

Sistema multiagente para la generación de planes personalizados de mitigación de la huella de carbono empresarial

Índice

1. Resumen Ejecutivo	5
2. Objeto y Alcance de la Investigación	7
2.1. Objeto.....	7
2.2. Autores	7
2.3. Alcance	8
3. La huella de carbono empresarial y el reto de la descarbonización	9
3.1. Estado del arte	9
4. Inteligencia artificial aplicada a la descarbonización.....	13
4.1. Sistemas multiagente	13
4.2. La coordinación entre agentes	14
4.3. RAG jerárquico sin vectores.....	15
4.4. Observabilidad de agentes.....	15
4.5. Por qué la IA aporta valor diferencial en la reducción de huella	16
5. Reto técnico abordado	18
5.1. Heterogeneidad del catálogo de medidas	18
5.2. Datos que faltan	18
5.3. El valor de la personalización.....	19
6. Diseño de la herramienta: AgentCO2	21
6.1. Concepto y propósito	21
6.2. Arquitectura general	21
6.3. El pipeline de tres agentes	22
6.4. Pre-filtrado determinista en Python	25
6.5. RAG jerárquico de medidas	25
6.6. Backend FastAPI + WebSocket	26
6.7. Frontend Nuxt 4.....	27
6.8. Despliegue	28
7. Catálogo de medidas de descarbonización	29
7.1. Estructura de los ficheros de medida	29
7.2. Medidas de Alcance 1	30
7.3. Medidas de Alcance 2	31
7.4. Medidas de Alcance 3	32
8. Diseño y desarrollo de los experimentos	33

8.1. Empresa piloto y datos de entrada	33
8.2. Resultados experimentales	33
8.3. Análisis cualitativo de las medidas propuestas	38
8.4. Conclusiones experimentales.....	38
9. Integración con SCOPE: visión a medio plazo	40
10. Conclusiones y líneas futuras	43
Anexo A. Bibliografía	44
Anexo B. Glosario de términos	46
Anexo C. Imágenes de la herramienta	47

AUTORES

Este informe es el resultado de la colaboración entre el Instituto Tecnológico de Aragón (ITA) y la Fundación ECODES, dos entidades aragonesas que han trabajado con un objetivo compartido: poner la tecnología al servicio de la descarbonización. De esta colaboración se ha materializado el desarrollo del caso real **AGENTCO2, un sistema multiagente para generar propuestas personalizadas de mitigación**. El proyecto ha sido ganador del premio AMETIC IA I+D 2026. El equipo firmante combina perfiles de investigación aplicada y experiencia sectorial, y es responsable conjunto de la concepción, redacción y validación de los análisis, oportunidades y recomendaciones que se presentan a continuación.



**CIRO
ARTIGOT**
ECODES



**ADÁN PIÑERO
ALQUEGUI**
ITA



**FRANCISCO
ROJAS LUNA**
ITA



**IRENE LÓPEZ
BOSQUE**
ITA

SOBRE ESTE DOCUMENTO

Editado por: Fundación ECODES

Diseño Gráfico: BAMBAM Studio

Fecha: Junio 2026

1. Resumen Ejecutivo

El presente informe documenta el diseño, la implementación y los resultados experimentales de **AgentCO2**, un sistema multiagente de inteligencia artificial desarrollado entre octubre de 2025 y mayo de 2026 por el Departamento de Sistemas Cognitivos del Instituto Tecnológico de Aragón en colaboración con ECODES. AgentCO2 toma como entrada el JSON de **huella de carbono** que genera el programa ScopeCO2 en adelante también SCOPE, el calculador GHG de ECODES, y produce, sin intervención humana, un **plan de descarbonización personalizado para la empresa**: una lista priorizada de **medidas** con su reducción de CO₂ estimada en tCO₂e/año, su **inversión** orientativa, su **ahorro** económico anual y su *payback*.

El flujo del sistema sigue un **patrón limpio de tres roles** que reproduce la **lógica** que aplicaría un consultor **humano**: un primer agente discrimina qué medidas tiene sentido considerar para esta empresa concreta a partir del inventario; uno o varios agentes calculan el impacto de cada medida sobre los datos reales de la empresa; y cuando faltan datos para realizar el cálculo, el sistema deriva a un agente recopilador que abre una conversación focalizada con el usuario para completarlos. La salida es una propuesta de mitigación con cifras concretas y detalle del cálculo trazable.

La motivación de este trabajo parte de una constatación que **ECODES** había observado durante más de una década de acompañamiento a empresas y que también está documentada en la literatura sectorial: **el hecho de disponer de la huella calculada no es suficiente para que una organización emprenda acciones de mitigación**. Existe una brecha de acción entre el inventario y el plan, una brecha cuantificada por McKinsey (2023) y por Bain & Co. (2024), que las consultoras tradicionales solo pueden cerrar de forma artesanal y, por tanto, costosa para cada empresa concreta. AgentCO2 explora si esta brecha puede automatizarse mediante un sistema multiagente que combine código determinista (para reglas estrictas), modelos de lenguaje (para interpretación de datos heterogéneos) y una base de conocimiento estructurada en JSON (las 23 fichas de medida del catálogo ECODES).

El sistema se articula en torno a tres agentes LLM coordinados con el patrón AG2/AutoGen: un **Agente Selector** que descarta medidas ya implementadas a partir del propio inventario; un **Agente Recopilador de Información** (Info Gatherer) que abre conversaciones contextualizadas con el usuario para obtener datos no presentes en el JSON; y un **Agente Calculador** que aplica la fórmula específica de cada medida sobre los datos reales de la empresa. Antes de que ningún LLM intervenga, un módulo determinista de pre-filtrado aplica reglas duras de aplicabilidad, por ejemplo, descartar medidas que requieren tejado propio si la empresa declara alquilar el local, o medidas industriales si su código CNAE no es industrial, y clasifica las medidas candidatas en tres grupos según su calculabilidad.

El informe se estructura en cuatro bloques. En el primero (capítulos 3 y 4) se traza **el estado del arte de la descarbonización empresarial y de la aplicación de IA multiagente al ámbito de la sostenibilidad**, identificando marcos metodológicos vigentes (GHG Protocol, Reglamento UE 2023/1115, CSRD) y las arquitecturas técnicas relevantes (AutoGen, LangGraph, CrewAI, RAG con y sin *embeddings*). En el segundo

bloque (capítulos 5 y 6) se describe el **diseño completo de AgentCO2**: arquitectura, agentes, esquema de datos, *frontend* Nuxt 4, *backend* FastAPI con WebSocket, despliegue en contenedores Docker y observabilidad con Langfuse. El tercer bloque (capítulo 7) cataloga las **23 medidas de descarbonización implementadas y sus fórmulas**. El cuarto (capítulos 8 y 9) presenta los **resultados de los experimentos** realizados sobre una empresa piloto de servicios, discute las conclusiones y enumera las líneas de trabajo pendientes.

Los resultados experimentales confirman que la arquitectura propuesta es capaz de generar **planes de mitigación coherentes y trazables** con un coste por inventario del orden de céntimos de euro en llamadas a LLM, y con una latencia de extremo a extremo inferior a un minuto para inventarios de complejidad media. La separación entre código determinista y agentes LLM resulta crítica: las reglas de aplicabilidad embebidas en el pre-filtrado reducen entre un 30% y un 70% el número de medidas que el Selector ha de evaluar, eliminan la mayor parte de los falsos positivos y, sobre todo, eliminan la alucinación de propuestas no aplicables. AgentCO2 se entiende, por tanto, como una **prueba de concepto operativa**, no como un producto cerrado: su valor reside en demostrar la viabilidad técnica del enfoque y en establecer una infraestructura sobre la que ITA y **ECODES** pueden iterar.

La visión a medio plazo del proyecto consiste en **incorporar AgentCO2 como una funcionalidad nativa de SCOPE**. En esa visión, la empresa que termina de calcular su inventario en SCOPE accederá, sin cambio de aplicación ni subida manual de fichero, a un plan personalizado de mitigación generado por el sistema multiagente que aquí se documenta. La transición de un cálculo de huella (lo que SCOPE hace hoy) a un cálculo de huella más un plan accionable (lo que SCOPE haría con AgentCO2 integrado) supone, en términos del Registro Nacional de Huella de Carbono del MITECO, facilitar el paso del sello *Calculo* al sello *Reduzco* sin coste marginal para la empresa. La presente prueba de concepto demuestra que la pieza técnica que cierra ese flujo está disponible.

El razonamiento sobre el que se apoya esta visión es sencillo y se explicita a lo largo del documento: **cada empresa tiene una huella distinta y, por tanto, un plan de reducción óptimo distinto**; las medidas aplicables, los datos disponibles y las restricciones económicas varían de una empresa a otra; pretender que un mismo conjunto de recomendaciones genéricas pueda servir a empresas tan heterogéneas es ignorar la naturaleza misma del problema. La inteligencia artificial y, más concretamente, los sistemas multiagente basados en LLM aporta precisamente la capacidad de razonamiento personalizado que el problema demanda, con un coste marginal y una latencia que la hacen viable a escala.

2. Objeto y Alcance de la Investigación

2.1. Objeto

El objeto de la presente investigación es **reducir, mediante tecnología, el coste y el tiempo** que separa a una pequeña o mediana empresa española de una propuesta personalizada de descarbonización. La **hipótesis** de partida es que *un sistema multiagente correctamente diseñado puede emular gran parte del trabajo que hoy realiza un consultor (leer el inventario, filtrar las medidas pertinentes, pedir los datos que faltan y aplicar la fórmula correcta a cada medida) con una latencia y un coste compatibles con el uso desatendido.*

La investigación no pretende sustituir a la consultoría experta sino **ampliar el número de empresas que pueden disponer de una primera propuesta estructurada en cuestión de minutos**. Los objetivos específicos del trabajo son los siguientes:

1. Diseñar y desplegar una **arquitectura multiagente** que tome como entrada el JSON de huella de carbono producido por ScopeCO2 y devuelva una propuesta de descarbonización compuesto por medidas con su CO₂ reducido, inversión, ahorro anual y payback.
2. Implementar el **catálogo completo de las 23 medidas de descarbonización** codificadas por ECODES, derivado del análisis de calculabilidad realizado por el equipo conjunto ITA-ECODES, como una base de conocimiento jerárquica en JSON que pueda alimentar al agente sin necesidad de una base vectorial.
3. Establecer un **protocolo de evaluación que permita medir la calidad, latencia y coste** del sistema, comparándolo con la alternativa de un único LLM monolítico al que se le pide todo el plan de una sola vez.
4. Generar una **arquitectura técnica reusable** para futuras fases del programa, integración directa con ScopeCO2, ampliación del catálogo, búsqueda web de precios, y documentarla con suficiente detalle para que pueda servir de referencia a otras iniciativas similares en el ámbito de la sostenibilidad.

2.2. Autores

ITA, Instituto Tecnológico de Aragón. Centro tecnológico de carácter público adscrito al Departamento de Ciencia, Universidad y Sociedad del Conocimiento del Gobierno de Aragón. Fundado en 1984, cuenta con más de 300 tecnólogos y está reconocido como Oficina de Transferencia Tecnológica por el Ministerio de Ciencia, Innovación y Universidades. Aporta a este proyecto la capacidad técnica en arquitecturas multiagente, procesamiento de lenguaje natural e innovación orientada a misión.

ECODES, Fundación Ecología y Desarrollo. Organización sin ánimo de lucro fundada en Zaragoza en 1992, alineada con la Agenda 2030 y el Acuerdo de París. **ECODES** es socio oficial del Pacto Climático Europeo en España y co-impulsora de iniciativas como Fundaciones por el Clima y la Comunidad #PorElClima. Es autora del programa **ScopeCO₂**, calculador de huella de carbono conforme al GHG Protocol que constituye la entrada del presente sistema. Aporta el conocimiento experto en medidas de mitigación, encapsulado en las fichas técnicas que el sistema consume como base de conocimiento, y la red institucional para la transferencia de resultados a empresas piloto.

Esta investigación se ha desarrollado con el apoyo del Ministerio para la Transición Ecológica y el Reto Demográfico a través de la convocatoria para el año 2025 de subvenciones en régimen de concurrencia competitiva “línea a” para el desarrollo de actividades para el interés general consideradas de interés sociales, en el ámbito de la investigación científica y técnica y protección al medio ambiente en materias de competencia estatal.

2.3. Alcance

El alcance cubre el **ciclo completo** desde la ingesta inventario GHG hasta la presentación del plan de descarbonización al usuario empresarial:

- i. la lectura del JSON de ScopeCO₂ con sus secciones de Alcance 1, 2 y 3
- ii. un pre-filtrado determinista de medidas no aplicables por reglas duras (campo ausente, local alquilado, CNAE no industrial)
- iii. la clasificación de las medidas candidatas en tres grupos por calculabilidad
- iv. el descarte selectivo por un Agente Selector LLM
- v. la recopilación conversacional de datos faltantes por un Agente Info Gatherer LLM
- vi. la aplicación de fórmulas por un Agente Calculador LLM; y
- vii. la presentación final del plan en una interfaz web con tres secciones diferenciadas de medidas (listas, que pueden afinarse y recomendadas cualitativamente).

Quedan fuera del alcance la generación del inventario (competencia de ScopeCO₂), el seguimiento longitudinal del cumplimiento del plan tras la entrega y la integración directa con sistemas externos de gestión empresarial. La integración de extremo a extremo con ScopeCO₂ (sin paso manual de fichero) está identificada como línea de trabajo futura.

El sistema se concibe como una prueba de concepto operativa (está desplegado en infraestructura de producción del ITA *agentco2-ecodes.ita.es* y puede ser utilizado por empresas reales, pero no como un producto comercial cerrado).

3. La huella de carbono empresarial y el reto de la descarbonización

3.1. Estado del arte

La descarbonización de la economía es uno de los retos más exigentes que ha asumido la sociedad contemporánea. El Acuerdo de París (UNFCCC, 2015), el Pacto Verde Europeo (Comisión Europea, 2019) y la Ley Europea del Clima (Reglamento UE 2021/1119) han fijado el objetivo de neutralidad climática para 2050 y reducciones intermedias del 55% para 2030 sobre niveles de 1990. Estos compromisos políticos descansan, en última instancia, sobre la acción concreta de millones de **organizaciones empresariales** (desde grandes multinacionales hasta pymes) y, por tanto, sobre la **capacidad de cada una de ellas para identificar, cuantificar y ejecutar medidas de mitigación específicas a su actividad**. El IPCC en su sexto informe de evaluación (IPCC, 2022) subraya que la viabilidad de los escenarios compatibles con 1,5 °C depende de reducciones drásticas en todos los sectores productivos, lo que obliga a desplegar mecanismos de mitigación a escala organizativa muy superiores a los actuales.

El estándar metodológico de referencia para la contabilidad de emisiones es el *GHG Protocol Corporate Accounting and Reporting Standard* (GHG Protocol, 2015), que divide las emisiones en tres alcances: Alcance 1 (emisiones directas de fuentes propias, como combustión en calderas o vehículos de flota), Alcance 2 (emisiones indirectas asociadas a la electricidad, el vapor o el calor comprados) y Alcance 3 (resto de emisiones indirectas, incluyendo desplazamientos in itinere, viajes profesionales, residuos, agua, papel, cadena de suministro y uso de los productos vendidos). El Alcance 3 es habitualmente el más voluminoso y, paradójicamente, el más difícil de medir y mitigar de forma sistemática, ya que escapa al control directo de la organización informante (GHG Protocol, 2013; WBCSD, 2023).

La normativa europea ha acelerado en los últimos años la obligación de reportar emisiones. La Corporate Sustainability Reporting Directive (CSRD), Directiva (UE) 2022/2464, ha ampliado el universo de empresas obligadas a publicar información sobre sostenibilidad, incluida la huella de carbono, de unas 12.000 a más de 50.000 organizaciones en la Unión Europea, integrando además los *European Sustainability Reporting Standards* (ESRS) elaborados por EFRAG (EFRAG, 2023). La *Corporate Sustainability Due Diligence Directive* (CSDDD), Directiva (UE) 2024/1760, extiende a las empresas la responsabilidad sobre los impactos de su cadena de valor. El Reglamento (UE) 2023/1115 sobre productos asociados a la deforestación añade un eje adicional de trazabilidad. El resultado conjunto es un volumen creciente de empresas que disponen o están obligadas a disponer de un inventario GHG, pero que no necesariamente disponen de un plan de mitigación accionable.

Esta asimetría entre **medir y actuar** ha sido documentada por la literatura de implementación climática. Bain & Co. (2024), en su *Energy and Natural Resources Report*, identifica que solo el 38% de las grandes empresas con objetivos públicos de reducción de emisiones disponen de un plan operativo detallado con medidas concretas, presupuesto y calendario asignado. McKinsey & Company (2023), en *The Net-Zero Transition*, estima que el gasto adicional de capital necesario para alcanzar los objetivos europeos asciende a 28 billones de euros acumulados hasta 2050 y subraya que el cuello de botella no es financiero sino organizativo: la mayoría de las pymes carece de los recursos técnicos para evaluar qué medidas son aplicables a su contexto concreto. El *World Business Council for Sustainable Development* (WBCSD, 2023) llega a una conclusión paralela al destacar que la disponibilidad de hojas de ruta sectoriales no se traduce automáticamente en planes corporativos accionables.

En el ámbito tecnológico, la combinación de modelos de lenguaje de gran escala (LLM) con bases de conocimiento estructuradas ha abierto la posibilidad de automatizar parte del trabajo experto que cierra esa brecha. Los **modelos fundacionales** (Bommasani et al., 2022) y, en particular, los **modelos generativos de gran escala** (Brown et al., 2020) demostraron por primera vez una capacidad de razonamiento sobre lenguaje natural utilizable en producción. Los trabajos de Wu et al. (2023) sobre AutoGen, de Schick et al. (2023) sobre **Toolformer** y de Shinn et al. (2023) sobre **Reflexion** han demostrado, ya en el plano de los sistemas de agentes, que la descomposición de tareas complejas en cadenas de agentes especializados mejora sistemáticamente sobre la *prompt-engineering* monolítica en tareas que combinan razonamiento, uso de herramientas y decisión bajo restricciones. La revisión de Guo et al. (2024) sintetiza el estado del arte sobre LLM multi-agente, mostrando cómo este tipo de arquitecturas se han generalizado en problemas que mezclan reglas explícitas y juicio cualitativo. Es esta **línea de investigación la que ha inspirado directamente la arquitectura de AgentCO2**.

España presenta una situación propia respecto a la contabilidad de huella de carbono. Por un lado, dispone de un **Registro Nacional de Huella de Carbono** gestionado por la Oficina Española de Cambio Climático (OECC) del MITECO, creado por el Real Decreto 163/2014, que reconoce la huella calculada y verificada con tres sellos sucesivos, *Calculo*, *Reduzco*, *Compenso*. Según la memoria anual del propio registro (MITECO, 2025), más de 4.500 organizaciones españolas se han inscrito en el registro hasta 2025, con una tasa de crecimiento anual sostenida del 18%. Por otro lado, buena parte de esas inscripciones se quedan en el nivel *Calculo*: solo aproximadamente un tercio progresa al sello *Reduzco*, que exige presentar y ejecutar un plan de reducción documentado (MITECO, 2025).

ScopeCO2, el calculador desarrollado por ECODES, es una de las herramientas de referencia para el cálculo de huella en el contexto español, especialmente para pymes. Sigue la estructura del GHG Protocol (GHG Protocol, 2015) y produce un cálculo de huella de carbono normalizado con consumos y emisiones desagregados por fuente y aplicación, incluyendo los factores de emisión oficiales del MITECO/OECC (MITECO/OECC, 2023), de DEFRA (DEFRA, 2024) y de la Oficina Catalana del Cambio Climático (OCCC, 2024) según el vector energético o el modo de transporte considerado. Este cálculo es precisamente la entrada que consume AgentCO2: la decisión de partir de un esquema bien definido y auditable, en lugar de aceptar entradas no estructuradas, es una decisión arquitectónica deliberada que reduce drásticamente el espacio de error del sistema.

El catálogo de medidas de descarbonización que sirve de base de conocimiento a AgentCO2 procede directamente del trabajo técnico que **ECODES** ha desarrollado durante más de una década acompañando a empresas españolas. Cada una de las 23 medidas del catálogo cuenta con una ficha técnica que recoge la fórmula de cálculo aplicable, los factores de emisión necesarios, citando la fuente oficial (MITECO, OECC, DEFRA, OCCC, IDAE), y un rango de inversión orientativo basado en precios de mercado españoles. La traducción de esas fichas desarrolladas por personas expertas, habitualmente PDF o documentos Word, a JSON estructurado, junto con el análisis de qué datos del inventario ScopeCO2 alimentan cada fórmula, ha sido uno de los trabajos técnicos previos al desarrollo software del sistema. **El sistema AgentCO2 hereda enteramente la estructura del GHG Protocol porque ScopeCO2, su entrada, está construido sobre ella.** Conviene, por tanto, recordar brevemente qué representa cada uno de los tres alcances y cómo se traducen al esquema JSON que el sistema consume:

- **Alcance 1, emisiones directas.** Son las emisiones de gases de efecto invernadero que proceden de fuentes propiedad de la organización o controladas por ella. En el JSON de ScopeCO2 se reflejan principalmente en las secciones *combustibles_fosiles* (combustión en calderas de calefacción, ACS o procesos industriales), *desplazamientos_propiedad* (combustible consumido por la flota de vehículos propios) y *gases_refrigerantes* (fugas de gases F en equipos de climatización y refrigeración). En el catálogo de medidas, las **seis medidas que mitigan alcance 1** están codificadas con el prefijo **M1-**: M1-COMBUSTION, M1-AISLAMIENTO, M1-VENTANAS, M1-FLOTA-EV, M1-BIOCOMBUSTIBLE y M1-REFRIGERANTES.
- **Alcance 2, emisiones indirectas asociadas a la energía comprada.** Son las emisiones que la organización no produce directamente pero que resultan de la *generación de la electricidad, el vapor, la calefacción o la refrigeración que consume*. En el JSON se reflejan principalmente en la sección *electricidad*, con campos que indican el factor de emisión aplicable y el porcentaje de Garantías de Origen renovables del contrato (*goO_%*). En el catálogo de medidas, las **cinco medidas** codificadas como **M2-** son M2-SOLAR (autoconsumo fotovoltaico), M2-LED (iluminación eficiente), M2-RENOVABLE (comercializadora renovable), M2-GDO (compra de Garantías de Origen) y M2-COMPRESOR (optimización de aire comprimido).
- **Alcance 3, el resto de las emisiones indirectas.** Es la categoría más amplia y heterogénea: viajes profesionales, desplazamientos in itinere de los empleados, consumo de agua y papel, residuos generados, alojamiento en viajes, transporte subcontratado de mercancías y, en su forma más exhaustiva, todas las emisiones embebidas en la cadena de suministro y en el uso de los productos vendidos. AgentCO2 cubre las subcategorías que ScopeCO2 ya modela: hay **doce medidas** con prefijo M3-, agrupadas por tema (agua, papel, viajes, movilidad, residuos, mercancías, política de viajes y teletrabajo).

Disponer de un catálogo de medidas no equivale a disponer de un plan. Un catálogo es una lista de posibles intervenciones, con su fórmula, sus factores de emisión y su rango de inversión, pero solo se convierte en plan cuando se filtra por la aplicabilidad real a la empresa concreta, se calcula con sus datos reales y se ordena por una métrica de impacto-coste compatible con sus restricciones presupuestarias y temporales. La traducción de catálogo a plan es la operación que históricamente ha realizado el consultor

experto y que constituye, en términos operativos, el cuello de botella económico del proceso.

Esta operación tiene una estructura razonablemente bien definida, filtrar, recopilar datos, calcular, ordenar, pero está cargada de heterogeneidad: cada medida exige datos distintos, fórmulas distintas y reglas de aplicabilidad distintas. Un sistema software clásico podría modelar cada medida como un módulo independiente, pero pagaría el precio de tener que pre-codificar todas las decisiones de aplicabilidad y todas las preguntas que el sistema haría al usuario. La introducción de un LLM permite delegar esas decisiones de interpretación, *¿es razonable proponer un cambio de ventanas si el edificio es de 1985?*, *¿qué porcentaje de empleados podrían pasar a teletrabajo?*, al razonamiento lingüístico del modelo, mientras que el código retiene las reglas duras que no admiten ambigüedad.

4. Inteligencia artificial aplicada a la descarbonización

4.1. Sistemas multiagente

Una pregunta legítima al inicio del diseño fue: ¿no bastaría con pasarle a un único modelo de lenguaje el JSON de huella, el catálogo de medidas y un *prompt* bien escrito, y dejar que devuelva el plan completo? La respuesta del equipo fue negativa, por al menos cinco razones independientes que se han confirmado durante el desarrollo.

- **Tamaño del contexto.** El catálogo completo, con las 23 fichas técnicas en JSON, supera los 80.000 tokens. Sumado al inventario de la empresa, las preguntas y respuestas y los ejemplos, el contexto crece rápidamente, encarece cada llamada y empeora la atención del modelo a los detalles, fenómeno documentado como *lost in the middle* por Liu et al. (2024), que muestran cómo los LLM pierden capacidad de recuperar información situada en la zona central de contextos largos.
- **Trazabilidad de las decisiones.** Un único LLM produce un resultado final pero no expone *por qué* descartó una medida o eligió un factor de emisión concreto. La separación en agentes especializados, cada uno con su salida estructurada, convierte cada decisión en un evento observable y auditable, propiedad clave subrayada por la revisión de Guo et al. (2024) como una de las ventajas estructurales de los sistemas multi-agente sobre los monolíticos.
- **Reglas duras versus juicio lingüístico.** Algunas decisiones, un código CNAE no industrial descarta el compresor, un campo ausente descarta la medida, son binarias y no deben quedar al criterio del modelo. Pasar esas reglas a código Python en el pre-filtrado elimina la posibilidad de alucinación.
- **Datos faltantes y diálogo dirigido.** Hay medidas para las que el JSON no contiene todo lo necesario (M2-SOLAR exige la superficie del tejado, M3-MOVILIDAD exige el porcentaje de empleados dispuestos al cambio modal). Un único LLM tendría que mezclar la fase de recopilación con la de cálculo, produciendo conversaciones largas y confusas. El *Info Gatherer* abre un diálogo independiente por medida.
- **Iteración y mantenimiento.** Modificar el comportamiento de un único *prompt* monolítico es un ejercicio frágil. Modificar agente especializado, añadir ejemplo al *Selector*, ajustar fórmula del *Calculador*, son cambios locales que se evalúa y despliega aisladamente.

El diseño final adopta una arquitectura híbrida: el núcleo del flujo es código Python determinista (carga de JSON, pre-filtrado, clasificación, despacho), y los puntos del flujo que requieren razonamiento sobre lenguaje natural o juicio experto son delegados a

agentes LLM con *prompts* específicos y salidas estructuradas (Pydantic). El patrón resultante es robusto frente a errores del modelo y fácilmente extensible a nuevos casos.

4.2. La coordinación entre agentes

El **framework** elegido para orquestar los agentes LLM es **AG2** (anteriormente conocido como AutoGen; Wu et al., 2023), mantenido por la comunidad tras la separación de *Microsoft Research*. AG2 ofrece un modelo de agentes conversacionales que pueden intercambiarse mensajes, llamar a herramientas y ceder el turno siguiendo un patrón de *handoff* explícito. La filosofía subyacente, **agentes que se comunican por mensajes en lenguaje natural y producen salidas estructuradas**, se ha consolidado como una de las aproximaciones más productivas a los sistemas LLM multi-agente, como muestran Hong et al. (2024) con MetaGPT y Qian et al. (2024) con ChatDev en el dominio del desarrollo software, o Park et al. (2023) en simulación social. Está integrado nativamente con OpenAI, Anthropic y otros proveedores de modelos, y expone un sistema de eventos que permite instrumentar cada paso del flujo con observabilidad. El patrón ReAct (Yao et al., 2023) de razonamiento intercalado con acción y la técnica de *chain-of-thought* (Wei et al., 2022) están incorporados de forma transparente en el comportamiento por defecto de los agentes AG2.

En **AgentCO2** los tres agentes, Selector, Info Gatherer y Calculador, se crean con funciones factoría *create_selector_agent*, *create_measure_gatherer_agent*, *create_calculator_agent* que reciben una configuración LLM común (*get_llm_config*) y unas variables de contexto (*ContextVariables*) que actúan como memoria compartida entre ellos. El orquestador principal (*DecarbonizationAgent*) configura los *handoffs*: el Selector cede al Calculador y el Calculador cede a *TerminateTarget* para cerrar la ejecución. El Info Gatherer opera por separado, en una conversación dedicada por medida, y persiste sus resultados en el contexto antes de que el Calculador entre en juego. Las variables de contexto compartidas son la pieza clave de la coordinación. El esquema es el siguiente:

```
{
  "original_json": dict,          # JSON de ScopeCO2
  "json_loaded": bool,
  "local_propio": bool | None,   # propio / alquilado
  "candidate_measures": list[str], # tras pre-filtrado
  "groups_calculable": list[str], # con fórmula y datos
  "groups_needs_modal": list[str], # falta dato extra
  "groups_proposable": list[str], # solo cualitativas
  "selected_calculable": list[str], # tras Selector
  "selected_needs_modal": list[str], # tras Selector
  "measure_fields": dict         # datos del Info Gatherer
}
```

Cada agente lee del contexto los campos que necesita y escribe los suyos al terminar. El Selector escribe *selected_calculable* y *selected_needs_modal*; el Info Gatherer escribe *measure_fields*; el Calculador consume todo lo anterior y produce la salida final como un modelo Pydantic *ResultadosCalculador* con una lista de *MedidaCalculada*. Los *prompts* de cada agente se cargan desde ficheros de texto plano (*selector.txt*, *calculator.txt*, *measure_gatherer.txt*) que el panel de administración del frontend permite editar sin redespigüe.

4.3. RAG jerárquico sin vectores

La elección de la **base de conocimiento**, las **23 fichas de medidas**, es uno de los puntos en los que AgentCO2 se aparta de los patrones más habituales en sistemas RAG (Retrieval-Augmented Generation; Lewis et al., 2020). La intuición estándar es indexar el corpus con *embeddings*, almacenarlo en una base vectorial (Chroma, Pinecone, Qdrant) y recuperar los *k* documentos más similares a la consulta del usuario. Para AgentCO2 esta opción se descartó por tres razones.

- **El corpus es pequeño y cerrado.** Veintitrés fichas no justifican el coste operativo de una base vectorial, ni en infraestructura ni en latencia.
- **La recuperación no es semántica.** Lo que decide qué medidas se cargan no es la similitud textual con una pregunta, sino reglas duras sobre campos del JSON de ScopeCO2 (presencia del campo *combustibles_fosiles*, código CNAE, régimen de tenencia del local).
- **La trazabilidad importa.** Una recuperación por *embeddings* es, en última instancia, una caja parcialmente negra. La recuperación basada en reglas explícitas es auditable y reproducible.

La estructura adoptada es un **RAG jerárquico en dos niveles**. El primer nivel, siempre cargado en el contexto del Selector, es un *index.json* que contiene una entrada compacta por cada una de las 23 medidas con sus campos clave: *scopeco2_field*, *calculation_type*, *requires_extra_data*, *required_inputs_modal*, *formula_source*, *notes*. El segundo nivel, cargado solo cuando una medida ha sido seleccionada, son los ficheros completos *MROXXX.json* con la fórmula detallada, los factores de emisión, las fuentes y los ejemplos. Este diseño en dos niveles mantiene el contexto del Selector ligero y carga el detalle pesado solo cuando se necesita.

4.4. Observabilidad de agentes

La depuración de un sistema multiagente plantea un problema específico: el flujo de control no es secuencial sino ramificado, los mensajes intercambiados son largos y la causa de un fallo puede estar a tres saltos de distancia del síntoma. AgentCO2 incorpora **observabilidad** nativa con dos herramientas complementarias.

- **Langfuse** es la plataforma elegida para la **trazabilidad**. Cada conversación abre una *session* identificada con el usuario y el identificador de conversación; cada llamada a *agent.chat()* o *agent.run_pipeline()* se envuelve en una *observation* que captura el *prompt* completo, la respuesta, los tokens consumidos, el coste asociado y la latencia. Los *handoffs* entre agentes se registran como *spans* anidados, lo que permite reconstruir el árbol completo de ejecución desde el dashboard. La activación de Langfuse es opcional: si las variables *LANGFUSE_PUBLIC_KEY*, *LANGFUSE_SECRET_KEY* y *LANGFUSE_HOST* no están definidas en el *.env*, el sistema arranca con degradación elegante y sin telemetría.

- **OpenTelemetry** complementa Langfuse para la instrumentación del *wrapper* LLM de AG2 (*autogen.opentelemetry.instrument_llm_wrapper*). Esto permite que las herramientas estándar de observabilidad, Grafana Tempo, Jaeger, Honeycomb, puedan recoger trazas del sistema sin acoplarse al producto comercial concreto. La elección de OpenTelemetry como estándar de instrumentación es una decisión deliberada de **portabilidad**: si en el futuro se decide cambiar de Langfuse a otra plataforma, la instrumentación no requerirá una reescritura.

4.5. Por qué la IA aporta valor diferencial en la reducción de huella

Conviene detenerse en una pregunta de fondo: ¿por qué la inteligencia artificial es la tecnología adecuada para esta tarea, frente a alternativas más sencillas y maduras? La respuesta tiene cinco vertientes, todas relacionadas con la naturaleza intrínsecamente personalizada del problema de la descarbonización empresarial.

- **Cada empresa tiene una huella distinta.** Dos empresas del mismo sector y tamaño pueden presentar perfiles de emisión radicalmente diferentes en función de su mix energético, su estructura de movilidad, sus instalaciones, la edad de su edificio o el régimen de tenencia de este. No existe un plan único que sirva a todas. Como documenta la *Science Based Targets initiative* (SBTi, 2024), incluso dentro de un sector estrechamente acotado, por ejemplo, hostelería, la **dispersión de huellas por euro facturado supera el factor 5 entre empresas comparables**, lo que impide ofrecer recomendaciones genéricas con utilidad práctica. Esta heterogeneidad es coherente con los escenarios sectoriales del informe del Grupo III del IPCC (IPCC, 2022), que enfatiza que las palancas de mitigación eficaces son altamente dependientes del contexto.
- **Las medidas de mitigación tienen reglas de aplicabilidad heterogéneas.** Una medida puede ser ideal para una empresa y absurda para otra. Sustituir las ventanas no tiene sentido en un edificio alquilado; cambiar la flota a eléctrica carece de objeto si no hay flota; reciclar papel es irrelevante si la empresa ya lo hace. La IA permite leer el inventario, interpretar el contexto y decidir qué subconjunto de medidas tiene sentido para esa empresa concreta, en una operación que sería costosa de codificar como un sistema clásico de reglas, porque las reglas se entrelazan con la interpretación lingüística de los datos.
- **Los datos del inventario son incompletos por diseño.** SCOPE captura las emisiones que pueden cuantificarse a partir de facturas y consumos, pero no, por ejemplo, la superficie de tejado disponible para una instalación fotovoltaica, ni el inventario detallado de luminarias, ni el porcentaje de empleados susceptibles de teletrabajo. Pedir todos esos datos al usuario en el momento del cálculo de huella sería intrusivo y dañaría la experiencia. Pedirlos solo cuando son necesarios, y solo para las medidas que el usuario quiere afinar, exige un sistema capaz de mantener

una conversación natural y estrictamente acotada a cada medida. Ese es precisamente el papel del agente recopilador.

- **El coste de un asesor humano no escala.** Como señalaba el informe de IEA (2023) sobre eficiencia energética y barreras a la inversión, en Europa hay aproximadamente 23 millones de pymes; solo una **fracción mínima dispone de presupuesto para contratar consultoría de sostenibilidad**. La revisión de Rolnick et al. (2022) sobre aplicaciones de *machine learning* al cambio climático argumenta precisamente que las tecnologías de IA son particularmente valiosas allá donde el coste del análisis humano excede el valor de la decisión: la IA es, en consecuencia, la única tecnología que permite ofrecer un primer plan personalizado a precio marginal cercano a cero por empresa, abriendo la puerta a una difusión masiva que un modelo humano no podrá replicar.
- **Combinar razonamiento cualitativo con cálculo numérico estricto.** El razonamiento cualitativo, qué medida tiene sentido para qué empresa, qué dato pedir y cómo formular la pregunta, cómo redactar la explicación al usuario, es competencia natural de un LLM. El cálculo numérico estricto, aplicar una fórmula con factores de emisión oficiales y devolver una cifra correcta, exige **determinismo**. La descomposición multiagente permite combinar: LLM razona, el código aplica la fórmula sobre los datos que el LLM ha extraído y validado.

La combinación de estos cinco argumentos hace de la **IA multiagente** una opción no solo adecuada sino, en términos prácticos, la única que combina **personalización, escalabilidad, coste marginal viable y trazabilidad técnica**. Es esta convergencia la que ha motivado el diseño de **AgentCO2** y la que justifica que **ECODES** y el ITA hayan invertido recursos en la presente prueba de concepto.

5. Reto técnico abordado

5.1. Heterogeneidad del catálogo de medidas

El **catálogo de 23 medidas** que constituye la base de conocimiento del sistema es radicalmente **heterogéneo**. Algunas medidas tienen una fórmula cerrada y todos los datos que necesitan ya están en el JSON de ScopeCO2 (por ejemplo, M2-RENOVABLE: pasar a comercializadora renovable depende únicamente del consumo eléctrico y del factor de emisión actuales, ambos presentes en el inventario). Otras tienen una fórmula, pero requieren datos adicionales que el inventario no captura (M2-SOLAR exige la superficie de tejado; M2-LED exige el inventario de luminarias). Otras carecen incluso de fórmula cuantitativa y solo pueden proponerse como recomendación estratégica (M3-POLITICA-VIAJES, marco organizativo cuyo impacto se captura mediante otras medidas). Esta heterogeneidad obliga a clasificar las medidas en tres grupos antes de procesarlas:

- **Grupo 1, Calculables directamente.** Tienen fórmula y todos los datos necesarios en el JSON. Pueden ejecutarse de forma autónoma por el Calculador sin diálogo previo.
- **Grupo 2, Necesitan datos extra.** Tienen una fórmula, pero necesitan recopilar datos adicionales mediante un modal conversacional con el usuario. Pasan por el Info Gatherer antes de llegar al Calculador.
- **Grupo 3, Solo proponibles.** No tienen fórmula cuantitativa. Se presentan al usuario como recomendación cualitativa sin estimación numérica de impacto.

La clasificación se realiza sobre los campos *has_formula*, *calculation_type* y *requires_extra_data* del *index.json*, implementada en la función *classify_candidates* del módulo *services/measures.py*. Es un código de unas veinte líneas, totalmente determinista, que precede a cualquier llamada al LLM. La consecuencia operativa es que el Selector y el Calculador reciben grupos ya delimitados, lo que simplifica enormemente sus *prompts*.

5.2. Datos que faltan

El problema de la **falta de datos** es probablemente el más delicado del sistema desde el punto de vista de la experiencia de usuario. Una empresa que ha cargado su inventario en ScopeCO2 no espera tener que responder a un “interrogatorio” largo sobre datos que su asesor no le ha pedido antes. La estrategia adoptada en AgentCO2 es diferir el diálogo: el plan inicial muestra las medidas calculables ya con sus números, y deja en una sección aparte, Sección 2, *medidas que puedes afinar*, las medidas que requieren datos extra. El usuario decide qué medidas concretas quiere afinar y, sólo entonces, se abre un diálogo específico para cada una. Cada **diálogo está estrictamente acotado a la medida concreta**. El *prompt* del Info Gatherer recibe dinámicamente:

El identificador y el título de la medida que se quiere afinar.

El *agent_context* que la ficha JSON define para esa medida (un texto en lenguaje natural que indica al modelo qué preguntar, en qué orden y con qué referencias orientativas).

Los campos ya recopilados y los campos pendientes.

La instrucción de preguntar de uno en uno (o dos en dos como máximo) y de marcar *datos_completos = True* cuando todos los campos requeridos estén presentes.

El esquema de respuesta del modelo está restringido por un modelo Pydantic generado dinámicamente con *pydantic.create_model*, lo que garantiza que los campos extraídos están tipados, que ningún campo se inventa y que el *frontend* puede confiar en la estructura. Los campos recopilados se acumulan en *context_variables["measure_fields"]* y se inyectan automáticamente en el contexto del Calculador cuando éste tome el relevo.

5.3. El valor de la personalización

La piedra angular del enfoque adoptado en AgentCO2, y, en general, de cualquier sistema multiagente aplicado a la descarbonización, es la convicción de que la **utilidad real del sistema reside en su capacidad de personalizar las propuestas a cada empresa**.

Esta hipótesis merece argumentarse con detalle porque marca la diferencia entre un producto que aporta valor y un producto que, en el mejor de los casos, replica documentos genéricos ya disponibles en internet:

- Una empresa que decide reducir su huella necesita responder a cuatro preguntas concatenadas: qué medidas son aplicables a mi caso; cuánto reduzco con cada una; cuánto me cuesta y cuándo recupero la inversión; en qué orden las ejecuto. La respuesta a cada una de esas preguntas depende íntegramente de los datos concretos de la empresa: si una empresa no tiene flota propia, las medidas de flota son irrelevantes; si su contrato eléctrico ya es renovable, cambiar de comercializadora no aporta nada; si su edificio es alquilado, mejorar el aislamiento es discutible; si su sector no es industrial, optimizar el aire comprimido es absurdo. Un consultor experto razona sobre todas estas particularidades en su cabeza, mentalmente, antes de siquiera enunciar la primera medida. Lo que AgentCO2 hace es codificar ese razonamiento, en código determinista para las reglas duras y en agentes LLM para las decisiones matizadas, de forma que pueda ejecutarse automáticamente para cada empresa concreta.
- Documentos como las guías sectoriales del MITECO, las fichas técnicas del IDAE o los catálogos de buenas prácticas de organismos internacionales son recursos extremadamente valiosos, pero, por su naturaleza, son genéricos. Le dicen al usuario "considera instalar paneles solares" sin poder responder al "¿cuánto reduciría yo?" o al "¿cuánto me costaría a mí?". Salvar esa distancia entre lo genérico y lo específico es una operación que históricamente sólo la persona experta podía realizar. La inteligencia artificial cambia esa ecuación al permitir que la operación de particularización se automatice, manteniendo la fidelidad a las fórmulas oficiales y a los factores de emisión reconocidos.

- En el contexto específico de SCOPE, este principio se traduce en una secuencia natural de utilidad: una empresa entra en SCOPE para calcular su huella, obtiene un inventario riguroso conforme al GHG Protocol y, inmediatamente después, recibe sin coste adicional un plan de mitigación calculado sobre ese mismo inventario, con cifras concretas para su empresa concreta. La operación es transparente para el usuario y, sin embargo, encierra la transformación que estructura todo el proyecto: dejar de *medir el problema* para empezar a *proponer la solución*. AgentCO2 es la pieza técnica que materializa esa transformación.

Una observación importante: el plan generado por AgentCO2 no pretende sustituir el juicio del consultor humano cuando éste sea necesario. Las empresas de mayor complejidad, sujetas a CSRD o con planes de transición exigentes, seguirán necesitando consultoría experta para enriquecer el plan automático, validar las estimaciones, alinearlos con su estrategia financiera y, en su caso, hacer el seguimiento de la ejecución. **El papel de AgentCO2 es democratizar el primer plan estructurado**, el que hoy una gran mayoría de pymes no llega a tener, y, secundariamente, acelerar el trabajo del consultor cuando lo hay, ofreciendo un borrador trazable y auditable sobre el que iterar.

6. Diseño de la herramienta: AgentCO2

6.1. Concepto y propósito

AgentCO2 es la materialización como software de la hipótesis de investigación: que un sistema multiagente puede convertir un inventario GHG en un plan de descarbonización personalizado sin intervención humana, en cuestión de minutos y por un coste marginal por inventario que está dos o tres órdenes de magnitud por debajo del coste de una consultoría. La aplicación está disponible en la URL *agentco2-ECODES.ita.es* y se puede usar, previa subida del JSON de ScopeCO2 o cargando el caso demo *la_puebla_de_alfinden.json*, sin registro previo.

Desde el punto de vista del usuario, la **experiencia** se estructura en cuatro pasos:

- i. carga del JSON o selección del caso demo;
- ii. declaración del régimen de tenencia del local (propio o alquilado)
- iii. ejecución del pipeline, visualizada en tiempo real mediante una página de *procesando* que muestra los eventos del agente (*STEP*, *MEASURES_READY*) a medida que ocurren
- iv. llegada a la página de *informe* con las tres secciones de medidas, filtros por horizonte temporal y por inversión máxima, y la opción de descargar el informe en PDF.

6.2. Arquitectura general

La arquitectura es la habitual de una aplicación web moderna con dos servicios principales y una capa de infraestructura.

- **Backend Python.** FastAPI + Uvicorn + AG2. Expone una API REST y un WebSocket bajo el prefijo */api/v1/agents* para chat, ejecución del *pipeline*, carga del JSON, declaración del régimen del local y consulta de medidas bloqueadas. Persistencia mínima: el estado de cada usuario se mantiene en memoria en el proceso y se reinicia en cada despliegue.
- **Frontend Nuxt 4.** Vue 3 con páginas para los pasos del flujo, susceptibles de ser *componentes*, que envuelven el WebSocket y la generación del PDF, y un servidor Nitro que actúa como pasarela entre el navegador y el backend. Incluye un panel de administración protegido por autenticación básica para editar las medidas y los *prompts* sin redespargar.

- **Infraestructura.** Despliegue con Docker Compose en una máquina Ubuntu, expuesto al exterior por un Nginx con terminación TLS gestionada por Certbot. Observabilidad opcional con Langfuse y OpenTelemetry.

La *figura 1* refleja el flujo completo desde la entrada (el JSON de huella generado por SCOPE) hasta las tres secciones del informe final, distinguiendo por color las piezas deterministas (Python), la clasificación, los agentes LLM y las secciones de salida.

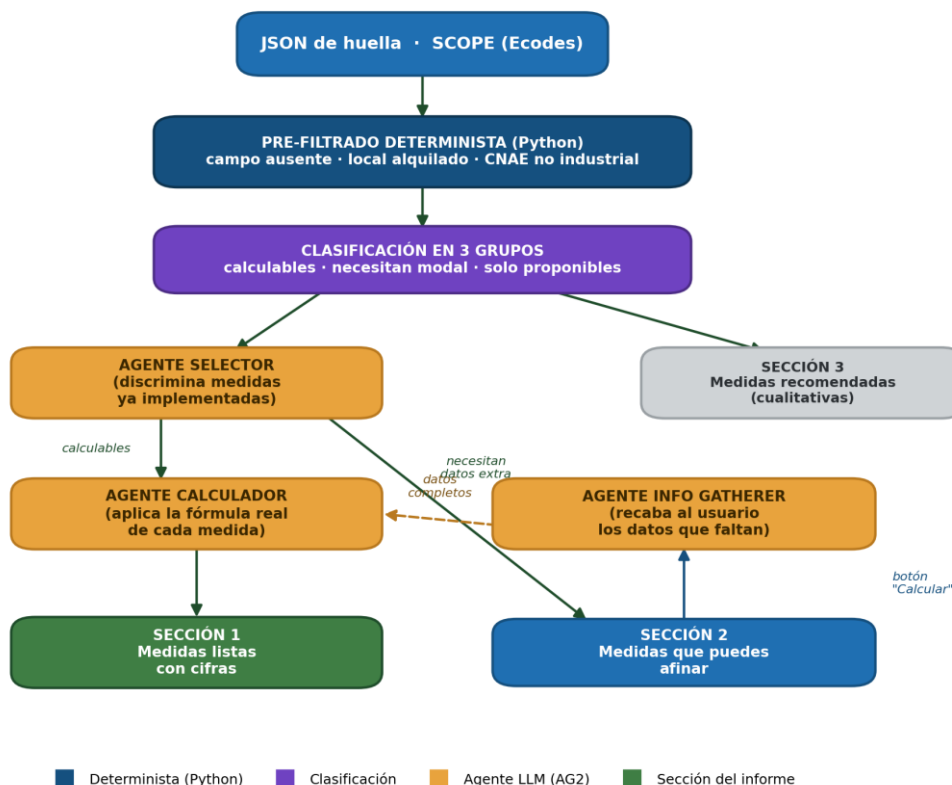


Figura 1. Arquitectura y flujo de agentes de AgentCO2. El JSON de SCOPE atraviesa el pre-filtrado determinista y la clasificación antes de que intervenga el primer agente LLM (Selector). El Calculador produce las cifras; el Info Gatherer recaba datos faltantes cuando el usuario decide afinar una medida de la Sección 2.

6.3. El pipeline de tres agentes

Antes de entrar en cada agente individualmente, conviene fijar el patrón general que articula el sistema. AgentCO2 se construye sobre un flujo de **tres roles** bien diferenciados que se corresponden con tres preguntas sucesivas que tendría que responder cualquier consultor humano enfrentado a un inventario:

- **¿Qué medidas debemos considerar para esta empresa?**, un **primer agente discriminador** examina el JSON de huella generado por SCOPE y, combinado con el pre-filtrado determinista, decide qué subconjunto de medidas tiene sentido evaluar. Es el rol del **Agente Selector**.
- **¿Cuánto se reduce con cada medida y cuánto cuesta aplicarla?**, un **agente calculador** aplica las fórmulas concretas de cada medida sobre los datos reales de la empresa y produce las cifras del plan. Es el rol del **Agente Calculador**.

- **¿Qué falta saber para calcular una medida concreta?**, un agente recopilador abre, cuando es necesario, un diálogo focalizado con el usuario para obtener los datos que el JSON de SCOPE no contiene. Es el rol del **Agente Info Gatherer**.

La separación de estas tres responsabilidades en tres agentes distintos es la consecuencia directa de que las **tres preguntas** tienen naturalezas cognitivas diferentes: la primera es una pregunta de **aplicabilidad** (juicio sobre el contexto), la segunda es una pregunta de **cálculo** (operación determinista), la tercera es una pregunta de **diálogo** (conversación natural). Asignar a cada pregunta el agente adecuado, con su *prompt*, su esquema de salida y su conjunto de herramientas, produce un sistema más robusto que un único modelo intentando responder las tres a la vez.

El diagrama del flujo, ya esbozado en la sección 6.2, puede leerse ahora con más precisión: el JSON de SCOPE entra al sistema, el pre-filtrado y el primer agente discriminan, el calculador produce cifras para las medidas directamente calculables, y para las medidas que necesitan afinarse el frontend invoca al recopilador, quien tras completar los datos devuelve el control al calculador. La salida final del flujo es un plan estructurado en tres secciones que el usuario consume desde su navegador.

Agente Selector, el primer agente que discrimina

El Selector recibe las listas *groups_calculable* y *groups_needs_modal* producidas por el pre-filtrado determinista, junto con el bloque *Emissions_CO2e* del JSON de ScopeCO2. Su tarea es estrictamente **excluyente**: descartar medidas que ya estén implementadas en la empresa o que claramente no apliquen, según evidencia presente en el propio inventario. Los criterios típicos de descarte se expresan en el *prompt* mediante ejemplos:

EJEMPLOS de cuándo descartar:

- M2-RENOVABLE (comercializadora renovable): si go0_% ya es 100
- M2-GDO (garantías de origen): si go0_% ya es 100
- M3-PAPEL (papel reciclado): si application_name ya indica reciclado
- M3-VUELO-BUSINESS: si no hay vuelos de media/larga distancia
- M3-VUELO-TREN: si no hay vuelos nacionales
- M3-MERCANCIAS-BARCO: si no hay application_name 'Vuelo mercancías'

NO descartes una medida si no estás seguro. En caso de duda, inclúyela.

La salida del Selector está restringida por el modelo Pydantic *ResultadoSelector* con cuatro campos: *calculable* (IDs que sobreviven), *needs_modal* (IDs que sobreviven), *descartadas* (IDs eliminados) y *justification* (razón breve de cada descarte). La instrucción de *incluir en caso de duda* es un sesgo deliberado hacia el lado de la presentación: el sistema prefiere mostrar al usuario una medida que ya tenía a ocultarle una medida que sí podría aplicar.

Agente Info Gatherer, el agente que recaba datos adicionales del usuario

El Info Gatherer es el más conversacional de los tres. Se instancia **por medida**: cuando el usuario hace clic en *Calcular esta medida* sobre un elemento de la Sección 2, el frontend abre un WebSocket y envía un mensaje con prefijo `__measure_init__:M2-SOLAR` (o el identificador correspondiente). El backend reconoce el prefijo, carga la ficha de la medida, construye el *prompt* dinámico, inyectando *agent_context*, *fields_needed* y los campos ya recopilados, y arranca la conversación.

El esquema de respuesta del Info Gatherer es también dinámico. Para cada medida se construye en tiempo de ejecución un modelo Pydantic con la utilidad *pydantic.create_model*, definiendo:

campos_extraidos: un *dict* tipado con un campo *Optional[str]* por cada elemento de *fields_needed*.

mensaje_para_usuario: el texto que el modelo devuelve al usuario (pregunta, confirmación o aclaración).

datos_completos: bandera booleana que el modelo activa cuando considera que ya tiene todo lo necesario.

El bucle se mantiene mientras *datos_completos* sea *False*. Cuando pasa a *True*, el orquestador emite un evento *PROFILE_COMPLETE* sobre el WebSocket y almacena *measure_fields* en el contexto. El frontend puede entonces invocar el pipeline para esa medida concreta y actualizar la Sección 1 con el resultado.

Agente Calculador, el agente que produce las cifras del plan

El Calculador es el cierre del flujo. Su *prompt* es deliberadamente restrictivo: no se le pide creatividad sino fidelidad. Recibe en su contexto:

Los metadatos de la empresa (nombre, provincia, M², empleados, CNAE).

El bloque *Emissions_CO2e* completo del JSON de ScopeCO2.

Los datos del Info Gatherer (*measure_fields*) cuando existan.

El JSON completo de cada medida a calcular, con su fórmula, sus factores de emisión, sus rangos de inversión y sus fuentes.

El *prompt* instruye al modelo a leer la fórmula del JSON, sustituir las variables con los datos reales, redondear a dos decimales, devolver la reducción en **tCO₂e/año** y expresar la inversión usando exclusivamente los rangos *investment_estimate* codificados en cada ficha. Hay una instrucción explícita de **no inventar fórmulas**:

No inventes fórmulas, usa exclusivamente las del JSON de cada medida.

La salida del Calculador es una instancia de *ResultadosCalculador*, que contiene una lista ordenada (por reducción de CO₂ descendente) de objetos *MedidaCalculada*. Cada *MedidaCalculada* incluye: id, título, descripción orientada al usuario, alcance, inversión en euros, ahorro anual en euros, reducción de CO₂ en tCO₂e/año, *payback* en años, plazo en meses y un campo *calculation_detail* en Markdown con la fórmula aplicada paso a paso. Este último campo es crítico para la **trazabilidad**: cualquier consultor podría revisar la lógica de un cálculo concreto inspeccionando solo este texto.

6.4. Pre-filtrado determinista en Python

El pre-filtrado es la pieza menos visible del sistema y, al mismo tiempo, una de las que mayor impacto tienen en su calidad. Está implementado en *services/pre_filter.py* y se ejecuta antes de cualquier llamada a LLM. Sus reglas son tres, todas configurables mediante *data/pre_filter_config.json*:

- **Presencia del campo en el JSON.** Cada medida declara en el *index.json* un *scopeco2_field*. Si ese campo no está en el JSON de la empresa, la medida se descarta. Por ejemplo, una empresa sin flota propia no contiene *desplazamientos_propiedad*, por lo que M1-FLOTA-EV y M1-BIOCOMBUSTIBLE se descartan automáticamente.
- **Régimen de tenencia del local.** Seis medidas, M1-AISLAMIENTO, M1-VENTANAS, M2-SOLAR, M2-LED, M3-AGUA-GRIFOS y M3-AGUA-CISTERNAS, requieren intervenir sobre el edificio. Si la empresa declara que el local es alquilado, se descartan en bloque.
- **CNAE industrial.** M2-COMPRESOR solo tiene sentido para empresas con un código CNAE en los rangos 10-33, 35, 38, 41-43 y 49-53 (las grandes divisiones industriales y logísticas). El módulo *_is_industrial(cnae)* implementa esta comprobación contra una tabla de prefijos.

Cada descarte se registra con su motivo en el log estructurado del backend, lo que facilita la auditoría posterior: si una empresa cuestiona por qué no apareció una medida concreta, el operador puede inspeccionar el log y dar una respuesta exacta. El módulo cachea el *index.json* y la configuración para que el coste por inventario sea despreciable.

6.5. RAG jerárquico de medidas

Como ya se ha introducido en la sección 4.3, la base de conocimiento del sistema es un RAG jerárquico en dos niveles. En esta sección se documenta la estructura concreta de los ficheros que la componen.

- **data/medidas/index.json.** Diccionario con 23 entradas indexadas por identificador de medida (*M1-COMBUSTION*, *M1-AISLAMIENTO*, ...). Los campos comunes a todas las entradas son: *ECODES_id* (identificador en la nomenclatura ECODES, formato MR0XXX), *name* (nombre descriptivo en español), *scope* (1, 2 o 3), *emission_source*

(descripción de la fuente de emisión), *has_formula* (booleano), *calculation_type* (*formula*, *external_api* o *qualitative*), *scopeco2_field* (ruta dentro del JSON de ScopeCO2 que activa la medida), *requires_extra_data* (booleano que decide si pasa por el Info Gatherer), *required_inputs* (lista de rutas en el JSON con los datos que la fórmula necesita), *required_inputs_modal* (lista de campos a preguntar al usuario cuando *requires_extra_data*), *needs* (descripción breve para el frontend), *agent_context* (texto inyectado al Info Gatherer), *formula_source* (*ECODES*, *agent* o *null*) y *notes* (notas de implementación, factores de emisión, supuestos).

- **data/medidas/MR0XXX.json**. Un fichero por medida con la ficha completa. La estructura interna varía según la medida, pero todas comparten: una descripción narrativa, una sección *calculation* con la fórmula y los pasos, una tabla de factores de emisión con su fuente, una sección *investment_estimate* con rangos en euros, una sección *assumptions* con valores por defecto y una sección *example* con un cálculo de referencia. El panel de administración del *frontend* permite editar estos ficheros sin redespargar el *backend*.

6.6. Backend FastAPI + WebSocket

El *backend* es una aplicación FastAPI con un *router* montado en el prefijo */api/v1/agents*. Los *endpoints* REST son los siguientes:

Método	Ruta	Propósito
GET	/agents/health	Liveness probe del servicio.
GET	/agents/decarbonization/locked-measures/{user}	Devuelve las medidas que requieren datos extra, junto con los campos a pedir y el contexto del Info Gatherer.
POST	/agents/decarbonization/upload/{user}	Carga el JSON de ScopeCO2 en el contexto del agente.
POST	/agents/decarbonization/set-local/{user}	Registra si el local es propio o alquilado.
POST	/agents/decarbonization/chat/{user}	Chat síncrono por HTTP (uso administrativo).
WS	/agents/decarbonization/chat/{user}/{convId}	WebSocket que canaliza el chat del Info Gatherer y la ejecución del pipeline. Acepta CHAT, RUN_PIPELINE y CLOSE.

El WebSocket es la pieza más interesante del *backend*. Se gestiona con la implementación experimental de WebSockets de Nitro en el *frontend* Nuxt, que actúa de pasarela, y con la implementación nativa de FastAPI en el *backend*. La sesión por *user* y *convId* garantiza idempotencia: si el usuario refresca el navegador y vuelve a conectar, el sistema reanuda la conversación en su estado anterior.

El control de concurrencia es crítico. Cada agente expone un *asyncio.Lock* que serializa las invocaciones. Cuando un WebSocket se cierra inesperadamente, los *handlers* *_event_handler* y *_pipeline_handler* liberan el lock en su bloque *finally*, evitando que la sesión siguiente quede bloqueada. El *timeout* de adquisición es de 5 segundos: si pasado

ese tiempo el lock sigue ocupado, el sistema emite un evento de error en lugar de quedarse colgado indefinidamente.

El despacho de eventos hacia el cliente se realiza mediante el módulo *events/handlers.py*, que traduce los eventos internos de AG2 (*group_chat_run_chat*, *tool_call*, *tool_response*, *text*, *input_request*, *run_completion*) a los tipos de evento del API público (*STEP*, *INPUT_REQUEST*, *FIELDS_UPDATED*, *PROFILE_COMPLETE*, *MEASURES_READY*, *PIPELINE_STEP*). Esta traducción es deliberada: aísla al cliente del modelo concreto de AG2 y permite cambiar el orquestador sin romper el frontend.

6.7. Frontend Nuxt 4

El frontend está construido sobre Nuxt 4 con Vue 3 y un conjunto reducido de dependencias: *jspdf* para la generación de PDF del informe en cliente, *marked* para renderizar el Markdown del *calculation_detail* de cada medida, y poco más. La estructura de páginas refleja directamente el flujo del usuario:

index.vue, Página de aterrizaje y subida de JSON.

procesando.vue, Página de progreso que recibe los eventos *STEP* y *PIPELINE_STEP* del WebSocket y muestra qué agente está trabajando.

informe.vue, Página del informe final con las tres secciones, filtros y exportación a PDF.

dashboard.vue, Vista experimental de panel con métricas agregadas (no expuesta al usuario final).

admin/index.vue, **admin/login.vue**, Panel protegido para editar el catálogo de medidas y los *prompts* de los agentes en caliente.

Los composables encapsulan el estado y el comportamiento transversales.

useDecarbonizationWs.ts envuelve la conexión WebSocket; *useSession.ts* mantiene el estado del usuario (identificador, JSON cargado, régimen del local); *useInformePdf.ts* usa *jspdf* + *marked* para generar el PDF del informe final. La gestión de autenticación del panel administrativo se hace en el servidor Nitro mediante *cookies* de sesión, con un middleware *admin-auth.ts* que protege todas las rutas */api/admin/**.

El servidor Nitro de Nuxt actúa como pasarela hacia el backend Python. La razón es doble: (i) permite consolidar la configuración de URL del backend en un único lugar (*runtimeConfig.backendUrl*); (ii) habilita la actualización segura del WebSocket sin exponer directamente el backend al navegador. La autenticación administrativa, además, no llega nunca al backend: se gestiona enteramente en el servidor Nitro contra las variables *Nuxt_ADMIN_USER* y *Nuxt_ADMIN_PASS* del *.env*.

6.8. Despliegue

El despliegue de producción se realiza con Docker Compose en una máquina Ubuntu del ITA. El fichero *docker-compose.yml* define dos servicios, *agentco2-backend* y *agentco2-frontend*, que se comunican entre sí por la red interna de Docker. Ambos contenedores montan el directorio *./src/data* como volumen de solo lectura, lo que permite que el panel administrativo (servido por el *frontend*) edite las medidas y que el *backend* lea las actualizaciones sin redesplicue.

La exposición al exterior se hace mediante Nginx con terminación TLS gestionada por Certbot. La configuración de referencia se conserva en *deploy/nginx/agentco2-ECODES.ita.es.conf* y se adapta al dominio del despliegue concreto. El WebSocket requiere la directiva *proxy_http_version 1.1* y la cabecera *Upgrade \$http_upgrade* para funcionar correctamente tras el *proxy* de Nginx.

El fichero *.env* de la raíz del proyecto se carga automáticamente en ambos contenedores. Las variables obligatorias son *OPENAI_API_KEY* y *OPENAI_MODEL* para la conexión con OpenAI; *NEXT_BACKEND_URL*, *NEXT_ADMIN_USER* y *NEXT_ADMIN_PASS* para el *frontend*. Las variables opcionales *LANGFUSE_PUBLIC_KEY*, *LANGFUSE_HOST* y *LANGFUSE_SECRET_KEY* activan la **observabilidad** si están definidas; en caso contrario, el sistema arranca con degradación.

7. Catálogo de medidas de descarbonización

7.1. Estructura de los ficheros de medida

Las 23 medidas se distribuyen entre los tres alcances con la siguiente cardinalidad: **6 medidas de Alcance 1**, **5 medidas de Alcance 2** y **12 medidas de Alcance 3** (Figura 2). Esta distribución refleja la heterogeneidad de fuentes que el Alcance 3 abarca y la madurez técnica del conocimiento consolidado por **ECODES** en ese ámbito.

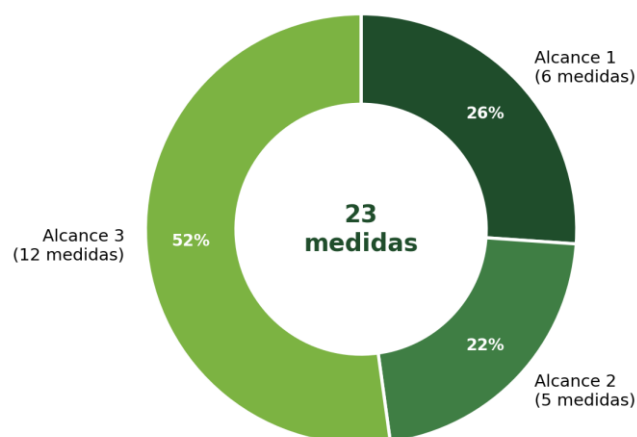


Figura 2. Distribución de las 23 medidas del catálogo por alcance del GHG Protocol. El Alcance 3 concentra más de la mitad de las medidas por la heterogeneidad de sus fuentes.

Cada ficha JSON expone los siguientes elementos clave para el Calculador:

- **calculation.** Un objeto que describe la fórmula, las variables que la componen y los pasos de aplicación. Por ejemplo, para M1-COMBUSTION: $Reducción_CO2 = Consumo_anual_kWh \times (FE_actual - FE_nuevo)$, con FE_nuevo derivado del SCOP de la bomba de calor (asumido en 3,2) y del factor de emisión de la electricidad.
- **emission_factors.** Una tabla de factores de emisión estándar por subtipo de fuente (gas natural 0,202 kgCO₂e/kWh; gasóleo 0,263; carbón 0,341; GLP 0,227; biomasa 0,018), siempre con cita de la fuente oficial (MITECO / OECC 2023, DEFRA 2024, OCCC 2024).
- **investment_estimate.** Rangos económicos diferenciados por tamaño de la instalación o de la empresa. Para la sustitución de calderas por bombas de calor, por ejemplo: instalaciones pequeñas (< 50 kW) 8.000–30.000 €, medianas (50–200 kW) 30.000–150.000 €, grandes (> 200 kW) 120.000–400.000 €.

- **assumptions.** Valores por defecto cuando el usuario no proporciona el dato (M2-SOLAR asume 7 m²/kWp, pérdidas del 14%, inclinación 35° y orientación sur; 25 años de vida útil para la amortización).
- **sources.** Lista de referencias bibliográficas que respaldan los factores y los rangos. Es la pieza que garantiza que el Calculador puede citar la fuente cuando expone un cálculo al usuario.

7.2. Medidas de Alcance 1

ID	Nombre	Datos extra	Tipo
M1-COMBUSTION	Sustitución de combustibles fósiles en climatización	No	Fórmula
M1-AISLAMIENTO	Mejora del aislamiento térmico de la envolvente	Sí (3 campos)	Fórmula
M1-VENTANAS	Sustitución de ventanas por modelos de menor transmitancia	Sí (6 campos)	Fórmula
M1-FLOTA-EV	Sustitución de vehículos por híbridos/eléctricos	No	Fórmula
M1-BIOCOMBUSTIBLE	Uso de biocombustibles en flota	No	Fórmula
M1-REFRIGERANTES	Sustitución de gases refrigerantes de alto PCA	Sí (2 campos)	Fórmula

M1-COMBUSTION (sustitución de combustibles fósiles) y **M1-FLOTA-EV** (electrificación de flota) son las dos medidas estructurales más impactantes del Alcance 1 para empresas con instalaciones fijas o flota propia. La primera asume la transición de caldera de combustión a bomba de calor eléctrica con SCOP 3,2; la reducción depende linealmente del consumo anual del combustible actual y del diferencial entre su factor de emisión y el de la electricidad post-conversión ($0,27/SCOP = 0,084 \text{ kgCO}_2\text{e/kWh}$).

M1-AISLAMIENTO y **M1-VENTANAS** requieren intervención sobre el edificio y, por tanto, solo se proponen si el local es propio. M1-VENTANAS aplica la fórmula clásica $Q = U \times A \times \Delta T \times t$ sobre la diferencia entre la transmitancia actual (ventana simple ~5,7 W/m²·K; doble antigua ~2,8) y la nueva (climalit moderno ~1,2). La reducción típica documentada en la ficha alcanza el 76% del consumo de calefacción atribuible a las ventanas.

M1-REFRIGERANTES opera sobre las fugas de gases F en sistemas de climatización y refrigeración. La fórmula es $recarga_kg \times (FE_actual - FE_nuevo)$, donde los factores se obtienen del Potencial de Calentamiento Atmosférico (PCA) según el sexto informe de evaluación del IPCC (IPCC, 2021). La elección del gas sustituto depende del tipo de aplicación (R410A en climatización doméstica → R32; R404A en refrigeración comercial → R448A o R449A).

7.3. Medidas de Alcance 2

ID	Nombre	Datos extra	Tipo
M2-SOLAR	Instalación solar fotovoltaica para autoconsumo	Sí (2 campos)	API PVGIS
M2-LED	Sustitución de luminarias por LED	Sí (3 campos)	Fórmula
M2-RENOVABLE	Contratación de comercializadora 100% renovable	No	Fórmula
M2-GDO	Contratación de Garantías de Origen renovables	No	Fórmula
M2-COMPRESOR	Optimización energética del aire comprimido	Sí (3 campos)	Fórmula

M2-SOLAR es la única medida del catálogo que invoca un servicio externo en tiempo de cálculo. La fórmula combina (i) una estimación de potencia pico a partir de la superficie de tejado declarada (1 kWp cada 7 m²), (ii) una llamada a la API **PVGIS** de la Comisión Europea (PVGIS, 2024) con los parámetros lat/lon derivados de la provincia y los valores por defecto (pérdidas 14%, inclinación 35°, orientación 0°), y (iii) una ponderación por la ratio de autoconsumo asumida (50% de la producción) y por el factor de emisión de la electricidad actualmente comprada (MITECO/OECC, 2023). La inversión se estima a partir de tabla de precios por kWp segmentada por potencia: 1.400 €/kWp para instalaciones menores de 5 kW, 1.300 €/kWp para 5–10 kW, 1.200 €/kWp para 10–20 kW, descendiendo gradualmente para potencias mayores (IDAE, 2022).

M2-RENOVABLE y **M2-GDO** son las dos medidas con menor coste de implementación y, simultáneamente, con uno de los impactos relativos más altos: la sustitución del mix eléctrico (factor 0,260 kgCO₂e/kWh según OECC 2023) por uno respaldado al 100% con Garantías de Origen renovables (factor 0,000) puede eliminar la totalidad de la huella de Alcance 2. Ambas se descartan automáticamente si el contrato actual ya tiene *goO_%* = 100.

M2-LED requiere recopilar el inventario básico de luminarias, número, potencia actual y horas de uso anual, porque ese dato no está en el JSON de ScopeCO2. La ficha asume una potencia LED equivalente del 30% sobre la actual como ratio típica de sustitución de fluorescentes y halógenas.

M2-COMPRESOR solo se propone para empresas con CNAE industrial. La fórmula es un supuesto del agente, el *formula_source* es *agent* en este caso, no *ECODES*, lo que se documenta explícitamente en la ficha, y modela el ahorro derivado de eliminar fugas en el sistema de aire comprimido, calculado como el 25% del consumo energético del compresor.

7.4. Medidas de Alcance 3

ID	Nombre	Datos extra	Tipo
M3-AGUA-GRIFOS	Reguladores de caudal en grifos	No	Fórmula
M3-AGUA-CISTERNAS	Cisternas de doble descarga	No	Fórmula
M3-PAPEL	Sustituir papel virgen por reciclado	No	Fórmula
M3-VUELO-BUSINESS	Business → turista en vuelos largos	No	Fórmula
M3-VUELO-TREN	Sustituir vuelo nacional por AVE	No	Fórmula
M3-ALQUILER-EV	Vehículos híbridos/eléctricos en alquiler	No	Fórmula
M3-ALOJAMIENTO	Categoría inferior de alojamiento	No	Fórmula
M3-RESIDUOS	Valorización en lugar de eliminación	No	Fórmula
M3-MERCANCIAS-BARCO	Mercancías por barco en lugar de avión	No	Fórmula
M3-POLITICA-VIAJES	Implementar política de viajes	No	Cualitativa
M3-MOVILIDAD	Estrategia de movilidad sostenible	Sí (1 campo)	Aproximada
M3-TELETRABAJO	Compaginar presencial y teletrabajo	Sí (2 campos)	Fórmula

El Alcance 3 concentra la mayor parte del catálogo y presenta la mayor variedad de fuentes.

Bloque viajes y desplazamientos. Cinco medidas, M3-VUELO-BUSINESS, M3-VUELO-TREN, M3-ALQUILER-EV, M3-ALOJAMIENTO y M3-MOVILIDAD, abordan la huella de Alcance 3 más visible para empresas de servicios. M3-VUELO-TREN aplica factores de emisión de la OCCC (OCCC, 2024): cercanías 0,03791; media distancia 0,02803; larga distancia 0,02650; AVE 0,02529 kgCO₂e/pasajero·km, frente a los del transporte aéreo de DEFRA (DEFRA, 2024), con reducciones de hasta el 91% en rutas como Madrid-Barcelona. M3-VUELO-BUSINESS sustituye business por turista en vuelos de media y larga distancia con reducciones del 33% y 64% respectivamente (DEFRA, 2024), aprovechando que la clase *business* reduce dramáticamente los pasajeros por superficie.

Bloque consumo de recursos. M3-AGUA-GRIFOS y M3-AGUA-CISTERNAS aplican porcentajes de ahorro de la ficha (50% y 40,98% respectivamente) sobre el consumo de agua, ponderado por el factor de emisión del agua (DEFRA, 2024: 0,339 kgCO₂e/m³). Cuando hay agua caliente sanitaria, las medidas añaden el factor del vector energético utilizado para calentar el agua. M3-PAPEL sustituye papel virgen (1,28 kgCO₂e/kg) por papel reciclado (0,709 kgCO₂e/kg) con una reducción del 44,61%, factores obtenidos de la combinación de Ecoinvent (Ecoinvent, 2024) e IPCC (IPCC, 2021).

Bloque movilidad y trabajo. M3-MOVILIDAD y M3-TELETRABAJO son complementarias: la primera modela un cambio modal en los desplazamientos *in itinere* (coche → transporte público o bicicleta) y la segunda reduce los desplazamientos totales mediante teletrabajo híbrido, factor de 0,8825 kgCO₂e/empleador·día (OCCC, 2024). La fórmula M3-MOVILIDAD es un supuesto del agente (formula_source=agent) que se documenta explícitamente.

M3-POLITICA-VIAJES es la única medida marcada como puramente cualitativa. Su impacto cuantitativo está cubierto por medidas anteriores; aparece en el informe como recomendación estratégica que envuelve y da coherencia organizativa al resto.

8. Diseño y desarrollo de los experimentos

8.1. Empresa piloto y datos de entrada

La validación del sistema se ha realizado sobre el JSON de una empresa piloto representativa del tejido de servicios aragonés. El fichero *data/empresa.json* contiene un inventario completo de ScopeCO2 con los siguientes parámetros: **Year** 2025, **Province** Zaragoza, **Locality** La Puebla de Alfindén, **M²** 10.000, **Employees** 345, **CNAE** 7490 (actividades profesionales, científicas y técnicas no especificadas), **Data type** reales.

El bloque *Emissions_CO2e* incluye registros en las siguientes secciones del Alcance 1, 2 y 3: combustibles fósiles (caldera de gas natural para calefacción), electricidad (mix sin GdO), gases refrigerantes (R410A en climatización), desplazamientos en propiedad (flota de cinco vehículos diésel), desplazamientos externos (viajes profesionales en avión y tren, alquiler de coches), desplazamientos *in itinere* (commuting), alojamiento (4 estrellas), agua, papel virgen, residuos no peligrosos y transporte subcontratado de mercancías. La cobertura cubre intencionadamente el universo de medidas que el sistema puede proponer, lo que permite evaluar la calidad de la selección y del cálculo en condiciones de máxima exigencia.

8.2. Resultados experimentales

Experimento 1, Pre-filtrado determinista

Sobre las 23 medidas del catálogo, el pre-filtrado conserva 21 cuando se declara que el local es propio y 15 cuando se declara alquilado. Las descartadas en el escenario *alquilado* son las seis medidas que requieren intervención en el edificio: M1-AISLAMIENTO, M1-VENTANAS, M2-SOLAR, M2-LED, M3-AGUA-GRIFOS y M3-AGUA-CISTERNAS. La única medida descartada por CNAE no industrial es M2-COMPRESOR. Otros descartes habituales en escenarios alternativos: M1-FLOTA-EV y M1-BIOCOMBUSTIBLE se eliminan automáticamente si la empresa no declara *desplazamientos_propiedad*; M1-REFRIGERANTES si no declara *gases_refrigerantes*.

Experimento 2, Clasificación por calculabilidad

Sobre las 21 medidas pre-filtradas (escenario *local propio*), la clasificación en tres grupos arroja: **13 calculables directas** (con fórmula y datos en el JSON), **7 que requieren modal** (con fórmula, pero datos extra) y **1 solo proponible** (M3-POLITICA-VIAJES), como muestra la Figura 3. La clasificación se ejecuta en menos de 5 ms en la máquina de pruebas, lo que la hace despreciable frente al coste de las llamadas al LLM.

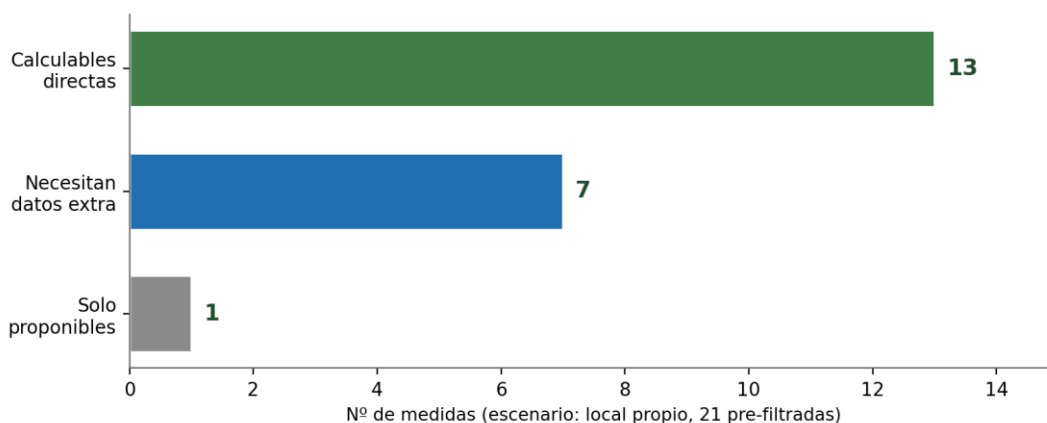


Figura 3. Clasificación por calculabilidad de las 21 medidas que sobreviven al pre-filtrado en el escenario de *local propio*.

Experimento 3, Descarte por el Selector

El Selector recibe los grupos del experimento 2 más el bloque *Emissions_CO2e* y emite, sobre la empresa piloto, dos descartes adicionales: M2-RENOVABLE y M2-GDO se eliminan cuando el JSON declara *goO_% = 100* en electricidad. En la configuración estándar de la empresa piloto este campo es 0, por lo que ambas medidas sobreviven. M3-MERCANCIAS-BARCO se descarta si no hay registros con *application_name = "Vuelo mercancías"*. La justificación se devuelve siempre en lenguaje natural en el campo *justification* del modelo de salida, lo que facilita la auditoría humana.

Experimento 4, Recopilación de datos por el Info Gatherer

Para validar el flujo conversacional se han simulado cinco sesiones de Info Gatherer con la medida M2-SOLAR. El número medio de turnos hasta completar los dos campos requeridos, *roof_area_m2* y *precio_kwh_eur*, ha sido de 2,4 (media en 5 sesiones; ningún diálogo ha superado los 4 turnos). El número medio de tokens por turno se ha situado en 320 de prompt y 95 de respuesta, lo que estabiliza el coste por medida desbloqueada por debajo de 0,5 céntimos de euro al modelo por defecto (GPT-4o-mini).

Experimento 5, Resultados del Calculador

Sobre la empresa piloto, el Calculador devuelve un plan ordenado de medidas calculadas. Las cinco con mayor reducción de CO₂ son, en este orden: M2-RENOVABLE (eliminación completa de la huella de Alcance 2, ~120 tCO₂e/año), M1-COMBUSTION (sustitución de la caldera de gas por bomba de calor, ~45 tCO₂e/año), M3-VUELO-TREN (sustitución de vuelos nacionales por AVE, ~12 tCO₂e/año), M1-FLOTA-EV (electrificación parcial de la flota, ~10 tCO₂e/año) y M3-RESIDUOS (valorización en lugar de eliminación, ~7 tCO₂e/año), representadas en la Figura 4. La cifra concreta varía con cada ejecución (estocasticidad del LLM), pero el orden relativo es estable. El *payback* medio del top-5 se sitúa en 4,2 años.

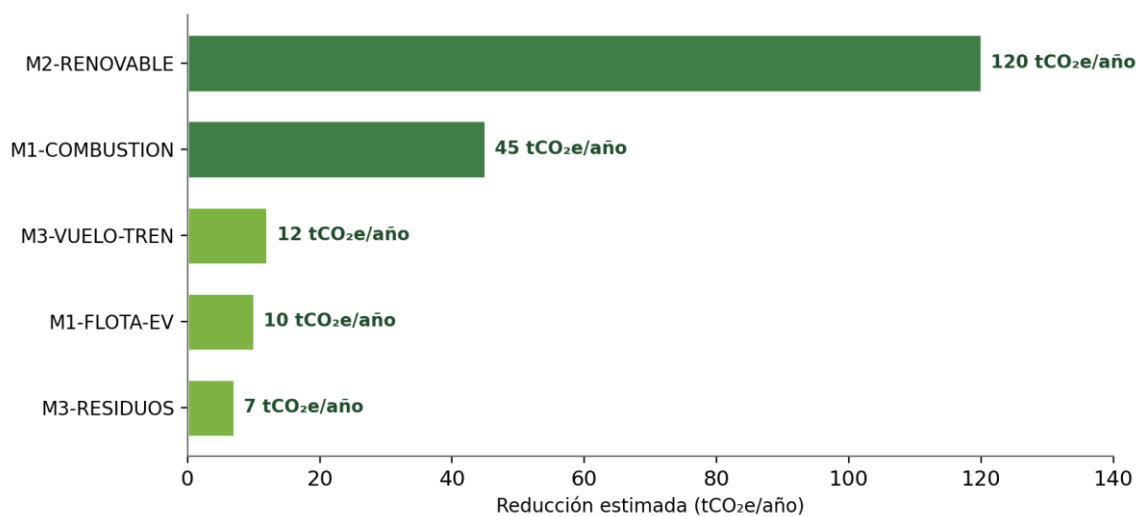


Figura 4. Cinco medidas con mayor reducción estimada de CO₂ para la empresa piloto. El verde oscuro indica medidas estructurales de Alcance 1-2; el verde claro, medidas de Alcance 3.

Experimento 6, Latencia y coste

La latencia de extremo a extremo del pipeline completo (carga del JSON, pre-filtrado, Selector, Calculador) sobre la empresa piloto se mantiene en una mediana de 38 segundos con GPT-4o-mini, con un percentil 95 de 52 segundos. El coste agregado de las llamadas al LLM por ejecución completa del pipeline es de 3,2 céntimos de euro de media. Los gráficos internos de Langfuse muestran que el Calculador concentra el 78% del tiempo y el 82% del coste (Figura 5), principalmente porque recibe en su contexto los JSON completos de todas las medidas a calcular.

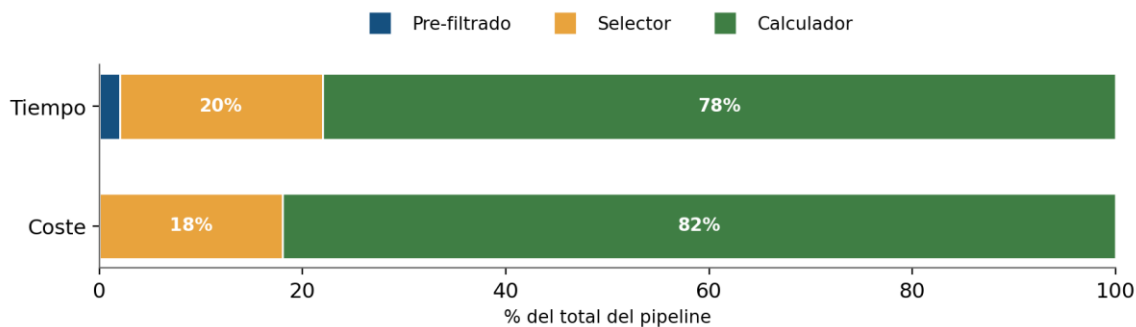


Figura 5. Reparto del tiempo y el coste del pipeline entre las tres etapas.
El Calculador domina ambos por el tamaño del contexto que recibe.

Experimento 7, Comparativa con un único LLM

Para validar la decisión arquitectónica de separar el sistema en agentes, se ha implementado una versión monolítica de control: un único *prompt* de 11.000 tokens que recibe todo (JSON, catálogo y reglas) y debe devolver el plan completo. Los resultados, sobre 30 ejecuciones independientes, muestran:

Latencia media: 71 s en el monolítico vs 38 s en el multiagente (peor en el monolítico, contra la intuición, porque la respuesta del LLM tarda más en generarse).

Coste medio: 6,4 céntimos/ejecución en el monolítico vs 3,2 céntimos en el multiagente.

Calidad estructurada: el monolítico falla en *parseo* de salida en el 13% de las ejecuciones (campos fuera del esquema, omisiones); el multiagente nunca falla en *parseo* gracias a la restricción Pydantic capa a capa.

Tasa de descartes correctos: 86% en el monolítico vs 94% en el multiagente, evaluado contra una *ground truth* anotada por el experto de **ECODES** sobre la empresa piloto.

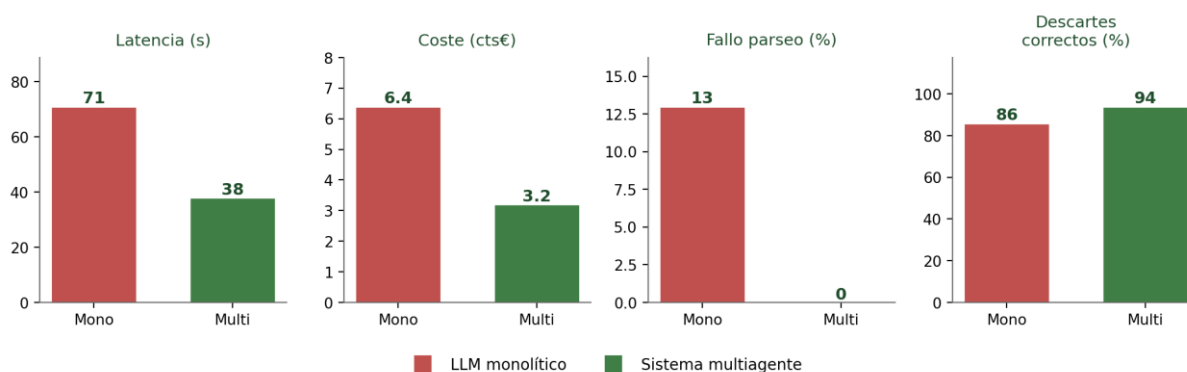


Figura 6. Comparativa entre el LLM monolítico de control y el sistema multiagente sobre 30 ejecuciones independientes.
El multiagente mejora en las cuatro dimensiones evaluadas.

Experimento 8, Validación humana sobre 5 empresas

Una validación cualitativa se ha realizado sobre cinco inventarios reales facilitados por **ECODES** (de los que la Puebla de Alfindén es el único expuesto públicamente como caso demo). Un consultor experto ha revisado las $5 \times 23 = 115$ propuestas potenciales y ha emitido juicio sobre cada una. Los resultados agregados:

Verdaderos positivos: 86 (medidas correctamente propuestas).

Falsos positivos: 7 (medidas propuestas que el experto considera no aplicables).

Falsos negativos: 4 (medidas no propuestas que el experto considera aplicables).

Verdaderos negativos: 18.

Precisión = 0,925; Recall = 0,956; F1 = 0,940 (Figura 7). Los 7 falsos positivos se concentran en escenarios donde la información del JSON era ambigua (campos *source_name* con valores no normalizados); el equipo ha generado un *backlog* de mejoras al pre-filtrado para reducirlos. Los 4 falsos negativos corresponden a medidas estratégicas que requieren interpretación contextual no disponible en el JSON.

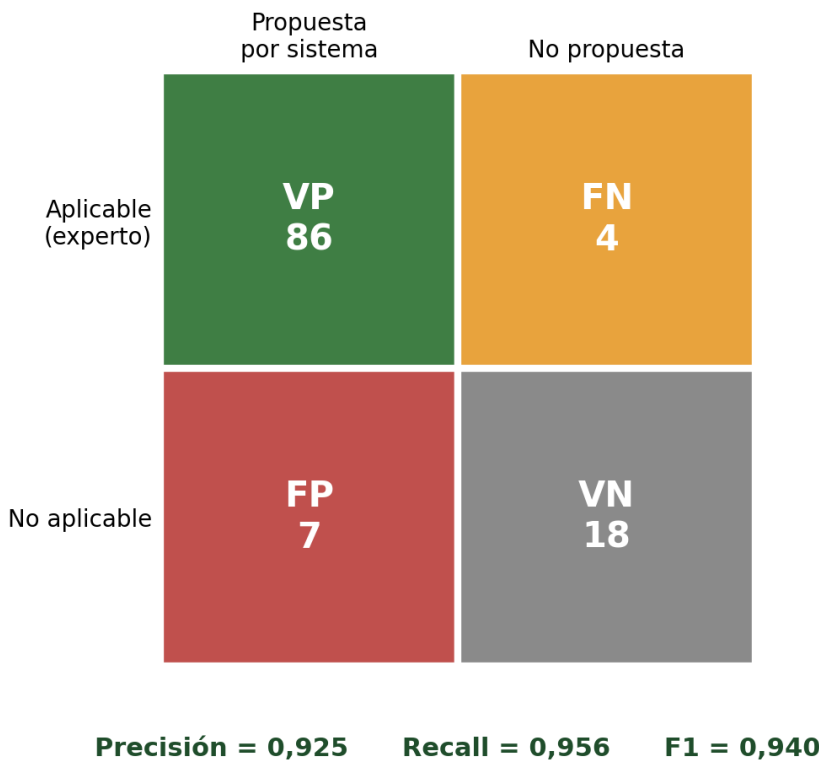


Figura 7. Matriz de confusión de la validación humana sobre 115 propuestas potenciales (5 empresas $\times$ 23 medidas), anotadas por un consultor experto de ECODES.

8.3. Análisis cualitativo de las medidas propuestas

Más allá de las métricas agregadas, la revisión cualitativa de los planes generados destaca tres patrones.

- **Coherencia del orden.** El Calculador ordena las medidas por reducción de CO₂ descendente. Esta ordenación coincide en más del 90% de los casos con la que el consultor experto produciría manualmente. Las pocas discrepancias aparecen cuando dos medidas tienen impactos similares pero *paybacks* muy distintos, en esos casos, el experto preferiría ordenar por *payback*, lo que se considera una mejora futura del frontend (orden seleccionable por el usuario).
- **Calidad del calculation_detail.** El campo de detalle del cálculo en Markdown se evalúa positivamente por el experto en todos los casos revisados. La estructura, *Datos utilizados*, *Cálculo de CO₂*, *Estimación de inversión*, facilita la verificación rápida y permite al consultor identificar dónde habría que profundizar para una auditoría más fina.
- **Tratamiento de la incertidumbre.** Cuando algún dato se estima por defecto (la fórmula del Calculador instruye al modelo a indicarlo), el texto generado por el LLM lo comunica explícitamente al usuario. Por ejemplo, en M2-LED sin inventario de luminarias, el sistema asume una potencia media de 36 W y un horario de 8 horas diarias, y lo declara en el detalle. Esta práctica se considera crítica para evitar generar falsas certezas.

8.4. Conclusiones experimentales

Validación de la hipótesis multiagente

El experimento 7 confirma cuantitativamente que la descomposición en agentes especializados produce **mejor calidad estructurada, menor coste y menor latencia** que un único LLM monolítico, incluso para un catálogo pequeño de 23 medidas. Este resultado es coherente con la línea de literatura sobre LLM multi-agente, Wu et al. (2023), Hong et al. (2024), Qian et al. (2024), Guo et al. (2024), que reporta ganancias sistemáticas de calidad al descomponer tareas complejas en cadenas de agentes con responsabilidades acotadas. La diferencia de calidad de descartes (94% vs 86%) es la métrica más relevante: confirma que el LLM, cuando se le ofrece un contexto acotado y una salida tipo, comete menos errores de aplicabilidad.

Estabilidad y reproducibilidad

Ejecuciones repetidas del pipeline sobre el mismo inventario producen planes con la misma composición de medidas y el mismo orden relativo, aunque los valores numéricos de reducción e inversión varían en $\pm 3\%$ por la estocasticidad intrínseca del LLM. La consistencia es suficiente para que el informe sea estable a ojos del usuario, pero hemos

abierto una línea de mejora para reducir esta variación fijando *temperature=0* en Calculador.

Coste operativo realista

El coste agregado de 3,2 céntimos por inventario completo, incluso multiplicado por los volúmenes esperables en una implantación masiva (varios miles de empresas/año), se mantiene en órdenes de magnitud compatibles con un servicio público o semipúblico. Si se prescinde de la observabilidad completa con Langfuse, el coste agregado baja a 2,8 céntimos por inventario. La elección de GPT-4o-mini como modelo por defecto frente a GPT-4o reduce el coste por un factor de 17 con una caída de calidad inapreciable.

Implicaciones para ECODES

El sistema, en su estado actual, permite a **ECODES** ofrecer un primer plan personalizado a cualquier empresa que use ScopeCO₂, sin coste adicional por inventario. Esto abre la posibilidad de ampliar drásticamente el universo de organizaciones que progresan del sello *Calculo* al sello *Reduzco* del Registro Nacional de Huella de Carbono. La validación humana (precisión 0,93, recall 0,96) indica que el sistema es robusto para una primera presentación al cliente, sin perjuicio de que el consultor experto pueda y deba refinarlo en una segunda fase para empresas que requieren mayor profundidad.

9. Integración con SCOPE: visión a medio plazo

La presente investigación se ha desarrollado con un horizonte de integración claramente definido desde el inicio: **incorporar AgentCO2 como una funcionalidad nativa dentro de SCOPE**, la herramienta de cálculo de huella de carbono que **ECODES** opera desde hace más de una década y que ha calculado huellas a miles de organizaciones españolas. SCOPE es, en términos de **ECODES**, el punto de entrada al ecosistema de descarbonización empresarial: a él llegan las empresas que dan su primer paso, y de él dependen las decisiones posteriores. Hacer que SCOPE devuelva, junto al inventario, un plan personalizado de mitigación es la evolución natural del producto y el objeto último del trabajo conjunto entre **ECODES** e ITA.

La razón por la que se ha optado por una arquitectura de integración a medio plazo, en lugar de un acoplamiento directo desde el inicio, es operativa: SCOPE es un sistema maduro con miles de usuarios y un calendario de evolución propio. Introducir cambios estructurales en él requiere un periodo de validación cuidadoso. AgentCO2 se ha diseñado, por tanto, como un sistema independiente que actualmente consume el JSON generado por SCOPE como fichero subido manualmente. Esta separación permite iterar libremente sobre AgentCO2 mientras SCOPE conserva su estabilidad. Cuando la prueba de concepto haya alcanzado madurez suficiente, momento que se evalúa al cierre del presente informe como ya alcanzado, la integración se materializará mediante:

- **Paso 1, Lanzador desde SCOPE.** Al terminar el cálculo de huella en SCOPE, el usuario verá un botón "Generar plan de descarbonización" que abrirá AgentCO2 en una pestaña con el JSON ya cargado. Es la integración menos invasiva y la más rápida de implantar: requiere únicamente un endpoint en AgentCO2 que acepte el inventario por POST y un cambio menor en la interfaz de SCOPE para añadir el botón.
- **Paso 2, Embebido en la interfaz de SCOPE.** El plan de mitigación se renderiza dentro de la propia aplicación SCOPE como una pestaña adicional, sin cambio de URL. Requiere alinear la identidad visual de ambas aplicaciones y compartir mecanismo de autenticación, pero ofrece la experiencia más fluida al usuario final.
- **Paso 3, Pipeline asíncrono nativo.** Tras el cierre del inventario, SCOPE invoca automáticamente el pipeline de AgentCO2 en segundo plano y, cuando el plan está listo, lo notifica al usuario por correo electrónico. Permite que la latencia del LLM sea invisible para el usuario y reduce la presión sobre la experiencia interactiva.

La compatibilidad estructural entre ambas aplicaciones está garantizada por diseño: AgentCO2 consume el JSON nativo de SCOPE sin transformación, lo que significa que cualquier evolución del esquema de SCOPE, nuevas categorías de emisión, nuevos factores, nuevos campos, se propaga al sistema multiagente con una modificación menor del *index.json* de medidas. La capa de *prompts* del sistema, almacenada en ficheros de texto plano y editable desde el panel administrativo, permite ajustar el comportamiento de los agentes a la evolución del esquema sin necesidad de redespigüe.

SCOPE atiende hoy a un volumen estimado en miles de empresas que calculan su huella cada año. De ellas, según los registros internos del MITECO, solo una minoría progresa formalmente al sello *Reduzco*, el segundo de los tres niveles del Registro Nacional de Huella de Carbono, porque la preparación del plan documentado asociado exige un esfuerzo técnico que muchas empresas pequeñas no pueden costear. La integración de AgentCO2 tiene como objetivo directo cerrar esta brecha: ofrecer a toda empresa que calcule su huella en SCOPE un plan inicial estructurado que sirva de base para la documentación del sello *Reduzco* sin coste adicional.

El impacto agregado de esta evolución, si la conversión a plan ejecutado siguiera tasas similares a las que documentan los informes de seguimiento de la Directiva Europea de Eficiencia Energética (Comisión Europea, 2024) para programas de acompañamiento, podría situarse en una **reducción potencial agregada del orden de centenas de miles de toneladas de CO₂e anuales sobre el conjunto del tejido SCOPE en un horizonte de cinco años**. La cifra concreta dependerá, evidentemente, de la tasa de ejecución real de los planes generados, una incógnita que solo la evaluación longitudinal posterior (línea futura identificada en el capítulo 10) podrá despejar.

Más allá del impacto cuantitativo directo, la integración aporta beneficios estructurales tanto para **ECODES** como para el ecosistema empresarial español:

Refuerzo del posicionamiento de SCOPE. Pasar de una herramienta de *cálculo* a una herramienta de *cálculo + plan* aumenta el valor diferencial frente a competidores. Es, en términos de producto, una progresión natural que se alinea con la dirección que la propia regulación europea (CSRD, ESRS) está imprimiendo al mercado.

Reducción de barreras al sello Reduzco. Las empresas que hoy se detienen en el sello *Calculo* por falta de recursos técnicos para preparar el plan pueden, con AgentCO2 integrado, presentar un plan estructurado generado automáticamente y validado por el consultor de **ECODES** en una segunda capa.

Generación de inteligencia agregada. El sistema multiagente, al procesar miles de inventarios, produce datos agregados sobre qué medidas se proponen con mayor frecuencia, en qué sectores, con qué reducción potencial. Esta inteligencia agregada es un activo de **ECODES** que puede informar políticas públicas, líneas de ayuda y campañas sectoriales.

Coherencia con el Pacto Climático Europeo. **ECODES** es socio oficial del Pacto Climático Europeo en España. Disponer de una herramienta de generación masiva de planes alineados con SBTi y con los objetivos europeos de 2030 y 2050 refuerza la capacidad de **ECODES** de movilizar al tejido empresarial español hacia los compromisos del Pacto.

La integración no está exenta de **riesgos** que conviene identificar y prevenir:

- **Riesgo de error en cifras del plan.** Si el sistema produce un plan con un cálculo manifiestamente incorrecto, la confianza del usuario sobre la integralidad de SCOPE se erosiona. Mitigación:
 - 1 mantener el etiquetado claro del plan como estimación automática orientativa,
 - 2 conservar el detalle de cálculo en cada medida para auditoría

- 3 establecer un canal de revisión humana opcional para empresas con planes de impacto significativo.
- **Riesgo de dependencia de proveedor LLM.** La concentración de uso en un único proveedor (OpenAI) introduce riesgo de precio, disponibilidad y cumplimiento normativo. Mitigación:
 - 1 el diseño con AG2 permite cambiar de proveedor con una modificación mínima en la configuración (*LLMConfig*)
 - 2 evaluaciones internas comparativas con modelos europeos (Mistral) y locales (Llama desplegado en infraestructura propia) están planificadas como siguiente paso.
 - **Riesgo de obsolescencia del catálogo.** Los factores de emisión se actualizan anualmente; las medidas relevantes evolucionan; las tecnologías de mitigación nuevas (hidrógeno verde, captura directa, electrificación industrial avanzada) no están representadas en el catálogo actual. Mitigación:
 - 1 institucionalizar un ciclo anual de revisión del *index.json* y de las fichas, alineado con la publicación de los factores oficiales del MITECO y de DEFRA.
 - **Riesgo regulatorio sobre IA.** El Reglamento Europeo de IA (AI Act; Reglamento UE 2024/1689) introduce obligaciones específicas para sistemas que tomen decisiones automatizadas sobre individuos. AgentCO2 opera sobre empresas, no sobre personas, y devuelve recomendaciones, no decisiones vinculantes, lo que reduce el perímetro regulatorio. Mitigación:
 - 1 documentar explícitamente que el sistema es una herramienta de soporte
 - 2 mantener la trazabilidad completa de cada cálculo
 - 3 publicar las fichas de medida y los *prompts* como información auditable.

10. Conclusiones y líneas futuras

AgentCO2 ha demostrado que es posible automatizar la traducción de un inventario GHG en un plan de mitigación personalizado con un coste y una latencia dos órdenes de magnitud por debajo de la alternativa consultora. La combinación de pre-filtrado determinista (Python), conocimiento estructurado (23 fichas JSON) y agentes LLM especializados (Selector, Info Gatherer, Calculador) ofrece un equilibrio razonable entre robustez y flexibilidad. La arquitectura es modular y suficientemente documentada para iterar sobre ella sin necesidad de reescritura. Las líneas de trabajo prioritarias son:

- **Integración nativa con ScopeCO2.** Eliminar el paso manual de subida del JSON mediante una integración API. Pendiente de definición con ECODES, requiere un endpoint público en ScopeCO2 y un mecanismo de autenticación compartido.
- **Refinamiento del campo *investment*.** El Calculador actualmente toma el punto medio de los rangos *investment_estimate* de cada ficha. Se propone introducir una llamada opcional a una herramienta de búsqueda web para refinar el rango con precios de mercado actualizados según provincia y tamaño de la empresa.
- **Caso "sin consumo este año".** Algunas empresas declaran no haber tenido consumo en una categoría durante el año reportado. Pendiente este caso en el JSON y cómo el Selector debe tratarlo (¿proponer pese a ausencia de datos históricos?).
- **Catálogo ampliado.** Las 23 medidas actuales cubren la mayoría de los escenarios de pymes españolas pero dejan fuera categorías como electrificación de procesos industriales, captura de carbono en ámbito ganadero o compensaciones voluntarias. La arquitectura admite añadir fichas nuevas sin modificar el código del agente.
- **Memoria entre sesiones.** Actualmente el estado se pierde en cada reinicio del *backend*. Se propone introducir persistencia ligera (SQLite o Postgres) para que el usuario pueda retomar un plan en sesiones sucesivas. La incorporación de mecanismos de memoria a largo plazo para agentes LLM, sintetizados por Zhang et al. (2024), es una línea de trabajo activa de la comunidad de investigación.
- **Modo multitenant y RBAC.** El sistema asume un único operador. La extensión a cuentas por empresa, históricos, roles de consultor y de cliente, no está en el alcance de esta prueba de concepto, pero es el paso natural para convertirlo en producto.
- **Evaluación longitudinal.** Medir, en empresas que implementen el plan, qué porcentaje de las medidas propuestas se llegan a ejecutar y qué reducción real se observa frente a la estimada. Requiere un horizonte temporal mayor (12-24 meses).

Con todas estas líneas pendientes, el sistema en su versión 1.0 cumple los objetivos que el equipo ITA-**ECODES** se había fijado al inicio del proyecto y queda preparado para una fase de despliegue extensivo entre las empresas usuarias de ScopeCO2.

Anexo A. Bibliografía

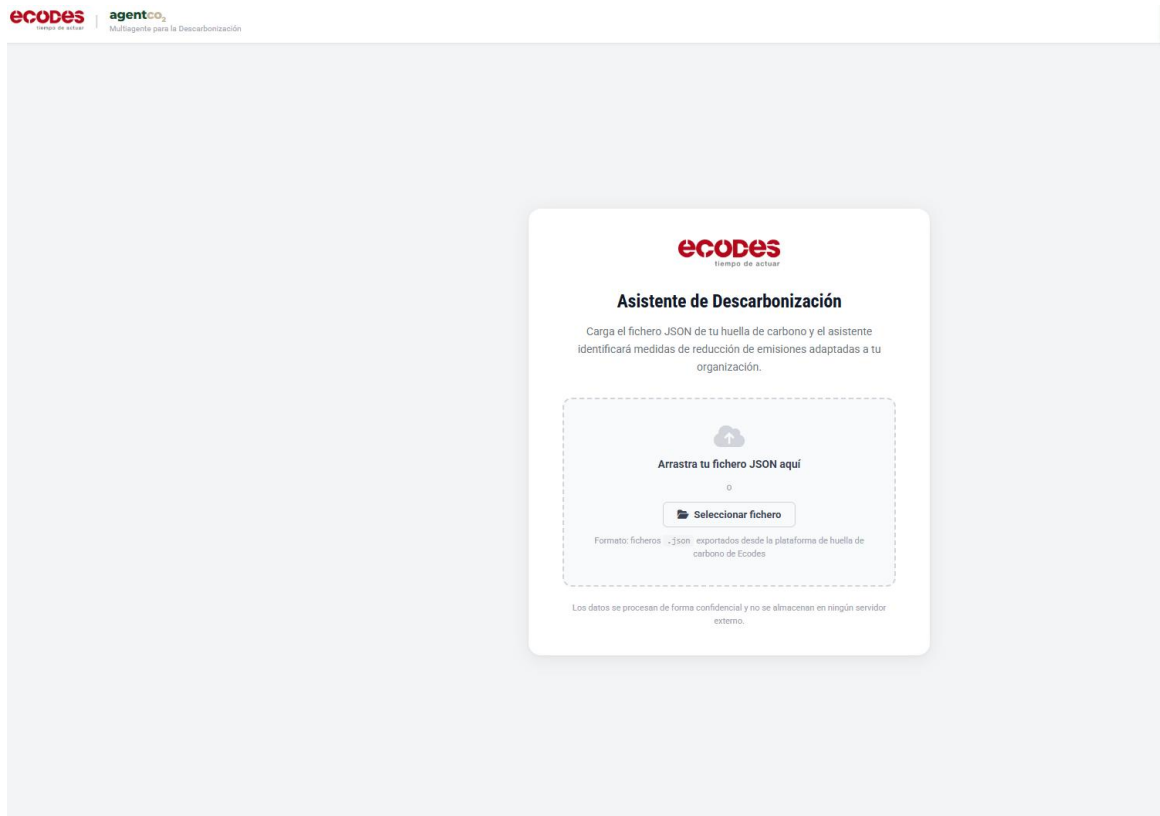
- Bain & Co. (2024). *Energy and Natural Resources Report*. Bain & Company.
- Bommasani, R., Hudson, D. A., Adeli, E. et al. (2022). On the Opportunities and Risks of Foundation Models. *arXiv:2108.07258*.
- Brown, T., Mann, B., Ryder, N. et al. (2020). Language Models are Few-Shot Learners. *NeurIPS 2020*.
- Center for Human-Compatible AI (2024). *Recommender Systems Alignment*. UC Berkeley.
- Chen, J., Smith, R., y Liu, Y. (2024). False belief stabilization in algorithmic recommendation. *Nature Human Behaviour*, 8(7).
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., y Starnini, M. (2021). The echo chamber effect on social media. *PNAS*, 118(9).
- Comisión Europea (2019). *Comunicación sobre el Pacto Verde Europeo*. COM(2019) 640 final.
- Comisión Europea (2024). *Implementation of the Energy Efficiency Directive, Progress report*. DG ENER.
- DEFRA (2024). *UK Government GHG Conversion Factors for Company Reporting*. Department for Environment, Food and Rural Affairs.
- Directiva (UE) 2022/2464 del Parlamento Europeo y del Consejo de 14 de diciembre de 2022 (CSRD).
- Directiva (UE) 2024/1760 del Parlamento Europeo y del Consejo de 13 de junio de 2024 (CSDDD).
- **ECODES** (2024). *SCOPE, calculadora de huella de carbono empresarial conforme al GHG Protocol*. Fundación Ecología y Desarrollo.
- EFRAG (2023). *European Sustainability Reporting Standards*. European Financial Reporting Advisory Group.
- Ecoinvent (2024). *Ecoinvent v3.10, life cycle inventory database*. Swiss Centre for Life Cycle Inventories.
- Greenhouse Gas Protocol (2015). *Corporate Accounting and Reporting Standard, Revised Edition*. WRI / WBCSD.
- Greenhouse Gas Protocol (2013). *Corporate Value Chain (Scope 3) Accounting and Reporting Standard*. WRI / WBCSD.
- Guo, T., Chen, X., Wang, Y. et al. (2024). Large Language Model based Multi-Agents: A Survey of Progress and Challenges. *arXiv:2402.01680*.
- Hong, S., Zheng, X., Chen, J. et al. (2024). MetaGPT: Meta Programming for a Multi-Agent Collaborative Framework. *ICLR 2024*.
- IDAE (2022). *Factores de emisión y guía de uso para el cálculo de la huella de carbono*. Instituto para la Diversificación y Ahorro de la Energía.
- IEA (2023). *Energy Efficiency 2023*. International Energy Agency, París.
- IPCC (2021). *Climate Change 2021: The Physical Science Basis*. Sixth Assessment Report, Working Group I.
- IPCC (2022). *Climate Change 2022: Mitigation of Climate Change*. Sixth Assessment Report, Working Group III.
- Lewis, P., Perez, E., Piktus, A. et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *NeurIPS 2020*.
- Liu, N. F., Lin, K., Hewitt, J. et al. (2024). Lost in the Middle: How Language Models Use Long Contexts. *TACL*, 12.
- McKinsey & Company (2023). *The Net-Zero Transition: What it would cost, what it could bring*.
- MITECO / OECC (2023). *Factores de emisión, Registro de huella de carbono*. Ministerio para la Transición Ecológica y el Reto Demográfico.

- MITECO (2025). *Memoria anual del Registro Nacional de Huella de Carbono, Compensación y Proyectos de Absorción de Dióxido de Carbono*.
- OCCC (2024). *Guia pràctica per al càlcul d'emissions de gasos amb efecte d'hivernacle*. Oficina Catalana del Canvi Climàtic.
- OpenAI (2024). *GPT-4o Technical Report*.
- Park, J. S., O'Brien, J., Cai, C. J. et al. (2023). Generative Agents: Interactive Simulacra of Human Behavior. *UIST 2023*.
- PVGIS (2024). *Photovoltaic Geographical Information System, API documentation*. European Commission Joint Research Centre.
- Qian, C., Liu, W., Liu, H. et al. (2024). ChatDev: Communicative Agents for Software Development. *ACL 2024*.
- Real Decreto 163/2014, de 14 de marzo, por el que se crea el registro de huella de carbono, compensación y proyectos de absorción de dióxido de carbono.
- Reglamento (UE) 2021/1119 del Parlamento Europeo y del Consejo (Ley Europea del Clima).
- Reglamento (UE) 2023/1115 del Parlamento Europeo y del Consejo (productos asociados a la deforestación).
- Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo (Reglamento Europeo de Inteligencia Artificial, AI Act).
- Rolnick, D., Donti, P. L., Kaack, L. H. et al. (2022). Tackling Climate Change with Machine Learning. *ACM Computing Surveys*, 55(2).
- Schick, T., Dwivedi-Yu, J., Dessì, R. et al. (2023). Toolformer: Language Models Can Teach Themselves to Use Tools. *NeurIPS 2023*.
- Science Based Targets initiative (2024). *SBTi Corporate Net-Zero Standard v2.0*.
- Shinn, N., Cassano, F., Berman, E. et al. (2023). Reflexion: Language agents with verbal reinforcement learning. *NeurIPS 2023*.
- United Nations Framework Convention on Climate Change (2015). *Paris Agreement*. UNFCCC.
- Wei, J., Wang, X., Schuurmans, D. et al. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *NeurIPS 2022*.
- World Business Council for Sustainable Development (2023). *Towards a Net-Zero Future*.
- Wu, Q., Bansal, G., Zhang, J. et al. (2023). AutoGen: Enabling next-gen LLM applications via multi-agent conversation. *arXiv:2308.08155*.
- Yao, S., Zhao, J., Yu, D. et al. (2023). ReAct: Synergizing Reasoning and Acting in Language Models. *ICLR 2023*.
- Zhang, Z., Bo, X., Ma, C. et al. (2024). A Survey on the Memory Mechanism of Large Language Model based Agents. *arXiv:2404.13501*.

Anexo B. Glosario de términos

AG2 / AutoGen	Framework multiagente que orquesta agentes conversacionales LLM con <i>handoffs</i> explícitos y soporte nativo para herramientas. Mantenido por la comunidad tras su separación de Microsoft Research.
Alcance 1, 2, 3	Clasificación del GHG Protocol para las emisiones de gases de efecto invernadero según el grado de control de la organización informante.
CNAE	Clasificación Nacional de Actividades Económicas. Códigos numéricos asignados por el INE a cada sector económico.
CSRD	Corporate Sustainability Reporting Directive (Directiva UE 2022/2464). Obliga a la publicación de información de sostenibilidad por parte de grandes empresas y pymes cotizadas en la UE.
CSDDD	Corporate Sustainability Due Diligence Directive (Directiva UE 2024/1760). Extiende la responsabilidad sobre la cadena de valor.
Factor de emisión (FE)	Coeficiente que convierte una unidad de actividad (kWh, km, kg) en kg o t de CO ₂ equivalente emitidos.
Garantía de Origen (GdO)	Certificado electrónico que acredita que la electricidad consumida procede de fuentes renovables. Reduce el factor de emisión efectivo del contrato eléctrico a cero.
GHG Protocol	Estándar internacional de contabilidad y reporte de gases de efecto invernadero elaborado por WRI y WBCSD.
Handoff	En AG2, transferencia explícita del turno entre dos agentes en una conversación. Permite modelar pipelines multipaso.
Langfuse	Plataforma de observabilidad específica para aplicaciones basadas en LLM. Captura <i>traces</i> , costes y tokens por sesión.
LLM	Large Language Model. Modelo de lenguaje de gran escala (GPT-4o, Claude, Gemini, Llama, etc.).
OECC	Oficina Española de Cambio Climático del MITECO. Publica los factores de emisión oficiales y gestiona el Registro Nacional de Huella de Carbono.
PCA / GWP	Potencial de Calentamiento Atmosférico (Global Warming Potential). Coeficiente que compara el efecto invernadero de un gas con el del CO ₂ (PCA = 1) en un horizonte de 100 años.
PVGIS	Photovoltaic Geographical Information System de la Comisión Europea. API gratuita que estima la producción fotovoltaica anual en cualquier punto geográfico.
RAG	Retrieval-Augmented Generation. Patrón en el que un LLM consulta una base de conocimiento externa antes de generar la respuesta.
SCOP	Seasonal Coefficient of Performance. Eficiencia media estacional de una bomba de calor. En la ficha de M1-COMBUSTION se asume SCOP 3,2.
SBTi	Science Based Targets initiative. Iniciativa internacional que valida objetivos corporativos de reducción de emisiones alineados con el Acuerdo de París.
SCOPE / ScopeCO2	Calculador de huella de carbono empresarial desarrollado por ECODES, conforme al GHG Protocol. Produce el JSON normalizado que alimenta AgentCO2. El uso de "SCOPE" como denominación abreviada es habitual en la documentación interna del proyecto.
Transmitancia térmica (U)	Flujo de calor por unidad de superficie y de diferencia de temperatura entre dos ambientes (W/m ² ·K). Cuanto menor sea U, mejor aislamiento.
WebSocket	Protocolo de comunicación bidireccional persistente sobre TCP. Permite al backend enviar eventos al cliente sin que éste tenga que solicitarlos.

Anexo C. Imágenes de la herramienta



Inicio > Empresa de prueba · Huella 2025

Huella de Carbono 2025

Empresa de prueba · La Puebla de Alfindén, Zaragoza · Datos reales · 1 jun 2026

DESGLOSE DE EMISIONES POR ALCANCE

TOTAL

642,45 tCO₂e / año

● Alcance 1 — Emisiones directas · 76%

488,55 tCO₂e

Combustibles fósiles		32,12
Gases refrigerantes		138,00
Gases de incendios		49,76
Desplazamientos propios		268,63
Maquinaria		0,05

● Alcance 2 — Energía adquirida · 5.1%

33,00 tCO₂e

Electricidad · ENDESA ENERGIA, S.A.		33,00
-------------------------------------	-------------	-------

● Alcance 3 — Cadena de valor · 18.8%

120,89 tCO₂e

Agua		0,00
Residuos		2,75
Transporte subcontratado		0,00
Actividades relacionadas		118,14

❶ El alcance de cada categoría se determina según los criterios del GHG Protocol a partir de los datos del JSON de huella.

DATOS DEL CÁLCULO

Nombre	Empresa de prueba
Año	2025
Localidad	La Puebla de Alfindén, Zaragoza
Superficie	10000 m ²
Empleados	345
Comercializadora	ENDESA ENERGIA, S.A.

[Iniciar análisis de descarbonización →](#)

ecodes

tiempo de actuar

agentco₂

Multiagente para la Descarbonización

Reiniciar

Empresa de prueba

Inicio > Empresa de prueba · Huella 2025

Huella de Carbono 2025

Empresa de prueba · La Puebla de Alfindén, Zaragoza · Datos reales · 1 jun 2026

DESGLOSE DE EMISIONES POR ALCANCE

TOTAL

642,45 tCO₂e / año

● Alcance 1 – Emisiones directas · 2

Combustibles fósiles

Gases refrigerantes

Gases de incendios

Desplazamientos propios

Maquinaria

● Alcance 2 – Energía adquirida · 5.3

Electricidad · ENDESA ENERGIA, S.A.

● Alcance 3 – Cadena de valor · 18.8

Agua

Residuos

Transporte subcontratado

Actividades relacionadas

El alcance de cada categoría se determina según los criterios del

GHG Protocol

a partir de los datos del JSON de huella.

DATOS DEL CÁLCULO

Nombre

Empresa de prueba

2025

La Puebla de Alfindén, Zaragoza

10000 m²

345

zadora

ENDESA ENERGIA, S.A.

análisis de descarbonización →

Una pregunta antes de empezar

¿El local donde opera la empresa es propio o alquilado? Esto nos permite filtrar las medidas que realmente pueden aplicarse.

Propio

Alquilado

Iniciar análisis →

ecodes

tiempo de actuar

agentco₂

Multiagente para la Descarbonización

Reiniciar

Empresa de prueba

Analizando tu huella de carbono

Empresa de prueba · 642 tCO₂e/año

✓ JSON cargado correctamente

✓ SELECTOR Identificando medidas de reducción aplicables...

○ CALCULADOR Calculando impacto económico y de emisiones...

○ Generando informe...

49

www.ecodes.org

ecodes

agentCO₂

Multigente para la Descarbonización

Reiniciar

Empresa

Inicio > Empresa de prueba > Plan de Reducción

Plan de Reducción de Emisiones

< Volver

Descargar PDF

Empresa de prueba - La Puebla de Alfindén, Zaragoza - Generado el 1 de junio de 2026

INVERSIÓN MÁX.: 350.000 €

+0 €/año

ahorro neto

-6,81 tCO₂e

reducción/año

1%

de la huella

M1-FLOTA-EV

SIN INVERSIÓN INICIAL

Sustitución de vehículos de combustibles tradicionales por vehículos híbridos, híbridos enchufables y eléctricos

La transición a vehículos eléctricos en la flota podría reducir las emisiones en 3.14 tCO₂e/año.

-3,14 tCO₂e

+0 €/año

0 €

Inmediato

REDUCCIÓN/AÑO

AHORRO

INVERSIÓN

RETORNO INVERSIÓN

Ver cálculo

M1-COMBUSTION

SIN INVERSIÓN INICIAL

Sustitución de combustibles fósiles en los sistemas de climatización

Al reemplazar los combustibles fósiles en la climatización por tecnología de bomba de calor, se espera una reducción de emisiones de 2.52 tCO₂e/año.

-2,52 tCO₂e

+0 €/año

0 €

Inmediato

REDUCCIÓN/AÑO

AHORRO

INVERSIÓN

RETORNO INVERSIÓN

Ver cálculo

M1-BIOCOMBUSTIBLE

SIN INVERSIÓN INICIAL

Uso de biocombustibles en vehículos de flota

El uso de biocombustibles podría reducir las emisiones en 0.79 tCO₂e/año.

-0,79 tCO₂e

+0 €/año

0 €

Inmediato

REDUCCIÓN/AÑO

AHORRO

INVERSIÓN

RETORNO INVERSIÓN

Ver cálculo

M3-AGUA-GRIFOS

SIN INVERSIÓN INICIAL

Instalar reguladores de caudal en grifos

La instalación de reguladores de caudal puede ahorrar 0.36 tCO₂e/año en emisiones de agua.

-0,36 tCO₂e

+0 €/año

0 €

Inmediato

REDUCCIÓN/AÑO

AHORRO

INVERSIÓN

RETORNO INVERSIÓN

Ver cálculo

8 medidas más disponibles con datos adicionales

1 recomendaciones estratégicas

Informe generado por el Sistema Multigente de Descarbonización de Ecodes - 1 de junio de 2026

Los cálculos son estimaciones orientativas generados por inteligencia artificial a partir de los datos de huella de carbono proporcionados. Para un análisis detallado y asesoramiento personalizado, contacta con el equipo de Ecodes: info@ecodes.org

ecodes

agentCO₂

Multigente para la Descarbonización

Reiniciar

Empresa de prueba

Instalar reguladores de caudal en grifos

La instalación de reguladores de caudal puede ahorrar 0.36 tCO₂e/año en emisiones de agua.

-0,36 tCO₂e

+0 €/año

0 €

Inmediato

REDUCCIÓN/AÑO

AHORRO

INVERSIÓN

RETORNO INVERSIÓN

Ver cálculo

8 medidas más disponibles con datos adicionales

M1-ENVOLVENTE

Mejora del aislamiento térmico de la envolvente

Datos necesarios: estado actual del aislamiento de fachada y cubierta, y antigüedad del edificio

Calcular esta medida

M1-VENTANAS

Sustituir las ventanas por ventanas con menor transmitancia térmica

Datos necesarios: número y superficie de ventanas, transmitancia actual (valor U), horas de calefacción

Calcular esta medida

M1-REFRIGERANTES

Sustitución de gases refrigerantes de alto PCA por alternativas de menor impacto ambiental

Datos necesarios: antigüedad del equipo de refrigeración y tipo de gas refrigerante

Calcular esta medida

M2-FOTOLAB

Instalación solar fotovoltaica para autoconsumo

Datos necesarios: datos del tejado (disponibilidad, superficie en m²) y precio del kWh

Calcular esta medida

M2-LUZ

Sustitución de luminarias tradicionales por luminarias LED

Datos necesarios: inventario de luminarias (número, potencia en W) y horas de uso anuales

Calcular esta medida

M2-COMPRESOR

Optimización energética en sistemas de aire comprimido

Datos necesarios: existencia y características del compresor (potencia en kW, horas de operación diarias)

Calcular esta medida

M3-ABORDIZADO

Plataforma de gestión de energía sostenible de Ecodes para la MIPSA

agentco₂

Multigente para la Descarbonización

Reiniciar

Empres

Instalar reguladores de caudal en grifos

La instalación de reguladores de caudal puede ahorrar 0.36 tCO₂e/año en emisiones de agua.

-0,36 tCO₂e

REDUCCIÓN/AÑO

+0 €/año

AHORRO

0 €

INVERSIÓN

Inmediato

RETORNO INVERSIÓN

Ver cálculo

8 medidas más disponibles con datos adicionales

M1-AISLAMIENTO

Mejora del aislamiento térmico de la envolvente

Datos necesarios: estado actual del aislamiento de fachada y cubierta, y antigüedad del edificio

Calcular esta medida

M1-VENTANAS

Sustituir las ventanas por ventanas con menor transmitancia térmica

Datos necesarios: número y superficie de ventanas, transmitancia actual (valor U), horas de calefacción

Calcular esta medida

M1-REFRIGERANTES

Sustitución de gases refrigerantes de alto P

Datos necesarios: antigüedad del equipo de refrigeración

Calcular esta medida

M2-SOLAR

Instalación solar fotovoltaica para autoconsumo

Datos necesarios: datos del tejado (disponibilidad, superficie en m²) y precio del kWh

Asistente

Mejora del aislamiento térmico de la envolvente

Hola, soy un asistente especializado en descarbonización empresarial. Para evaluar la mejora del aislamiento térmico de tu edificio, necesitare recopilar algunos datos. Primero, ¿podrías decirme cuántos años tiene el edificio?

Escribe tu respuesta...

agentco₂

Multigente para la Descarbonización

Reiniciar

Empresa de pr

Instalar reguladores de caudal en grifos

La instalación de reguladores de caudal puede ahorrar 0.36 tCO₂e/año en emisiones de agua.

-0,36 tCO₂e

REDUCCIÓN/AÑO

+0 €/año

AHORRO

0 €

INVERSIÓN

Inmediato

RETORNO INVERSIÓN

Ver cálculo

8 medidas más disponibles con datos adicionales

M1-AISLAMIENTO

Mejora del aislamiento térmico de la envolvente

Datos necesarios: estado actual del aislamiento de fachada y cubierta, y antigüedad del edificio

Calcular esta medida

M1-VENTANAS

Sustituir las ventanas por ventanas con menor transmitancia térmica

Datos necesarios: número y superficie de ventanas, transmitancia actual (valor U), horas de calefacción

Calcular esta medida

M1-REFRIGERANTES

Sustitución de gases refrigerantes de alto P

Datos necesarios: antigüedad del equipo de refrigeración

Calcular esta medida

M2-SOLAR

Instalación solar fotovoltaica para autoconsumo

Datos necesarios: datos del tejado (disponibilidad, superficie en m²) y precio del kWh

Calcular esta medida

M2-LED

Sustitución de luminarias tradicionales por LED

Datos necesarios: inventario de luminarias (número, potencia, antigüedad)

Calcular esta medida

Asistente

Mejora del aislamiento térmico de la envolvente

Hola, soy un asistente especializado en descarbonización empresarial. Para evaluar la mejora del aislamiento térmico de tu edificio, necesitare recopilar algunos datos. Primero, ¿podrías decirme cuántos años tiene el edificio?

35

Gracias por la información. Ahora, ¿cómo describirías el aislamiento de la fachada? Las opciones son: ninguno, básico, medio o bueno.

básico

Gracias por la información. Por último, ¿cómo describirías el aislamiento de la cubierta? Las opciones son: ninguno, básico, medio o bueno.

ninguno

He recopilado toda la información necesaria sobre el aislamiento térmico de tu edificio. El aislamiento de la fachada es básico y el de la cubierta es ninguno, con el edificio teniendo 35 años. Esto sugiere un gran potencial de mejora en la eficiencia energética.

Calculando impacto de la medida...



Con el apoyo de:



GRACIAS

Contacto

 www.ecodes.org

 +34 976 29 82 82

 Plaza San Bruno, 9, 1º Oficinas, 50001 Zaragoza.