INTERACCION
Revista Digital de AIPO
Asociación Interacción Persona-Ordenador


DOI: https://doi.org/10.65234/interaccion.147

# Reimaginando la agencia de datos en sistemas agénticos: Un modelo triádico HAD para la interacción humano-datos en arquitecturas de agentes IA

## *Reimagining data agency in agent-based systems: A triadic HAD model for human-data interaction in AI agent architectures*

[(1)] Iván Durango [0009-0004-2041-7732], [(1)] José A. Gallud [0000-0002-6616-8055], [(1)] Víctor M. R. Penichet [0000-0003-1125-9344]

[(1)] Universidad de Castilla-La Mancha, Grupo de investigación ISE, Albacete, España.
{ivan.durango@uclm.es} / {jose.gallud@uclm.es} / {victor.penichet@uclm.es}

**Resumen.** Los agentes autónomos de IA están transformando la relación de las personas con sus datos, actuando como intermediarios que perciben, razonan y operan sobre la información con supervisión humana limitada. Este artículo investiga la convergencia de los principios de Interacción Humano-Datos (HDI) y las arquitecturas modernas de agentes IA, tomando como referencia el Agent Development Kit (ADK) de Google. A partir de los fundamentos de HDI —legibilidad, agencia y negociabilidad— introducimos un modelo triádico Humano-Agente-Datos (HAD) que captura los desafíos de las experiencias de datos mediadas por agentes. El análisis expone tensiones entre autonomía y control, transparencia y eficiencia de delegación, y gobernanza individual frente a colectiva. Traducimos estas ideas en siete principios de diseño centrados en el usuario y un conjunto de extensiones arquitectónicas para ADK (delegación graduada, responsabilidad epistémica y permisos negociables), y discutimos sus implicaciones para el RGPD, el AI Act y la ENIA 2.0. Se incluye un escenario ilustrativo, una comparativa de plataformas y un protocolo de evaluación piloto con usuarios que operacionaliza las métricas de legibilidad y preservación del control.

***Palabras clave:*** Interacción Humano-Datos, Agentes IA, Autonomía de Agentes, Agencia de Datos, Transparencia, Sistemas Multi-Agente, Diseño Centrado en el Usuario.

**Abstract.** Autonomous AI agents are transforming how people relate to their data, acting as intermediaries that perceive, reason about, and operate on information with limited human oversight. This article investigates the convergence of Human-Data Interaction (HDI) principles and modern AI agent architectures, using Google's

[Recibido: 19-04-26 | Aceptado: 01-06-26 | Publicado: 29-09-26 ]

Agent Development Kit (ADK) as a reference. Building on the HDI foundations — legibility, agency, and negotiability— we introduce a triadic Human-Agent-Data (HAD) model that captures the challenges of agent-mediated data experiences. The analysis exposes tensions between autonomy and control, transparency and delegation efficiency, and individual versus collective governance. We translate these insights into seven user-centred design principles and a set of architectural extensions for ADK (graduated delegation, epistemic accountability, and negotiable permissions), and discuss their implications for the GDPR, the AI Act, and ENIA 2.0. We include an illustrative scenario, an agent-platform comparison, and a pilot user-evaluation protocol that operationalises the legibility and control-preservation metrics.

**Keywords**: Human-Data Interaction, AI Agents, Agent Autonomy, Data Agency, Transparency, Multi-Agent Systems, User-Centred Design.

## 1. INTRODUCCIÓN

Una revolución silenciosa está en marcha en la relación entre las personas y sus datos. Donde los usuarios antes manipulaban la información mediante comandos explícitos, una nueva clase de entidades computacionales actúa ahora en su nombre: agentes de IA que perciben entornos, razonan sobre objetivos y ejecutan acciones con creciente independencia (Wang et al., 2023; Xi et al., 2023). El cambio de la manipulación directa a la autonomía delegada no es meramente técnico; reescribe el contrato entre los humanos y sus sistemas de datos, planteando preguntas urgentes sobre control, transparencia y responsabilidad.

La Interacción Humano-Datos (HDI), introducida por Mortier et al. (2014) y ampliada por Durango et al. (2024a), aporta tres principios rectores: legibilidad (hacer comprensibles los procesos de datos), agencia (empoderar el control del usuario) y negociabilidad (permitir la renegociación continua de permisos). Estos principios fueron concebidos para un mundo en el que el humano sigue siendo el principal tomador de decisiones. Cuando un agente autónomo decide qué API invocar, qué datos recuperar y cómo resumir los resultados —antes de que el usuario vea una sola salida— el modelo HDI tradicional necesita repensarse. El Agent Development Kit (ADK) de Google (Google Developers, 2025) ilustra el estado del arte con composición jerárquica de agentes y un rico ecosistema de herramientas, pero, como la mayoría de los marcos, carece de mecanismos explícitos para preservar la agencia del usuario sobre los datos. Esta brecha motiva nuestro trabajo.

Realizamos cuatro contribuciones: (i) un modelo triádico Humano-Agente-Datos (HAD) que extiende la teoría HDI para dar cuenta de intermediarios autónomos, con su representación gráfica; (ii) un análisis de cómo legibilidad, agencia y negociabilidad deben evolucionar en contextos multi-agente; (iii) un modelo de diseño de interacción concreto para integrar principios HDI en ADK, con patrones centrados en el usuario, extensiones arquitectónicas y un protocolo de evaluación piloto; y (iv) la conexión del modelo HAD con el marco regulatorio europeo (RGPD, AI Act y ENIA 2.0). En conjunto, ofrecen una base para construir sistemas de agentes simultáneamente autónomos y responsables.

## 2. TRABAJO RELACIONADO

### 2.1. Interacción Humano-Datos

La HDI surgió del reconocimiento de que los paradigmas clásicos de HCI resultan insuficientes ante la recolección ubicua de datos y la toma de decisiones algorítmica (Mortier et al., 2014). Sus tres pilares

abordan, respectivamente, la opacidad de las cadenas de procesamiento, el desequilibrio de poder entre individuos y controladores, y la rigidez del consentimiento único (Crabtree & Mortier, 2015; Mortier et al., 2015). Estudios recientes examinan la HDI en sus distintos contextos (Durango et al., 2024b) y amplían su alcance hacia la IA generativa (Durango et al., 2025); el marco integral de Durango et al. (2024a) la extiende con seis componentes adicionales —equidad, responsabilidad, privacidad, participación, centrado en el humano y consentimiento informado—. Sin embargo, el desafío específico de la interacción con datos mediada por agentes permanece en gran medida inexplorado.

### 2.2. Arquitecturas de Agentes IA

Los agentes modernos marcan una transición de herramientas reactivas a entidades capaces de percepción, razonamiento y acción (Wang et al., 2023). Entre los patrones clave figuran el bucle ReAct (Yao et al., 2023) y la coordinación multi-agente mediante topologías jerárquicas, entre pares o híbridas (Guo et al., 2024). La autonomía abarca un amplio espectro, desde la supervisión total hasta la operación completamente autónoma (Luo et al., 2025), generando una tensión documentada entre eficacia y control humano (Wang et al., 2020).

### 2.3. Google ADK y Plataformas de Agentes

ADK proporciona infraestructura para sistemas multi-agente con composición jerárquica, agentes de flujo de trabajo deterministas (secuencial, paralelo, bucle), agentes LLM no deterministas, estado de sesión compartido e integraciones de herramientas (Google Developers, 2025; Google, 2025). Sus herramientas de evaluación se centran en el rendimiento de tareas y su documentación enfatiza la experiencia del desarrollador. Ninguna aborda sistemáticamente el control del usuario, la transparencia o la gobernanza de datos, las preocupaciones centrales de HDI.

### 2.4. Transparencia y Explicabilidad de Agentes

Chakraborti et al. (2018) distinguen explicabilidad, legibilidad, predecibilidad y transparencia en agentes autónomos. El marco HAII de Sundar (2020) examina la psicología de la agencia mecánica percibida, y South et al. (2025) proponen delegación autenticada mediante extensiones de OAuth 2.0, pero sin fundamentación en HDI. La investigación reciente sobre dashboards de transparencia algorítmica muestra que los usuarios con acceso a explicaciones interactivas exhiben mayor confianza calibrada y menores tasas de delegación excesiva (Wang et al., 2020), evidencia que fundamenta nuestras propuestas de Capa de Legibilidad y Gestor de Delegación.

### 2.5. Interfaces de Usuario para Agentes IA

El diseño de interfaces para sistemas agénticos es un área emergente en HCI. Las directrices de Amershi et al. (2019) establecieron dieciocho principios para la interacción humano-IA; en particular, los de "iniciativa mixta" —donde tanto el usuario como el sistema pueden tomar la iniciativa— resultan directamente aplicables al modelo de delegación graduada que proponemos. Los estudios de confianza en delegación automatizada (Sundar, 2020) subrayan además que la confianza no es estática, sino que se construye, erosiona y reconstruye según la consistencia entre las acciones del agente y las expectativas del usuario, un fenómeno que el modelo HAD aborda mediante sus propiedades de opacidad de mediación y difusión de responsabilidad.

## 3. EL MODELO HUMANO-AGENTE-DATOS (HAD)

### 3.1. De Díadas a Tríadas

La HDI clásica trata la interacción humano-datos como diádica: una persona interactúa con datos a través de una interfaz. Este esquema se rompe cuando un agente autónomo se interpone entre ambos. Proponemos un modelo triádico Humano-Agente-Datos (HAD) compuesto por tres entidades: actores humanos con intenciones, valores y derechos sobre datos; agentes intermediarios con autoridad delegada y razonamiento autónomo; y recursos de datos percibidos, procesados y sobre los que se actúa.

Como muestra la Figura 1, seis vías de interacción conectan estas entidades: delegación de objetivos (Humano → Agente), operaciones autónomas (Agente → Datos), retroalimentación ambiental (Datos → Agente), comunicación de resultados (Agente → Humano), inspección directa (Humano → Datos) y conciencia de prácticas (Datos → Humano). La estructura triádica genera propiedades emergentes ausentes en modelos diádicos: opacidad de mediación (razonamiento del agente oculto al usuario), difusión de responsabilidad (dificultad para atribuir resultados a intenciones humanas frente a decisiones del agente) y emergencia colectiva (prácticas de datos que surgen de interacciones entre agentes en lugar de un diseño explícito).

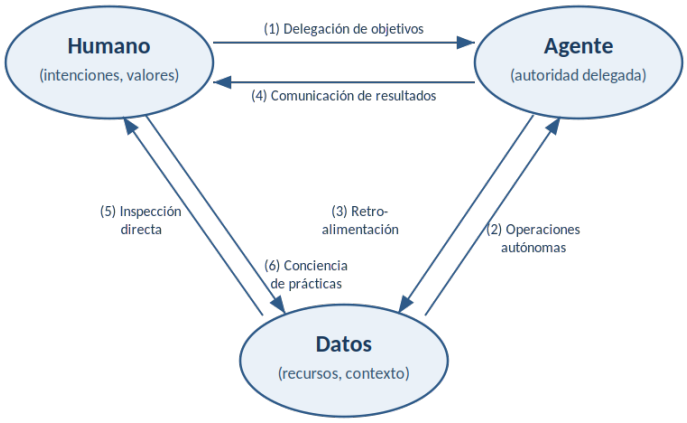


**Figura 1.** Modelo triádico HAD: seis vías de interacción entre Humano, Agente y Datos. Las flechas indican la dirección del flujo de información o control.

### 3.2. Escenario Ilustrativo

Considérese una investigadora que pide a su agente ADK: "Prepara un resumen de mis correos sobre el proyecto X de los últimos 6 meses y propón reuniones con los colaboradores más activos." El agente accede al buzón, invoca un sub-agente de resumen LLM, llama a la API de calendario y prepara invitaciones. En el modelo diádico HDI clásico, la investigadora tendría que controlar cada acción manualmente. En el modelo HAD, el agente opera con autonomía acotada (nivel 3), genera una traza de legibilidad operativa ("He accedido a 47 correos, seleccionado 12 remitentes únicos, propuesto 3 reuniones") y presenta las invitaciones para aprobación prospectiva antes de enviarlas; cada acción queda registrada con su justificación. Sin el modelo HAD, estas propiedades no existen por defecto en los marcos actuales.

### 3.3. Tres Formas de Agencia

El principio de agencia de HDI debe descomponerse en tres formas dentro de HAD. La agencia humana sobre datos engloba la capacidad del usuario para comprender, controlar y negociar

prácticas de datos pese a la intermediación del agente (derechos de decisión, anulación, revocabilidad de la delegación y visibilidad de las acciones). La agencia operativa del agente captura su capacidad para razonar y actuar autónomamente dentro de un ámbito delegado (límites de autonomía, rango de capacidades, criterios de decisión y adaptación). La agencia colectiva distribuida se refiere a las prácticas emergentes de la coordinación multi-agente, donde ningún actor individual determina por completo los resultados (Duan et al., 2025).

Equilibrar estas formas crea tensiones inherentes: maximizar la agencia humana restringe la autonomía del agente y limita la eficacia; maximizar la autonomía erosiona el control humano; y habilitar la agencia colectiva introduce imprevisibilidad que puede chocar con los derechos individuales. Resolver estas tensiones exige modelos de gobernanza que preserven los valores humanos mientras aprovechan las capacidades de los agentes.

### 3.4. Configuraciones Epistemológicas

Sobre el marco de Coeckelbergh (2025) acerca de la agencia epistémica, identificamos tres desafíos. La opacidad epistémica genera una brecha entre la intención del usuario y el procesamiento de datos: el sujeto conoce el resultado (what) pero no el proceso lógico-causal subyacente (how/why). La gobernanza de la formación de creencias surge porque los agentes moldean la arquitectura cognitiva del usuario al preseleccionar y jerarquizar la información, erosionando su control sobre los propios procesos deliberativos. Y la dilución de la responsabilidad epistémica aparece cuando la producción colectiva de conocimiento mediante agentes distribuidos fragmenta la autoría intelectual y dificulta la trazabilidad ética y epistemológica.

### 3.5. Fenomenología de la Experiencia

La mediación por agentes transforma la experiencia vivida de la interacción con datos (Limerick et al., 2014). La intencionalidad se vuelve delegada: los usuarios pretenden que algo se haga en lugar de cómo hacerlo, creando una distancia entre intención y acción que reduce el sentido de propiedad. La mediación efectiva puede requerir una opacidad productiva —ocultar detalles para permitir la concentración en objetivos—, lo que tensiona con la exigencia de legibilidad; el diseño debe ofrecer la información relevante bajo demanda sin interrumpir el flujo de trabajo. Delegar el control crea vulnerabilidad, pues el agente podría actuar contra los intereses del usuario (Hancock et al., 2020); la confianza se convierte en el eje y se construye, erosiona y reconstruye a través de cada interacción.

## 4. INTERSECCIONES HDI–AGENTE

### 4.1. Legibilidad en Arquitecturas Multi-Agente

Proponemos cuatro dimensiones para evaluar la legibilidad de los sistemas agénticos. La legibilidad intencional refiere a la capacidad de discernir los objetivos de cada agente; la legibilidad del comportamiento, a la predictibilidad de las acciones mediante señales observables, armonizando flujos deterministas con agentes LLM estocásticos; la legibilidad operativa, a la comprensión de los procedimientos de ejecución, incluida la selección de herramientas; y la legibilidad de coordinación, a la inteligibilidad de las interacciones inter-agenciales mediante delegación, estados compartidos o invocaciones explícitas.

### 4.2. Agencia y Autonomía Delegada

La delegación plantea una tensión fundamental: el beneficio de la autonomía requiere ceder control, mientras que HDI exige preservarlo. Proponemos cinco estadios de delegación graduada, adaptados de Luck y d'Inverno (1995): (1) ejecución supervisada, con intervención humana directa y persistente;

(2) autonomía supervisada, condicionada a validación humana en operaciones críticas; (3) autonomía acotada, restringida a marcos predefinidos de acceso; (4) autonomía condicional, sujeta a revocación dinámica de permisos; y (5) delegación completa, con mínimas restricciones.

### 4.3. Dimensiones Temporales de la Transparencia

Distinguimos tres modalidades temporales de transparencia: retrospectiva, que reconstruye la causalidad de acciones pretéritas mediante registros de procedencia; prospectiva, que comunica acciones proyectadas, riesgos y alternativas antes de su ejecución; y continua, que ofrece conciencia situacional persistente mediante indicadores de progreso. Para ser efectiva, la transparencia debe ser multimodal: narrativas en lenguaje natural para perfiles no técnicos, trazas de ejecución para desarrolladores y modelos causales para auditoría.

### 4.4. Gobernanza Colectiva y Responsabilidad Distribuida

Los sistemas multi-agente introducen desafíos de gobernanza que trascienden la escala individual (Dastani et al., 2005). La emergencia de comportamientos colectivos puede generar beneficios sistémicos, pero suele involucrar datos individuales sin consentimiento granular, y la fragmentación del proceso entre múltiples agentes diluye la cadena de mando. Por ello la gobernanza debe fundamentarse en la trazabilidad de los flujos de información y en la capacidad de propagar restricciones de permisos de forma holística por toda la red agencial.

## 5. IMPLEMENTACIÓN DE HDI EN ADK

### 5.1. Siete Principios de Diseño Centrados en el Usuario

En sintonía con Amershi et al. (2019), destilamos siete principios de diseño de interacción alineados con HDI:

**P1 — Legibilidad por Diseño.** descripciones estandarizadas de agentes, registro automático de operaciones, captura obligatoria de justificaciones para decisiones no deterministas y explicaciones jerárquicas.

**P2 — Delegación Graduada.** niveles de autonomía configurables con flujos de aprobación, restricciones de alcance y límites temporales.

**P3 — Uso Transparente de Herramientas.** notificaciones previas para operaciones sensibles, explicaciones de justificación y resúmenes de resultados.

**P4 — Permisos Negociables.** permisos dinámicos, modificables en tiempo de ejecución, propagados por las jerarquías y sujetos a negociación bidireccional.

**P5 — Coordinación Auditable.** registros completos de eventos, visualización de cadenas de delegación y detección de patrones emergentes.

**P6 — Responsabilidad Epistémica.** procedencia del conocimiento rastreable mediante atribución de fuentes, documentación de razonamiento e indicadores de confianza.

**P7 — Interfaces de Gobernanza Colectiva.** configuración de políticas a nivel de sistema, descubrimiento de agentes y controles de emergencia.

### 5.2. Extensiones Arquitectónicas

Proponemos cuatro extensiones a ADK: el Objeto de Contexto HDI, que extiende InvocationContext con el nivel de delegación, dominios de datos sensibles y límites de permisos propagados por las jerarquías; la Capa de Legibilidad, con capacidades de explicar acción, visualizar jerarquía, trazar decisión y predecir comportamiento; el Gestor de Delegación, que verifica si se necesita aprobación y

gestiona la negociación bidireccional ante solicitudes de acceso ampliado; y el Sistema de Auditoría, que registra acciones con justificaciones e invocaciones de herramientas con sus parámetros y resultados.

### 5.3. Marco de Evaluación y Métricas

Proponemos cuatro categorías de métricas: de legibilidad (cobertura de explicaciones, comprensión del usuario, precisión de predecibilidad); de agencia (preservación del control, granularidad de la delegación, tasa de éxito en negociaciones); de responsabilidad (completitud de auditoría, trazabilidad de procedencia); y a nivel de sistema (detección de comportamiento emergente, transparencia multi-agente, indicadores de confianza).

### 5.4. Protocolo de Evaluación Piloto con Usuarios

Para operacionalizar las métricas, diseñamos un protocolo piloto con un diseño entre sujetos y tres condiciones: (C1) agente con comportamiento por defecto de ADK; (C2) sistema con P1–P3 (legibilidad, delegación graduada, herramientas transparentes); y (C3) sistema con P1–P7 completos. La tarea ecológica —"gestiona mis documentos de proyecto usando el agente"— se aplica a n = 20 participantes por condición (N = 60), profesionales del conocimiento con experiencia variable en IA. Las medidas primarias son la comprensión de las acciones del agente (test post-tarea, umbral ≥ 70 %), la tasa de control preservado (anulaciones o modificaciones exitosas) y la conciencia situacional (SART adaptado a HDI); como medidas secundarias se incluyen la confianza calibrada (HCTS), el tiempo hasta recuperar el control y el número de errores de delegación no detectados.

## 6. DISCUSIÓN

### 6.1. Comparación entre Plataformas de Agentes

La elección de ADK como marco de referencia requiere justificación. La Tabla 1 compara ADK, LangChain y AutoGen en las dimensiones más relevantes para el modelo HAD.

**Tabla 1.** Comparativa de plataformas de agentes respecto a dimensiones HAD.

| Dimensión | ADK (Google) | LangChain | AutoGen (MS) |
|---|---|---|---|
| Transparencia nativa | Trazas básicas de ejecución | LangSmith (externo) | Sin mecanismo nativo |
| Delegación graduada | No explícita | No explícita | Parcial (human-in-the-loop) |
| Gobernanza multi-agente | Jerárquica (AgentTeam) | Parcial (LCEL) | Convencional (GroupChat) |
| Documentación de herramientas | Alta (tool schemas) | Alta | Media |
| Adopción en producción | Alta (Agentspace, Vertex AI) | Muy alta (ecosistema amplio) | Alta (Microsoft 365 Agents) |
| Alineación HDI explícita | Ninguna | Ninguna | Ninguna |

Los tres frameworks presentan brechas equivalentes en alineación HDI explícita, lo que refuerza la generalizabilidad del modelo HAD más allá de ADK. La elección de ADK se justifica por su composición jerárquica explícita —que facilita el Objeto de Contexto HDI—, su documentación de herramientas de alta calidad (compatible con P3) y su creciente adopción en producción empresarial.

### 6.2. Implicaciones para el Diseño de Sistemas

Las tensiones identificadas tienen consecuencias prácticas. La tensión entre autonomía y control implica que no se puede maximizar la eficiencia del agente sin considerar la agencia del usuario: un

agente que opera como caja negra viola los principios de HDI y erosiona la confianza. La delegación graduada ofrece una vía intermedia, adaptándose al contexto. Además, los siete principios forman un sistema interrelacionado: la legibilidad por diseño (P1) habilita la delegación graduada (P2), que requiere herramientas transparentes (P3) y permisos negociables (P4), mientras que la coordinación auditable (P5) y la responsabilidad epistémica (P6) sustentan las interfaces de gobernanza colectiva (P7).

### 6.3. Implicaciones Regulatorias y Contexto Europeo

El RGPD se fundamenta en un consentimiento informado que asume interacción directa entre el sujeto y el controlador; la introducción de agentes autónomos exige transformarlo en un proceso continuo y negociado, precisamente lo que habilita nuestro principio de permisos negociables (P4). El AI Act converge con nuestro análisis: su enfoque basado en riesgo se alinea con la delegación graduada, donde mayor riesgo implica mayor supervisión. En España, la ENIA 2.0 sitúa como eje una IA centrada en las personas, alineada con la protección de datos y la transparencia; los principios de legibilidad operativa y permisos negociables operacionalizan ese compromiso con la IA explicable y la soberanía digital. El contexto iberoamericano, con estructuras colectivistas y robustas regulaciones de datos, hace especialmente relevante la "negociabilidad", donde el control puede extenderse a comunidades de usuarios.

### 6.4. Limitaciones

Reconocemos varias limitaciones. La propuesta carece todavía de validación empírica, aunque el protocolo de la Sección 5.4 constituye la hoja de ruta inmediata. El análisis arquitectónico se centra en ADK, si bien los principios HAD son transferibles a cualquier framework, pues las brechas HDI son equivalentes. Las métricas requieren operacionalización detallada, que abordará el estudio piloto. Por último, la captura exhaustiva de justificaciones puede introducir sobrecargas de rendimiento que comprometan la usabilidad.

## 7. FUTURAS LÍNEAS DE INVESTIGACIÓN

La prioridad inmediata es ejecutar el protocolo de evaluación para validar empíricamente los principios HAD, complementado con el prototipado de baja fidelidad de la Capa de Legibilidad y el Gestor de Delegación (wireframes o mockups) para evaluar la comprensibilidad del modelo con usuarios reales. En paralelo, estudios fenomenológicos (entrevistas experienciales y diario), pueden iluminar cómo la delegación afecta al sentido de propiedad sobre las prácticas de datos.

A más largo plazo, conviene examinar las dinámicas de poder que la autonomía de los agentes redistribuye (Mahroof et al., 2025); la gobernanza colectiva mediante cooperativas de datos y bienes comunes digitales (Duncan, 2023); las perspectivas transculturales, dado que los marcos actuales de HDI reflejan suposiciones predominantemente occidentales; la HDI temporal en agentes de larga ejecución (legibilidad sostenida, permisos evolutivos, retención de memoria); y la integración con la seguridad de IA, mediante verificación formal de restricciones HDI y monitorización de violaciones.

## 8. CONCLUSIONES

Este artículo ha desarrollado fundamentos teóricos y un modelo de diseño de interacción para integrar los principios de Interacción Humano-Datos en arquitecturas de agentes IA, tomando ADK como referencia. Nuestro argumento central es que los sistemas agénticos introducen un cambio cualitativo en las relaciones humano-datos que exige más que ajustes incrementales a los marcos existentes. Hemos contribuido con un modelo triádico HAD (Figura 1) con sus tres formas de agencia

y seis vías de interacción; un escenario ilustrativo; siete principios de diseño con extensiones arquitectónicas concretas para ADK; una comparativa de plataformas que demuestra la transferibilidad del modelo; un protocolo de evaluación piloto que operacionaliza las métricas; y un análisis del encuadre regulatorio europeo (RGPD, AI Act y ENIA 2.0).

El análisis revela tensiones que no pueden eliminarse, sino navegarse: entre autonomía y control, transparencia y abstracción eficiente, derechos individuales e inteligencia colectiva. Sin una integración deliberada de los principios HDI, los marcos de agentes en producción corren el riesgo de reproducir los problemas que HDI fue creada para resolver. Al fundamentar el diseño de agentes en la teoría HDI y en la experiencia de usuario, podemos realizar la promesa de los agentes autónomos preservando la agencia humana, la dignidad y un control significativo sobre nuestro futuro con los datos.

**Agradecimientos:** Este trabajo ha sido desarrollado en el marco del Grupo de investigación ISE de la Universidad de Castilla-La Mancha.

## REFERENCIAS

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. En *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)* (pp. 1–13). Association for Computing Machinery.

Chaffer, T. J., Goldston, J., & Sebastian, G. (2024). On the ethos of AI agents: An ethical technology and holistic oversight system. arXiv. https://doi.org/10.48550/arXiv.2412.17114

Chakraborti, T., Kulkarni, A., Sreedharan, S., Smith, D. E., & Kambhampati, S. (2018). Explicability? Legibility? Predictability? Transparency? Privacy? Security? The emerging landscape of interpretable agent behavior. arXiv. https://doi.org/10.48550/arXiv.1811.09722

Coeckelbergh, M. (2025). AI and epistemic agency: How AI influences belief revision and its normative implications. *Social Epistemology*. Advance online publication.

Crabtree, A., & Mortier, R. (2015). Human data interaction: Historical lessons from social studies and CSCW. En *Proceedings of the European Conference on Computer Supported Cooperative Work (ECSCW)*.

Dastani, M., Arbab, F., & de Boer, F. (2005). Coordination and composition in multiagent systems. En *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '05)* (pp. 439–446). Association for Computing Machinery.

Duan, W., Lu, J., & Xuan, J. (2025). Multi-agent coordination across diverse applications: A survey. arXiv. https://doi.org/10.48550/arXiv.2502.14743

Duncan, J. (2023). Data protection beyond data rights: Governing data production through collective intermediaries. *Internet Policy Review*, *12*(3).

Durango, I., Gallud, J. A., & Penichet, V. M. R. (2024a). Human-data interaction framework: A comprehensive model for a future driven by data and humans. arXiv. https://doi.org/10.48550/arXiv.2407.21010

Durango, I., Gallud, J. A., & Penichet, V. M. R. (2025). The data dance: Choreographing seamless partnerships between humans, data, and GenAI. *International Journal of Data Science and Analytics*, *20*(4), 3613–3640.

Durango, I., Penichet, V., & Gallud, J. A. (2024b). Exploring human-data interaction: An AI-enhanced systematic mapping. *Universal Access in the Information Society*, *24*, 1059–1076.

Google. (2025). *Agent Development Kit documentation*. https://google.github.io/adk-docs/

Google Developers. (2025). Agent Development Kit: Easy to build multi-agent applications. *Google Developers Blog*. https://developers.googleblog.com/en/agent-development-kit-easy-to-build-multi-agent-applications/

Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., & Zhang, X. (2024). Large language model based multi-agents: A survey of progress and challenges. arXiv. https://doi.org/10.48550/arXiv.2402.01680

Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, *25*(1), 89–100.

Jackson, R. B., & Williams, T. (2021). A theory of social agency for human-robot interaction. *Frontiers in Robotics and AI*, *8*.

Limerick, H., Coyle, D., & Moore, J. W. (2014). The experience of agency in human-computer interactions: A review. *Frontiers in Human Neuroscience*, *8*, 643.

Luck, M., & d'Inverno, M. (1995). A formal framework for agency and autonomy. En *Proceedings of the First International Conference on Multiagent Systems (ICMAS '95)* (pp. 254–260). AAAI Press.

Luo, J., Zhang, W., Yuan, Y., Zhao, Y., Yang, J., Gu, Y., Wu, B., Chen, B., Qiao, Z., Long, Q., Tu, R., Luo, X., Ju, W., Xiao, Z., Wang, Y., Xiao, M., Liu, C., Yuan, J., Zhang, S., … Zhang, M. (2025). Large language model agent: A survey on methodology, applications and challenges. arXiv. https://doi.org/10.48550/arXiv.2503.21460

Mahroof, K., Weerakkody, V., & Hussain, Z. (2025). Navigating power dynamics in the public sector through AI-driven algorithmic decision-making. *Government Information Quarterly*. Advance online publication.

Mortier, R., Haddadi, H., Henderson, T., McAuley, D., & Crowcroft, J. (2014). Human-data interaction: The human face of the data-driven society. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2508051

Mortier, R., Haddadi, H., Henderson, T., McAuley, D., & Crowcroft, J. (2015). Human-data interaction. En M. Soegaard & R. F. Dam (Eds.), *The encyclopedia of human-computer interaction* (2.ª ed.). Interaction Design Foundation.

Onyeukaziri, J. N. (2023). Action and agency in artificial intelligence: A philosophical critique. *Philosophia: International Journal of Philosophy*, *24*(1), 74–90.

South, T., Marro, S., Hardjono, T., Mahari, R., Deslandes Whitney, C., Greenwood, D., Chan, A., & Pentland, A. (2025). Authenticated delegation and authorized AI agents. arXiv. https://doi.org/10.48550/arXiv.2501.09674

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74–88.

Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y., & Wang, Q. (2020). From human-human collaboration to human-AI collaboration: Designing AI systems that can work together with people. En *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)* (pp. 1–6). Association for Computing Machinery.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.-R. (2023). A survey on large language model based autonomous agents. arXiv. https://doi.org/10.48550/arXiv.2308.11432

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., … Gui, T. (2023). The rise and potential of large language model based agents: A survey. arXiv. https://doi.org/10.48550/arXiv.2309.07864

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. En *Proceedings of the International Conference on Learning Representations (ICLR)*.