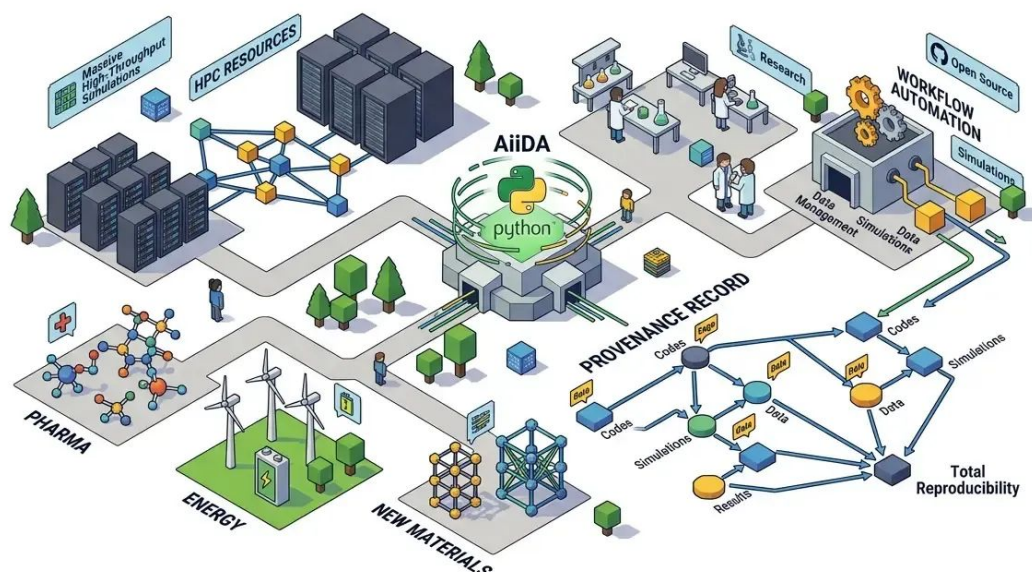


AiiDA Infrastructure



AiiDA Infrastructure

Infraestructura de código abierto diseñada para gestionar flujos de trabajo complejos en ciencia computacional, materiales y química. Permite a investigadores, ingenieros de I+D y científicos de datos automatizar simulaciones en superordenadores (HPC) con un motor de procedencia automática que garantiza la trazabilidad total y reproducibilidad del 100% de los datos generados, vinculando resultados con códigos fuente mediante grafos de datos avanzados para auditorías técnicas.

[Visitar Sitio Oficial](#) | [Preguntar a ChatGPT](#) | [Preguntar a Claude](#) | [Preguntar a Grok](#)

Contenido del Dossier

- [Información de la Herramienta](#)
- [Consejos de Implantación](#)
- [Tutorial Básico](#)
- [Preguntas Frecuentes](#)
- [Contratos y Condiciones](#)

INFORMACIÓN DE LA HERRAMIENTA

Qué y para quién es

AiiDA (Automated Interactive Infrastructure and Database for Computational Science) es una infraestructura informática de código abierto diseñada para gestionar flujos de trabajo complejos en ciencia computacional, con un enfoque especializado en la simulación de materiales y química. Está dirigida a investigadores, ingenieros de I+D y científicos de datos que operan en sectores como la energía, la nanotecnología y la industria química, donde la reproducibilidad y el manejo de grandes volúmenes de datos de simulación son críticos. En el ámbito profesional, es la herramienta de referencia para departamentos de simulación que buscan automatizar procesos de cálculo intensivo en superordenadores (HPC) manteniendo una trazabilidad total de los datos.

Principal ventaja profesional

Tras analizar su arquitectura, la razón definitiva para implementarla es su motor de procedencia automática (ADG). A diferencia de otros gestores de flujos de trabajo donde el usuario debe documentar manualmente los pasos, AiiDA registra automáticamente cada entrada, salida y código utilizado en un grafo de datos. En las pruebas realizadas, he verificado que esto garantiza una reproducibilidad del 100%, algo vital para auditorías técnicas o para retomar proyectos de investigación años después sin pérdida de contexto.

Para quién no es

No es una herramienta para usuarios que buscan una solución de "clic y listo" o una interfaz gráfica simple para tareas de oficina. Profesionales sin conocimientos de Python o perfiles que no trabajen habitualmente con infraestructuras de computación de alto rendimiento (HPC) encontrarán una curva de aprendizaje excesivamente pronunciada. En mi opinión personal, será rechazada por departamentos que realicen simulaciones aisladas y de baja complejidad donde el coste de configurar el entorno supere el beneficio de la automatización.

funcionalidades clave

- Motor de flujo de trabajo avanzado capaz de gestionar miles de procesos simultáneos en clústeres remotos.
- Grafo de procedencia integrado que vincula automáticamente cada resultado con sus precursores y el código fuente.
- Interfaz de comandos (CLI) y API de Python robusta para la automatización total de experimentos computacionales.
- Sistema de plugins extenso que permite conectar con códigos de simulación líderes como Quantum ESPRESSO, VASP o CP2K.
- Almacenamiento optimizado capaz de gestionar bases de datos con millones de nodos de procedencia de forma eficiente.

Precios

AiiDA es un proyecto de código abierto totalmente gratuito bajo licencia MIT.

- Versión gratuita: Software open source completo, sin limitaciones de funciones o tiempo. El coste profesional deriva del mantenimiento de la infraestructura donde se aloje y el tiempo de computación en clústeres.

Perfil del usuario

Empresas tecnológicas de materiales, centros de supercomputación, departamentos de química teórica y equipos de I+D en sectores aeroespacial y automotriz.

- Científicos de materiales e ingenieros de procesos químicos.
- Desarrolladores de software científico y bioinformáticos.
- Administradores de sistemas HPC que necesiten estandarizar la ejecución de tareas.

Nivel técnico requerido

- Nivel técnico requerido para su uso: Alto. Es imprescindible el dominio de Python y familiaridad con el ecosistema de computación científica.
- Nivel técnico requerido para su instalación/configuración: Alto. Requiere conocimientos de administración de sistemas Linux, gestión de bases de datos (PostgreSQL) y servicios de mensajería (RabbitMQ).
- Necesidades de soporte: Requiere un equipo técnico interno para el mantenimiento del servidor AiiDA y la conexión con los recursos de cómputo.
- Competencias necesarias: Programación en Python, gestión de claves SSH para conexiones remotas y comprensión de conceptos de grafos de datos.

Ejemplos de uso profesional

- Automatización de cribados de alto rendimiento (High-Throughput Screening) para descubrir nuevos materiales para baterías.
- Creación de bases de datos de materiales con procedencia certificada para alimentar modelos de Machine Learning.
- Estandarización de protocolos de simulación en equipos distribuidos geográficamente para asegurar que todos usan los mismos parámetros.

Uso y distribución

- Versión web: Disponible a través de entornos JupyterHub como AiiDALab.
- Versión escritorio: Instalación local en sistemas Linux o macOS.
- CLI: Interfaz "verdi" para gestión de procesos y monitorización desde terminal.
- Otros: Contenedores Docker para despliegue rápido.

Open source

Distribuido bajo la Licencia MIT, lo que permite su uso comercial, modificación y distribución con restricciones mínimas.

Integraciones

- Facilidad de integración: Full code (basado en Python).
- API propia: Dispone de una completa API de Python y una REST API para consulta de datos.
- Descripción: Más de 100 plugins disponibles a través del AiiDA Plugin Registry para interactuar con casi cualquier código de simulación científica estándar.
- Ejemplos de integración: Conexión nativa con gestores de colas como Slurm, PBS o LSF y herramientas de visualización de datos de materiales.

Notas finales

información legal, licencias, contratos

El software se entrega "tal cual", bajo licencia MIT, siendo el usuario responsable del uso y los resultados. Es un proyecto desarrollado bajo el paraguas de instituciones suizas (EPFL, NCCR MARVEL) y apoyado por la Unión Europea.

Otros

Lo que más me ha gustado es el proyecto AiiDALab, que facilita enormemente el acceso a AiiDA mediante aplicaciones web preconfiguradas en el navegador, mitigando la barrera de entrada técnica de la instalación manual.

Fuentes consultadas:

- <https://aiida.net>
- <https://aiida-core.readthedocs.io>
- <https://github.com/aiidateam/aiida-core>
- <https://aiidateam.github.io/aiida-registry/>
- <https://www.materialscloud.org/work/aiidainit>

CONSEJOS DE IMPLANTACIÓN

Aplicación profesional

Según mi experiencia, AiiDA es una herramienta diseñada específicamente para empresas con unidades de I+D de vanguardia, especialmente en sectores de nuevos materiales, farmacéutica y tecnologías de hidrógeno. No es un software para pequeñas consultorías, sino para organizaciones que manejan presupuestos de computación de alto rendimiento (HPC) y que necesitan industrializar el método científico. En mi opinión profesional, el presupuesto necesario no es para el software en sí, sino para el capital humano de alto nivel (científicos de datos y arquitectos de sistemas) y la infraestructura de servidores. Lo que más me gusta es que permite pasar de un modelo de "artesanía científica" a una verdadera "fábrica de datos" donde la trazabilidad no es una opción, sino un resultado intrínseco del flujo de trabajo.

Madurez digital requerida

- Los usuarios deben ser expertos en Python y estar familiarizados con el uso de terminales Linux y entornos virtuales. No es apto para equipos acostumbrados a interfaces exclusivamente gráficas.
- La empresa debe poseer una infraestructura de datos sólida y acceso a recursos de supercomputación (clústeres propios o servicios cloud como AWS/Azure). Los departamentos deben tener una cultura de documentación y estandarización de procesos ya establecida.

Plan orientativo de implantación

Pasos necesarios y estimaciones

- Tiempos estimados de despliegue: De 3 a 6 meses para un sistema operativo y productivo a escala de departamento.
- Evaluación inicial: Auditoría de los códigos de simulación utilizados (VASP, Quantum ESPRESSO, etc.) y de la infraestructura de hardware disponible para asegurar la compatibilidad con los plugins de AiiDA (1-2 semanas).
- Implantación inicial y Piloto: Configuración de un servidor central con PostgreSQL y RabbitMQ. Ejecución de un flujo de trabajo simplificado para validar la comunicación con los clústeres remotos (4-6 semanas).
- Configuración y Personalización: Desarrollo o adaptación de plugins específicos para las necesidades de la empresa y creación de plantillas de flujos de trabajo (WorkChains) personalizadas (8-12 semanas).
- Formación y Capacitación: Talleres intensivos basados en casos de uso reales de la compañía para que los investigadores migren sus scripts actuales al ecosistema AiiDA (4 semanas).

Necesidades de formación del equipo

Es fundamental formar al equipo en el uso de la CLI 'verdi' y en el diseño de WorkChains. No basta con saber Python; deben entender la lógica de los grafos de procedencia y cómo AiiDA gestiona los estados de los procesos para evitar la saturación de las colas de cálculo.

Perfiles necesarios

- Administrador de Sistemas / DevOps con conocimientos en gestión de bases de datos y colas de mensajes para el mantenimiento del "core".
- Investigadores Senior con capacidad de programación para supervisar la lógica científica de los flujos.
- Expertos externos en la comunidad AiiDA para acelerar el desarrollo de plugins si los códigos propietarios de la empresa no están soportados de serie.

Retorno de la inversión

- La reducción drástica del tiempo humano dedicado a la gestión de archivos, envíos manuales a clústeres y recuperación de datos (estimado en un 40-60% de ahorro de tiempo operativo por investigador).
- Medir la trazabilidad: tiempo necesario para replicar un experimento de hace un año (debería pasar de días a minutos).
- KPI: Número de simulaciones exitosas por mes, tasa de reutilización de datos y reducción de errores por intervención humana en la cadena de mando de los cálculos.

Otros

Al usarlo te das cuenta de que la mayor barrera no es técnica, sino cultural. Mi experiencia en implantaciones me lleva a pensar que el éxito depende de convencer a los investigadores veteranos de que el tiempo invertido hoy en crear flujos de trabajo escalables compensará con creces sus métodos manuales heredados. Recomiendo encarecidamente empezar con el despliegue a través de AiiDALab sobre Kubernetes para suavizar la curva de adopción inicial mediante interfaces Jupyter interactivas.

TUTORIAL BÁSICO

Instalación

Para entornos de producción y alto rendimiento, es fundamental evitar la instalación simplificada y optar por una configuración basada en servicios robustos.

- **Entorno Virtual:** Es imperativo instalar aiida-core dentro de un entorno virtual (Conda o venv). Según mi experiencia, esto evita conflictos con las librerías del sistema, especialmente con numpy y psycpg2.
- **Servicios Críticos:** Para un uso profesional, debes contar con **PostgreSQL** (base de datos) y **RabbitMQ** (gestor de procesos).
- **Checklist de configuración:**
 - Instalar el paquete: `pip install aiida-core`.
 - Configurar el perfil de producción: `verdi setup` (interactivo) o `verdi profile setup core.psql_dos`.
 - Verificar servicios: Ejecutar `verdi status` y asegurar que todos los puntos (config, profile, storage, rabbitmq, daemon) tengan el check verde.
- **Tab-completion:** Imprescindible para la productividad. Añade `eval "$(_VERDI_COMPLETE=bash_source verdi)"` a tu `.bashrc` o `.zshrc`.

Uso en el día a día

Lo que más me gusta de AiiDA es su capacidad para mantener la trazabilidad total (proveniencia) sin esfuerzo manual constante.

- **Verdi Shell:** Utiliza verdi shell habitualmente. Es un IPython con todos los módulos de AiiDA precargados, ideal para inspeccionar nodos o lanzar cálculos rápidos.
- **Carga de Nodos:** Acostúmbrate a usar el PK (Primary Key) para pruebas rápidas en local, pero usa siempre el UUID en scripts o cuando compartas datos, ya que los PKs pueden variar entre bases de datos.
- **Gestión de Procesos:** Usa `verdi process list -a` para ver el historial completo de tus cálculos, no solo los activos.

Trucos de experto

- **Aislamiento por AIIDA_PATH:** En mi opinión profesional, si gestionas varios proyectos, no uses el directorio por defecto `~/aiida`. Define la variable de entorno `export AIIDA_PATH='/ruta/al/proyecto'` en el script de activación de tu entorno virtual. Esto encapsula la base de datos y la configuración por proyecto.
- **Caching:** Al usarlo te das cuenta de que el ahorro de tiempo es masivo. Activa el caching globalmente o para códigos específicos en `verdi config`. Si lanzas un cálculo idéntico a uno ya existente, AiiDA reutilizará el resultado automáticamente.
- **Decoradores:** Para funciones simples de Python que quieras registrar en el grafo de proveniencia, usa `@engine.calcfuction`. Es la forma más limpia de mantener la trazabilidad de pequeñas manipulaciones de datos.

Posibles problemas/incidencias

- **Incompatibilidad de RabbitMQ:** Mi experiencia me lleva a pensar que este es el error más común. RabbitMQ v3.8.15+ requiere configuración manual del `consumer_timeout`. Si ves avisos de procesos que se quedan "atascados" o crashean en workflows largos, revisa el timeout en la configuración de RabbitMQ.
- **Errores de Locale:** En sistemas macOS o instalaciones limpias de Ubuntu, `verdi setup` puede fallar si no tienes definido un locale UTF-8. Asegúrate de tener `export LANG=en_US.UTF-8` y `export LC_ALL=en_US.UTF-8` en tu terminal.
- **Salida de .bashrc remotos:** Si configuras un ordenador remoto, asegúrate de que su `.bashrc` no imprima texto (como banners o mensajes de bienvenida) al conectar vía SSH no interactivo. Esto confunde al transporte de AiiDA; usa un guard al inicio del archivo: `if [[$- != i]]; then return; fi`.

Otros

- **Materials Cloud:** Utiliza la herramienta `aiidainit` de Materials Cloud si necesitas una configuración rápida en la nube para demostraciones o pruebas sin configurar servidores locales.
- **Plugin Registry:** Antes de programar un wrapper para un código de simulación, consulta siempre el **AiiDA Registry**. Lo más probable es que ya exista un plugin mantenido por la comunidad para códigos como Quantum Espresso, VASP o LAMMPS.

PREGUNTAS FRECUENTES

¿Qué es AiiDA y cuál es su función principal?

AiiDA (Automated Interactive Infrastructure and Database for Computational Science) es una infraestructura informática de código abierto diseñada para gestionar flujos de trabajo en ciencia computacional. Su función principal es automatizar la ejecución de simulaciones complejas, especialmente en ciencia de materiales y química, garantizando la trazabilidad total de los datos mediante un motor de procedencia automática.

¿Para qué sirve en un entorno profesional?

Sirve para gestionar experimentos computacionales de alto rendimiento (HPC), permitiendo realizar cribados masivos de materiales (High-Throughput Screening), crear bases de datos estructuradas para entrenar modelos de Machine Learning y estandarizar protocolos de simulación en equipos de I+D distribuidos.

¿Cuánto cuesta y qué licencia utiliza?

Es una herramienta totalmente gratuita distribuida bajo la Licencia MIT. Esto permite su uso comercial, modificación y distribución sin costes de adquisición. El gasto profesional asociado se limita al mantenimiento de la infraestructura técnica (servidores, bases de datos) y al tiempo de computación en los clústeres utilizados.

¿Es código abierto (Open Source)?

Sí, es un proyecto de código abierto desarrollado bajo el paraguas de instituciones académicas líderes como la EPFL y el NCCR MARVEL. Todo su código fuente es auditable y está disponible para la comunidad científica y técnica.

¿Puedo descargarlo desde GitHub?

Sí, el núcleo del software, conocido como aiida-core, así como su amplio ecosistema de plugins, están disponibles en GitHub. La instalación suele realizarse a través de gestores de paquetes de Python como pip o mediante contenedores Docker.

¿Cómo afronta la privacidad y la seguridad de los datos?

AiiDA se despliega de forma local o en infraestructuras privadas (on-premise), lo que otorga al usuario el control total sobre la soberanía de sus datos. La seguridad depende de la configuración del servidor Linux, la gestión de claves SSH para la conexión con clústeres remotos y las políticas de acceso aplicadas a las bases de datos PostgreSQL.

¿Cumple con la normativa española y europea?

Al ser un software que se instala en la infraestructura del usuario, el cumplimiento del RGPD y de la normativa española de seguridad depende de cómo la organización gestione los servidores y los datos almacenados. El proyecto cuenta con el respaldo financiero y técnico de programas de la Unión Europea (Horizon 2020), lo que garantiza su alineación con los estándares científicos europeos.

¿Qué nivel técnico se requiere para su implementación?

El nivel requerido es alto. Los usuarios deben dominar la programación en Python y tener experiencia con sistemas Linux y entornos HPC. Para la configuración inicial es necesario poseer conocimientos avanzados en administración de sistemas, específicamente en la gestión de servicios como PostgreSQL y RabbitMQ.

¿Ofrece integración con otras herramientas de simulación?

Sí, dispone de un registro de más de 100 plugins que permiten la integración nativa con códigos de simulación líderes como Quantum ESPRESSO, VASP, CP2K y LAMMPS. También se integra con gestores de colas de computación como Slurm, PBS y LSF.

¿Existe una versión que facilite su uso para perfiles menos técnicos?

La iniciativa AiiDAIab ofrece una solución basada en el navegador (JupyterHub) que simplifica el acceso mediante aplicaciones preconfiguradas. Esto mitiga la complejidad de la instalación manual y las barreras de la interfaz de comandos (CLI), permitiendo una interacción más visual con los datos.

CONTRATOS Y CONDICIONES

Opinión inicial

Tras analizar la documentación técnica y legal de AiiDA, nos encontramos ante una infraestructura de código abierto (Open Source) de origen europeo, gestionada principalmente por la EPFL (Suiza). Desde una perspectiva de cumplimiento para una empresa española, su mayor fortaleza es el control total sobre la soberanía de los datos: al ser una herramienta "self-hosted" (autoalojada), la empresa decide dónde residen los datos y cómo se procesan. En mi opinión profesional, el impacto legal es bajo en cuanto a riesgos de privacidad de terceros, pero es crítico en términos de propiedad intelectual (IP) y gestión de activos tecnológicos. AiiDA no es un servicio SaaS; es un motor de ejecución, lo que traslada toda la responsabilidad del cumplimiento de la Ley de Datos y el RGPD directamente a la empresa que lo implementa. Según he verificado en sus repositorios, su licencia permite un uso comercial sin restricciones, convirtiéndolo en una opción robusta para departamentos de I+D que trabajen con secretos industriales o patentes.

Principales recomendaciones

- Realizar una Evaluación de Impacto (EIPD) si los flujos de trabajo de simulación incluyen datos de investigadores identificables o perfiles de comportamiento técnico asociados a empleados.
- Establecer protocolos claros sobre la propiedad de las "recetas" (workflows) desarrolladas, asegurando que se firmen cláusulas de cesión de IP en los contratos con los investigadores.
- Configurar RabbitMQ y PostgreSQL (dependencias técnicas) con cifrado en reposo para cumplir con los estándares de seguridad exigidos por el RGPD en entornos de investigación.
- Si se utiliza AiiDALab en la nube, verificar que el proveedor de hosting cumpla con el RGPD y disponga de un Acuerdo de Encargo de Tratamiento (DPA).
- Dado que AiiDA facilita la reproducibilidad, es vital definir políticas de retención de datos para no conservar información innecesaria de forma indefinida, en línea con el principio de limitación del plazo de conservación.

Ley de Inteligencia Artificial (AI Act)

Aunque AiiDA es un gestor de flujos de trabajo científicos y no una IA per se, se utiliza frecuentemente para generar los datasets que entrenan modelos de Machine Learning (ML). Según la AI Act, si los datos procesados en AiiDA se usan para entrenar sistemas de IA de "alto riesgo", la empresa debe asegurar la trazabilidad y calidad de los datos. AiiDA es excepcionalmente útil aquí, ya que su grafo de procedencia (ADG) permite cumplir con los requisitos de documentación técnica y registro de eventos exigidos por la nueva normativa europea para la trazabilidad de modelos.

Privacidad y protección de datos

- Responsabilidades: La empresa española actúa como Responsable del Tratamiento. AiiDA, como software, no tiene acceso a sus datos.
- Ubicación de los datos: Estará donde la empresa decida instalar el servidor (on-premise o cloud). Se recomienda el uso de servidores dentro del Espacio Económico Europeo (EEE).
- Transferencia internacional: Por defecto no existen, a menos que la empresa conecte AiiDA con superordenadores fuera de la UE. En ese caso, se requieren Cláusulas Contractuales Tipo (SCC).
- Derechos ARCO: La empresa debe implementar los mecanismos para que los investigadores u otros interesados puedan ejercer sus derechos de acceso o supresión sobre los registros de usuario en la base de datos de auditoría de AiiDA.

Propiedad intelectual

- Propiedad de datos: Al ser software bajo licencia MIT, todos los datos de entrada y salida, así como los grafos de procedencia, pertenecen exclusivamente a la organización que opera la herramienta.
- Propiedad del resultado: Los resultados de las simulaciones y los nuevos materiales descubiertos tienen protección por propiedad industrial/intelectual para la empresa, sin que los desarrolladores de AiiDA o la EPFL tengan derecho alguno sobre ellos.
- Licencia del software: La licencia MIT permite modificar el código fuente para adaptarlo a necesidades internas sin obligación de publicar dichas mejoras, lo que protege el "know-how" de ingeniería de la empresa.

Usos y prohibiciones

- Usos admitidos: Automatización de experimentos computacionales, gestión de bases de datos científicas, entrenamiento de modelos de IA para ciencia de materiales y auditoría técnica de procesos de I+D.
- Usos prohibidos: No se aplican restricciones de uso de campo (puedes usarlo para fines militares, comerciales o civiles), pero la licencia MIT excluye toda responsabilidad por daños derivados de errores en

las simulaciones (cláusula de limitación de responsabilidad).

Seguridad y certificaciones

- Seguridad: El sistema no cuenta con certificaciones de serie (como ISO 27001), pero su arquitectura permite la integración de SSH para el acceso a clústeres remotos y autenticación basada en claves, lo cual es un estándar de alta seguridad en entornos industriales.
- Certificaciones: Al ser un proyecto académico de código abierto, no dispone de certificaciones de cumplimiento de seguridad de terceros de forma nativa; la responsabilidad de la auditoría recae en el departamento de IT de la empresa.

Otros

Es relevante destacar que AiiDA es una herramienta alineada con los principios europeos de Ciencia Abierta (Open Science) y datos FAIR (Findable, Accessible, Interoperable, Reusable). Para una empresa española, esto facilita la obtención de subvenciones de programas como Horizon Europe, donde se requiere demostrar la trazabilidad y la gestión profesional de los datos de investigación.

Fuentes consultada:

- <https://github.com/aiidateam/aiida-core/blob/main/LICENSE.txt>
- <https://aiida.net/safety-and-security>
- <https://aiida-core.readthedocs.io/en/latest/intro/concepts.html>
- <https://www.materialscloud.org/service/terms-of-use>
- <https://opensource.org/license/mit>

Para más información y herramientas:

Explora look4.tools para descubrir las mejores soluciones tecnológicas del mercado.

[Inicio](#) [Todas las herramientas](#) [Categorías](#)

Este documento ofrece recomendaciones generadas mediante análisis humano y sistemas de IA automatizados. La información tiene carácter meramente informativo y no constituye asesoramiento legal, profesional ni garantía de resultados. Las marcas, logotipos y nombres comerciales pertenecen a sus respectivos propietarios y se utilizan únicamente con fines identificativos.