un documento de investigación que analiza los datos de los casos de uso y explora las colaboraciones entre los investigadores y otros socios del proyecto. El proyecto tiene como objetivo finalizar los componentes técnicos para el final del período del proyecto, lo que permitirá una demostración parcial de los requisitos y elementos de diseño descritos en los entregables 3.2, 4.2 y 5.2.

Respecto a la actividad de publicación, se desarrollarán los siguientes artículos científicos:

- Un modelo de aprendizaje federado para la fijación de precios de datos en los mercados de datos comerciales.
- Programación de fragmentos que preserva la privacidad en una implementación de aprendizaje federado de BitTorrent
- Cómo equilibrar la precisión y la eficiencia en la optimización de algoritmos de programación de fragmentos para FLTorrent
- El potencial de la asignación inteligente de estacionamientos en las calles para reducir la congestión y el tiempo de búsqueda de estacionamiento
- RAG to Riches: una generación aumentada de recuperación eficiente y que preserva la privacidad para modelos de lenguaje grandes

Además, se seguirá trabajando en la creación de contenidos para la página web y las redes sociales sobre el proyecto. También se buscarán oportunidades de difusión este año y, de cara a 2025, de demostración de prototipos intermedios de los componentes técnicos, que deberían estar listos a finales de año.