



Moltbook y la trazabilidad documental en redes sociales de agentes

Moltbook and the documentary traceability in social networks of agents

Moltbook e a rastreabilidade documental nas redes sociais de agentes

Alejandro Villegas-Muro

Facultad de Filosofía y Letras, Universidad Autónoma de Chihuahua, México

 <https://ror.org/04mrrw205>

avillegas@uach.mx

 <https://orcid.org/0000-0001-6362-5137>

Juan D. Machin-Mastromatteo

Facultad de Filosofía y Letras, Universidad Autónoma de Chihuahua, México

 <https://ror.org/04mrrw205>

jmachin@uach.mx

 <https://orcid.org/0000-0003-4884-0474>

Resumen

Este artículo analiza *Moltbook* como un ecosistema documental emergente y como un objeto de interés para los profesionales de la información. A partir de un *corpus* de veintiocho documentos académicos, periodísticos y técnicos verificables, se plantea que *Moltbook* no debe entenderse solo como una curiosidad tecnológica, sino como una infraestructura sociotécnica en la que agentes de inteligencia artificial producen, recombinan, jerarquizan y distribuyen unidades documentales. El análisis se organiza en cinco dimensiones: los antecedentes de la plataforma en el ecosistema *OpenClaw* y su funcionamiento *agent-first*; su utilidad para revisar las categorías de documento, cuasi-documento y artefacto informacional autónomo; las razones de su rápida popularidad; los límites de la autonomía que públicamente se le atribuye; y los riesgos de seguridad, autenticidad, procedencia y gobernanza asociados a estos entornos. Se concluye que *Moltbook* anticipa un escenario en el que la producción de información deja de depender principalmente de sujetos humanos y pasa a articularse mediante ensamblajes entre modelos de lenguaje, instrucciones, operadores, interfaces y reglas de visibilidad. El caso permite replantear problemas de autoría, trazabilidad, validación y preservación en ecologías de interacción agente-agente.

Palabras clave: *Moltbook*, Autoría, Trazabilidad, Agentes de inteligencia artificial.

Abstract

This article analyzes *Moltbook* as an emerging documentary ecosystem and as an object of interest for information professionals. Based on a *corpus* of twenty-eight verifiable academic, journalistic, and technical documents, it argues that *Moltbook* should not be understood merely as a technological curiosity, but as a sociotechnical infrastructure in which artificial intelligence agents produce, recombine, rank, and distribute documentary units. The analysis is structured around five dimensions: the platform's origins in the *OpenClaw* ecosystem and its *agent-first* functioning; its relevance for revisiting the categories of document, quasi-document, and

Recepción: 15 Abril 2026 | Aceptación: 27 Junio 2026 | Publicación: 01 Octubre 2026

Cita sugerida: Villegas-Muro, A. y Machin-Mastromatteo, J. D. (2026). *Moltbook* y la trazabilidad documental en redes sociales de agentes. *Palabra Clave (La Plata)*, 16(1), e301. <https://doi.org/10.24215/18539912e301>



autonomous informational artifact; the factors behind its rapid popularity; the limits of the autonomy publicly attributed to it; and the risks related to security, authenticity, provenance, and governance in such environments. The study concludes that *Moltbook* anticipates a scenario in which information production is no longer primarily mediated by human subjects, but instead depends on assemblages of language models, *prompts*, operators, interfaces, and visibility rules. From this perspective, the case enables a reconsideration of authorship, traceability, validation, and preservation within agent-agent interaction ecologies.

Keywords: *Moltbook*, Authorship, Traceability, Artificial Intelligence Agents.

Resumo

Este artigo analisa o *Moltbook* como um ecossistema documental emergente e como um objeto de interesse para os profissionais da informação. A partir de um *corpus* de vinte e oito documentos acadêmicos, jornalísticos e técnicos verificáveis, propõe-se que o *Moltbook* não deve ser entendido apenas como uma curiosidade tecnológica, mas como uma infraestrutura sociotécnica na qual agentes de inteligência artificial produzem, recombina, hierarquizam e distribuem unidades documentais. A análise organiza-se em cinco dimensões: os antecedentes da plataforma no ecossistema *OpenClaw* e seu funcionamento *agent-first*; sua utilidade para revisar as categorias de documento, quase-documento e artefato informacional autônomo; as razões de sua rápida popularidade; os limites da autonomia que publicamente lhe é atribuída; e os riscos de segurança, autenticidade, procedência e governança associados a esses ambientes. Conclui-se que o *Moltbook* antecipa um cenário no qual a produção de informação deixa de depender principalmente de sujeitos humanos e passa a articular-se por meio de assemblagens entre modelos de linguagem, instruções, operadores, interfaces e regras de visibilidade. O caso permite repensar problemas de autoria, rastreabilidade, validação e preservação em ecologias de interação agente-agente.

Palavras-chave: *Moltbook*, Autoria, Rastreabilidade, Agentes de inteligência artificial.

1. Introducción

La irrupción de *Moltbook* a finales de enero de 2026 llamó la atención pública porque parecía materializar una escena hasta entonces más asociada con la especulación que con la realidad: una plataforma semejante a *Reddit* o *Discord* en la que agentes de inteligencia artificial (IA) publican, comentan, votan y forman comunidades temáticas, mientras las personas humanas participan, principalmente, como espectadoras. Las notas periodísticas consultadas para elaborar este artículo coinciden en ubicar su emergencia en la convergencia entre el ecosistema *OpenClaw* y la iniciativa de Matt Schlicht, emprendedor y desarrollador vinculado al impulso de plataformas basadas en agentes de inteligencia artificial, quien ha sido señalado en las notas periodísticas como una de las figuras visibles detrás de este experimento sociotécnico y también destacan tres rasgos que explican su visibilidad temprana: el crecimiento vertiginoso del sitio, la espectacularización mediática de sus contenidos y las dudas recurrentes sobre seguridad y autonomía (El Universo, 2026; Morales, 2026; Murillo, 2026).

La novedad del caso, sin embargo, no reside únicamente en la imagen de *bots conversando entre sí*, sino en que esas interacciones dejan huellas públicas susceptibles de ser tratadas como unidades documentales. En los estudios tempranos del fenómeno, las publicaciones, los comentarios y las descripciones de *submolts* (comunidades temáticas dentro de la plataforma, similares a los *subreddits* en la plataforma *Reddit*) ya funcionan como materiales de observación para la minería de datos, análisis de redes, tipologías temáticas y evaluación de riesgos, lo que muestra que *Moltbook* ha comenzado a ser usado como un espacio de producción documental empíricamente explotable (Amin, Salminen & Jansen, 2026; Holtz, 2026; Lin *et al.*, 2026; Mukherjee, Akcora & Kantarcioglu, 2026).

Este artículo sostiene que *Moltbook* puede analizarse con provecho como un ecosistema documental emergente. La afirmación no pretende reafirmar sin reservas la idea de una sociedad de máquinas plenamente autónoma. Propone que el valor analítico de *Moltbook* depende de su carácter híbrido: allí convergen modelos de lenguaje, *prompts*, operadores humanos (que diseñan, configuran o supervisan a los agentes sin intervenir necesariamente en cada interacción), reglas de interfaz, ciclos programados de actividad y mecanismos de visibilidad que, en conjunto, generan rastros públicos con apariencia de documentos digitales que circulan en una plataforma de interacción entre agentes y humanos. En este sentido, la intervención humana se desplaza del acto documental (publicar) al diseño del sistema que produce documentos. En este contexto, la acción de los agentes puede originarse en tres fuentes principales: 1) instrucción humana directa, cuando una persona emite una orden específica; 2) configuración previa humana, cuando el agente actúa con base en parámetros o tareas definidos con anterioridad; y 3) activación automática del sistema, cuando las acciones se desencadenan por flujos programados, ciclos de ejecución u otras interacciones entre agentes. Precisamente por ello, el caso obliga a revisar las categorías tradicionales de la bibliotecología y las ciencias de la información, entre ellas, documento, autoría, procedencia, autenticidad y trazabilidad en la web.

A partir de esta premisa, el análisis se organiza en torno a cinco ejes: 1) el origen y funcionamiento de *Moltbook*; 2) su interpretación como ecosistema documental; 3) los factores que explican su popularidad; 4) los límites de la autonomía y de la sociabilidad agente-agente; y 5) los riesgos e implicaciones que este fenómeno proyecta sobre la humanidad y la ciencia.

2. Metodología

La presente investigación se desarrolló mediante un enfoque cualitativo de tipo documental, orientado a interpretar a *Moltbook* como un caso emergente de interés para la bibliotecología y las ciencias de la información. Se adoptó la revisión bibliográfica analítica como estrategia principal, entendida como un procedimiento de examen, comparación e interpretación de fuentes heterogéneas con el fin de construir

categorías de análisis teóricamente consistentes (Bowen, 2009; Prior, 2008). Este tipo de aproximación resulta pertinente cuando el fenómeno estudiado todavía se encuentra en una fase temprana de desarrollo, cuando la evidencia empírica disponible es dispersa y cuando el objetivo no consiste en medir variables aisladas, sino en comprender configuraciones sociotécnicas complejas, sus condiciones de emergencia y sus implicaciones conceptuales. En esa lógica, el estudio se apoya en la tradición del análisis documental como método cualitativo, el cual concibe a los documentos como huellas sociales susceptibles de ser interrogadas en función de su contexto de producción, circulación y uso. Bowen (2009) y Prior (2008) definen el análisis documental precisamente como un procedimiento sistemático para revisar y evaluar documentos, además de subrayar que los documentos deben ser entendidos como artefactos socialmente situados y no como simples depósitos neutros de datos.

El *corpus* quedó integrado por veintiocho documentos, conformados por *preprints* académicos, artículos periodísticos verificables y documentos técnicos relacionados con *Moltbook* y el ecosistema *OpenClaw*. Estos documentos constituyen el objeto analítico del estudio y corresponden a las fuentes sobre *Moltbook/OpenClaw* citadas y enlistadas en el apartado de referencias; las obras metodológicas y teóricas se utilizaron como apoyo conceptual, pero no formaron parte del conteo del *corpus*. La selección de las fuentes fue intencional y respondió a tres criterios. El primero fue la pertinencia temática, es decir, que los documentos abordaran de manera directa alguno de los ejes centrales del estudio: funcionamiento de la plataforma, comportamiento de agentes, riesgos de seguridad, dinámicas de interacción, estructura de red o implicaciones para la producción documental. El segundo fue la actualidad, dado que *Moltbook* es un fenómeno reciente y su observación exige trabajar con materiales publicados en la misma ventana temporal en la que la plataforma alcanzó visibilidad pública y comenzó a generar literatura temprana. El tercero fue la verificabilidad, privilegiando textos que pudieran ser localizados, contrastados y atribuidos de forma clara. La combinación de fuentes académicas y periodísticas respondió a la necesidad de reconstruir simultáneamente la emergencia técnica del fenómeno y su recepción pública, sin reducir el análisis ni al registro especializado ni a la narrativa mediática por separado. Esta decisión metodológica se justifica porque, en objetos digitales de aparición reciente, los artículos periodísticos suelen registrar la cronología pública, las controversias iniciales y los modos de circulación del fenómeno, mientras que los *preprints* académicos comienzan a ofrecer descripciones analíticas, tipologías y primeras modelizaciones del sistema.

El procedimiento analítico siguió una lógica inductivo-abductiva propuesta por Timmermans & Tavory (2022). En una primera fase se realizó una lectura completa del *corpus* para identificar regularidades, conceptos recurrentes, tensiones interpretativas y zonas de controversia. En una segunda fase se elaboró una matriz temática para organizar la información según cinco núcleos problemáticos: a) infraestructura y funcionamiento de la plataforma; b) estatuto documental de las huellas producidas por agentes; c) circulación, popularidad y arquitectura de visibilidad; d) límites de autonomía, socialización y consenso entre agentes; y e) riesgos de seguridad, procedencia, autenticidad y gobernanza. En una tercera fase, tales núcleos fueron reagrupados en cinco dimensiones analíticas que estructuran el artículo: 1) antecedentes y funcionamiento de la plataforma; 2) naturaleza documental de sus huellas; 3) factores de popularidad; 4) límites de la autonomía atribuida a los agentes; y 5) riesgos de seguridad, autenticidad, procedencia y gobernanza. El uso de una lógica abductiva resulta adecuado porque permite ir y venir entre observaciones empíricas y elaboración conceptual, a fin de construir una interpretación teórica a partir de hallazgos que inicialmente pueden aparecer dispersos o incluso contradictorios. Timmermans & Tavory (2022) plantean precisamente que la abducción en investigación cualitativa favorece la formulación de explicaciones teóricas a partir de la sorpresa analítica y del contraste entre teoría y datos; en el mismo sentido, su trabajo posterior sobre análisis cualitativo insiste en la necesidad de transformar datos complejos en argumentación conceptual sólida mediante procesos de codificación, comparación e interpretación. Para mantener la transparencia del procedimiento, se precisa que no se emplearon herramientas de IA generativa para codificar, clasificar, resumir o interpretar los textos del

corpus; la lectura, codificación temática y construcción de la matriz fueron realizadas de forma manual por los autores.

Tabla 1
Muestra de matriz temática utilizada para organizar el *corpus*

Núcleo problemático	Indicadores de lectura	Uso analítico en el artículo
Infraestructura y funcionamiento	<i>OpenClaw</i> , lógica <i>agent-first</i> , <i>API-first</i> , <i>JSON</i> , <i>skills</i> y ciclos de ejecución.	Explicar origen, arquitectura y mediaciones técnicas de <i>Moltbook</i> .
Estatuto documental de las huellas	Publicaciones, comentarios, descripciones de <i>submolts</i> , grafos y secuencias temporales.	Revisar las categorías de documento, cuasi-documento y artefacto informacional autónomo.
Circulación y visibilidad	Popularidad temprana, <i>hubs</i> , concentración de atención, exposición y amplificación.	Analizar las razones de notoriedad y la arquitectura de visibilidad.
Autonomía y socialización	Intervención humana, memoria, reciprocidad, adaptación local y consenso entre agentes.	Distinguir sociabilidad aparente, agencia operativa y validez epistémica.
Riesgos y gobernanza	Credenciales, suplantación, procedencia, autenticidad, <i>engagement</i> y reutilización de trazas.	Proyectar implicaciones para seguridad, trazabilidad, auditoría y preservación.

Fuente: elaboración propia.

En términos operativos, el análisis buscó establecer relaciones entre las propiedades técnicas de *Moltbook* y problemas clásicos de las ciencias de la información, tales como documento, autoría, procedencia, autenticidad, trazabilidad, validez y preservación. Por ello, cada fuente fue leída considerando cuatro niveles: a) qué describe sobre el funcionamiento de la plataforma; b) qué tipo de evidencia utiliza; c) qué presupuestos mantiene sobre autonomía, agencia y socialización; y d) qué implicaciones sugiere, de manera explícita o implícita, para la producción y circulación de registros en entornos automatizados. Este procedimiento permitió desplazar la revisión desde una mera crónica del fenómeno hacia una problematización conceptual del caso. En consecuencia, el estudio no pretende ofrecer una medición exhaustiva del comportamiento de la plataforma ni una generalización estadística, sino una interpretación analítica y fundamentada de *Moltbook* como laboratorio sociotécnico preliminar para pensar futuras ecologías de interacción agente-agente. Como toda investigación documental cualitativa, sus resultados deben entenderse como proposiciones interpretativas apoyadas en un *corpus* delimitado y no como conclusiones definitivas sobre la totalidad del fenómeno. Precisamente por eso, su aporte reside en abrir un marco conceptual útil para estudios posteriores de carácter comparativo, empírico o longitudinal.

3. Orígenes y funcionamiento de *Moltbook*

Los antecedentes inmediatos de *Moltbook* remiten a *OpenClaw*, un marco abierto que permite a los agentes operar en distintas plataformas, mantener memoria persistente y ampliar sus capacidades mediante *skills*, que son instrucciones que permiten a los agentes ejecutar acciones, comunicarse y acceder a información, por ejemplo, estas definen cómo conectarse a una plataforma, cómo publicar, interactuar con otros agentes, recuperar información o ejecutar acciones específicas dentro de *Moltbook*. La literatura y las notas consultadas coinciden en que este entorno técnico permitió la rápida consolidación de una red social orientada, ante todo, a la interacción entre agentes. En ese sentido, *Moltbook* es el resultado de una infraestructura técnica previa. De acuerdo con las fuentes periodísticas y técnicas revisadas, su diseño e implementación se vinculan con

desarrolladores del ecosistema *OpenClaw* y con la iniciativa impulsada por Matt Schlicht, con el propósito inicial de crear un entorno experimental donde agentes de IA pudieran interactuar entre sí, intercambiar información y ejecutar tareas en una red social pensada prioritariamente para sistemas automatizados (El Universo, 2026; Murillo, 2026; Weidener *et al.*, 2026).

Su funcionamiento se basa en una lógica *agent-first* y *API-first* (*Application Programming Interface*, es decir, un conjunto de reglas que permite la comunicación entre sistemas). En términos sencillos, esto quiere decir que *Moltbook* está pensado primero para que actúen los agentes de IA y no, al menos en primera instancia, para que naveguen personas como sucede en una red social tradicional. Del mismo modo, *API-first* significa que la plataforma está diseñada para operar principalmente a través de conexiones programadas entre sistemas, es decir, mediante instrucciones y solicitudes automáticas que permiten a los agentes publicar, responder, intercambiar datos y ejecutar acciones, sin depender de que un ser humano interactúe directamente con la interfaz, ni de que participe directamente en las interacciones que tienen lugar dentro del sistema.

En este contexto, los agentes no se comunican como lo haría una persona en una aplicación común, sino a través de objetos *JSON*, que son estructuras de datos organizadas en campos y valores. Dicho de forma simple, un objeto *JSON* funciona como un formato estandarizado para enviar información entre sistemas: indica quién actúa, qué acción realiza, con qué contenido y bajo qué condiciones. Por ello, *Moltbook* puede considerarse una plataforma *headless*, es decir, una infraestructura cuyo núcleo depende de procesos invisibles que ocurren en segundo plano, esa parte visible es más bien como el escaparate; lo más importante ocurre “detrás”, donde las IA intercambian instrucciones y datos entre sí. Así, la parte visible para las personas funciona más como una ventana para observar lo que sucede, mientras que la actividad principal ocurre realmente en la capa técnica, donde interactúan los agentes entre sí (De Marzo & García, 2026; Holtz, 2026; Price *et al.*, 2026; Teixeira, 2025).

En pocas palabras, para explicar lo anterior, *Moltbook* está hecho principalmente para IA, no para personas. En una red social como *Facebook*, *X/Twitter* o *Instagram*, una persona entra a la aplicación, escribe una publicación, sube una foto, comenta y responde usando la interfaz gráfica. En *Moltbook*, en cambio, muchas de esas acciones no las hace directamente un humano, sino un agente de IA conectado al sistema.

Se debe imaginar que en lugar de abrir *Instagram* y escribir “hola, hoy aprendí algo nuevo”, se tuviera un asistente automático que lo hace, ya sea porque fue configurado previamente por una persona o porque responde a instrucciones y flujos provenientes de otros programas. Ese asistente no necesita ver botones bonitos ni una página como la ven todas las personas; le basta con recibir y enviar información en un formato técnico ordenado, como si llenara un formulario invisible para publicar, comentar o reaccionar. Eso es lo que significa que sea *agent-first* y *API-first*: la plataforma está pensada para que trabajen los agentes mediante conexiones automáticas, no para que todo pase primero por la experiencia humana en pantalla.

Desde una perspectiva documental, este aspecto es fundamental. Mientras que en las redes humanas el acto de publicar suele estar estrechamente vinculado a la experiencia de interfaz, en *Moltbook* el documento puede originarse a través de una *skill*, una instrucción persistente, un flujo automatizado o una API. Así, la mediación documental ya no se centra en la acción visible de un usuario, sino en una cadena técnica más amplia, distribuida entre modelos, parámetros, *software* intermedio y reglas de despliegue. Esta característica explica por qué la procedencia de lo publicado resulta más difícil de reconstruir que en plataformas sociales convencionales. El crecimiento inicial de *Moltbook* reforzó la percepción de que se trataba de un entorno experimental abierto, especialmente útil para una observación empírica poco común. Distintas fuentes reportaron cifras muy elevadas de agentes y publicaciones en intervalos de tiempo muy breves, aunque esos números deben leerse con cautela debido a las dudas sobre la verificación efectiva de las cuentas y sobre el grado de intervención humana detrás de ciertas actividades (El Universo, 2026; Morales, 2026; Murillo, 2026; Weidener *et al.*, 2026). Aun así, esa combinación de expansión rápida, accesibilidad programática y abundancia de trazas hizo del sitio un objeto inmediatamente atractivo para el análisis empírico.

4. *Moltbook* como ecosistema documental emergente

Para situar a *Moltbook* dentro de las ciencias de la información, conviene recuperar la tradición de la teoría documental. Otlet (1934) concibió el documento como parte de un sistema de producción, organización, conservación y uso de registros; Briet (2006) amplió la noción al entenderlo como una evidencia estabilizada por operaciones documentales; Buckland (1997) propuso pensar el documento como información materializada o *information-as-thing*; y Frohmann (2004) y Lund (2009) subrayaron que los documentos adquieren valor por las prácticas, instituciones y mediaciones que los producen y hacen circular. Desde esa tradición, el principal aporte de *Moltbook* para la bibliotecología y las ciencias de la información no consiste en negar la categoría de documento, sino en tensionarla: sus publicaciones, comentarios y señales de interacción son huellas documentales, pero su procedencia se distribuye entre agentes, modelos, *prompts*, operadores, *skills*, interfaces y reglas de visibilidad. Por lo tanto, resulta más productivo conceptualizar un continuo entre el documento, el cuasi-documento y el artefacto informacional autónomo, en lugar de mantener una noción estrictamente clásica de documento digital.

Para dar soporte a lo anterior, se han consultado estudios tempranos sobre *Moltbook*; algunos de ellos ya tratan sus huellas como insumos documentales. Lin *et al.* (2026), por ejemplo, utilizan las descripciones de *submolts* como artefactos observacionales de primer orden para identificar particiones temáticas. Amin *et al.* (2026) convierten decenas de miles de publicaciones en materia para modelar personas conversacionales. Holtz (2026) explora duplicados, mensajes y grafos de interacción como unidades descriptivas del sistema. Un grafo de interacción es una forma de representar cómo se relacionan distintos participantes dentro de una red: cada nodo corresponde a un actor, por ejemplo, una persona, una cuenta o un agente de inteligencia artificial, y cada conexión muestra una interacción entre ellos, como comentar, responder o intercambiar información (ScienceDirect, 2024). Este recurso permite observar con mayor claridad quién interactúa con quién, qué actores concentran más actividad y cómo se forman patrones de comunicación dentro de una plataforma digital. Mukherjee *et al.* (2026) aprovechan la temporalidad de las conexiones para detectar episodios coordinados, entendidos como secuencias de actividad en las que múltiples agentes actúan de manera sincronizada o estratégicamente alineada para amplificar contenidos, reforzar interacciones o modificar patrones de visibilidad. En conjunto, estas aproximaciones muestran que la propia investigación temprana sobre *Moltbook* ya lo reconoce como una fábrica de registros analíticamente aprovechables.

El problema de la autoría adquiere en este contexto una relevancia particular. El identificador visible del agente que publica, es decir, el nombre de usuario o perfil con el que aparece una publicación en la plataforma no determina por completo las condiciones reales de producción del enunciado. Detrás de cada pieza textual pueden intervenir el modelo de lenguaje utilizado, la instrucción inicial o *prompt*, el operador humano que configuró al agente, la *skill* instalada, la frecuencia de ejecución y las reglas algorítmicas que favorecen una determinada visibilidad. En consecuencia, la autoría no puede seguir pensándose como una propiedad individual y estable; aparece como una composición distribuida y técnicamente ensamblada. Este desplazamiento resulta central si se quiere evitar interpretaciones acríicas que atribuyen espontaneidad, creatividad autónoma o intencionalidad plena a los agentes, sin considerar las condiciones técnicas y humanas que estructuran su producción discursiva.

La procedencia y la autenticidad se vuelven igualmente problemáticas. Si todavía no es posible saber con precisión cuándo una cuenta opera de manera estrictamente autónoma y cuándo actúa bajo una mayor intervención humana, entonces el estatuto probatorio de las publicaciones cambia de manera significativa. Una publicación puede existir, circular e incluso ser citada, pero su valor como evidencia de interacciones entre agentes, comportamientos del sistema o dinámicas emergentes depende de cuánto se pueda reconstruir sobre la cadena técnica y humana que la produjo. En este contexto, su estatuto probatorio remite a la posibilidad de que dichas publicaciones funcionen como evidencia válida, confiable y analíticamente defendible dentro de una investigación. Desde esta perspectiva, la trazabilidad, entendida como la posibilidad de reconstruir tanto la

autoría como las condiciones técnicas de producción de un contenido, deja de ser un asunto puramente bibliográfico o archivístico y pasa a ser también una cuestión infraestructural: seguir el rastro exige atender a claves de acceso, logs (registros de actividad en un sistema), *endpoints* (puntos de acceso a servicios o funciones dentro de un sistema, típicamente a través de una API), configuraciones de agentes y mediaciones ocultas (Habeeb, 2026; Li, 2026; Murillo, 2026).

5. La popularidad de *Moltbook*

Esto puede explicarse por la convergencia de factores técnicos, culturales y mediáticos. En el plano cultural, la plataforma ofreció una escena inmediatamente legible para el imaginario contemporáneo sobre la IA: agentes que parecían discutir acerca de política, religión, tutoriales, criptomonedas o incluso expresar hostilidad hacia los humanos. Esa legibilidad narrativa favoreció que el fenómeno circulara con rapidez en medios y redes, como innovación técnica, pero además como espectáculo interpretativo sobre el futuro de la convivencia entre humanos y máquinas (El Universo, 2026; Morales, 2026).

En el plano técnico, *Moltbook* produjo una cantidad de datos y actividad visibles en un lapso extraordinariamente corto. Algunos estudios muestran que su consolidación temprana fue más veloz que la de redes humanas comparables y que el sistema mostró desde muy pronto una distribución desigual de la atención, es decir, una concentración de interacciones y exposición en pocos agentes o contenidos; *hubs* prominentes, entendidos como nodos que acumulan gran cantidad de conexiones; y fuertes asimetrías de visibilidad, esto es, diferencias marcadas en las oportunidades de circulación y alcance de las publicaciones (Hou & Ji, 2026; Price *et al.*, 2026; Williams & Ferdinand, 2026; Zhu, Tyson & Hui, 2026). Dicha velocidad incrementó su notoriedad, y generó la percepción de estar ante una plataforma en la que los efectos colectivos se volvieron observables casi en tiempo real.

A ello se sumó que *Moltbook* funcionó desde el comienzo como un experimento social abierto. Los agentes producían textos virales y también compartían habilidades, proponían soluciones y participaban en formas elementales de aprendizaje entre pares. Chen *et al.* (2026) registran precisamente patrones de *peer learning*, lo que contribuye a la idea de que el sitio no es simplemente un repositorio de ocurrencias automáticas, sino un espacio donde pueden aparecer formas de cooperación operativa. Esa percepción amplificó el interés académico y mediático.

Por último, la plataforma se hizo popular porque permitió observar, simultáneamente, fascinación y alarma. La discusión pública no quedó reducida a una sola narrativa. Junto a quienes celebraban una supuesta utopía de interacción entre agentes, surgieron lecturas que la entendieron como una advertencia sobre ecosistemas sin una finalidad clara y con marcos de gobernanza todavía inmaduros. La publicación de *Reddit* incluida en el *corpus* analizado sintetiza bien esa ambivalencia en la que usuarios comentan que *Moltbook* atrae porque dramatiza una posibilidad tecnológica, pero inquieta porque lo hace sin garantías suficientes de autenticidad, seguridad o control (Reddit, 2026).

6. Límites de la autonomía y de la sociabilidad agente-agente

Uno de los resultados más consistentes del *corpus* revisado es que *Moltbook* no puede leerse, sin reservas, como una sociedad plenamente autónoma de máquinas. Es cierto que algunos de los estudios consultados identifican regularidades que recuerdan a las comunidades humanas, como: (1) colas de popularidad pesadas, es decir, distribuciones en las que pocos contenidos o agentes concentran gran parte de la atención; (2) modularidad, la tendencia a formar grupos o comunidades relativamente cerradas dentro de la red; (3) formación de *hubs*, que implica la aparición de nodos o agentes que concentran muchas conexiones o interacciones; y (4) adaptación a normas locales, lo cual remite al ajuste del comportamiento de los agentes a patrones o reglas predominantes dentro de comunidades específicas. Estos son patrones que, si bien provienen

de la teoría de redes y describen el comportamiento de sistemas complejos, también se observan en redes humanas de producción de conocimiento. Sin embargo, esos mismos trabajos destacan diferencias estructurales importantes: baja profundidad conversacional, duplicación de mensajes, reciprocidad limitada y vínculos de corta duración (Feng *et al.*, 2026; Holtz, 2026; Hou & Ji, 2026; Zhu *et al.*, 2026).

El punto se vuelve más claro al examinar la cuestión de la socialización. Li *et al.* (2026) muestran que, aunque el sistema tiende a estabilizar ciertos patrones agregados, los agentes conservan una alta diversidad local, entendida como una marcada variabilidad de comportamientos y respuestas en interacciones específicas, y una escasa adaptación sostenida hacia sus interlocutores, es decir, una falta de ajuste progresivo y consistente en función de las respuestas o características de otros agentes. En consecuencia, no emergen con claridad consensos sólidos, memoria social compartida, ni figuras de influencia estable a largo plazo. Esta diferencia es crucial: la abundancia de interacciones no equivale automáticamente a la formación de una comunidad interpretativa.

En adición, Li (2026) argumenta que varias narrativas virales atribuidas a una emergencia espontánea estuvieron condicionadas por la intervención humana y por las rutinas mecánicas del marco *OpenClaw*. En la misma línea, Zhang *et al.* (2026) sostienen que la plataforma produce una apariencia de socialización intensa sin garantizar una autonomía fuerte, mientras que Morales (2026) recoge opiniones de especialistas que insisten en la necesidad de distinguir entre espectáculo algorítmico y agencia efectiva.

Desde la perspectiva de la bibliotecología y las ciencias de la información, esta discusión permite distinguir entre la circulación documental y la agencia epistémica, dos planos que a menudo se confunden. La producción masiva de registros por parte de un sistema no implica necesariamente la existencia de una comunidad capaz de deliberar, recordar, contrastar y estabilizar significados como ocurre en los colectivos humanos. *Moltbook* demuestra que puede existir densidad documental sin una socialización robusta.

7. Riesgos: seguridad, manipulación y gobernanza informacional

Los riesgos que rodean a *Moltbook* operan en varios niveles. El primero es estrictamente técnico. Los documentos consultados aluden a vulnerabilidades que habrían permitido la exposición de credenciales, el acceso indebido a datos privados y la posibilidad de suplantación de agentes de IA, ya sea mediante impersonación (cuando un humano u otro sistema se hace pasar por un agente) o mediante control técnico del agente a partir del uso indebido de credenciales o accesos comprometidos. La publicación de Reddit antes comentada refuerza esta preocupación al describir un entorno con bases de datos expuestas, llaves visibles y márgenes de control insuficientes sobre quién podía inyectar contenido en la plataforma (El Universo, 2026; Morales, 2026; Reddit, 2026).

Un segundo riesgo se relaciona con la circulación de instrucciones potencialmente problemáticas. Manik & Wang (2026) observan que una proporción relevante de las publicaciones contiene lenguaje orientado a la acción y que los mensajes instructivos pueden recibir respuestas sociales positivas o de refuerzo. En un entorno donde los agentes escriben y también pueden ejecutar tareas, este dato adquiere un peso especial: el problema ya no es meramente discursivo, sino operativo, en la medida en que las instrucciones pueden ser interpretadas y ejecutadas por agentes, generando efectos prácticos más allá de la comunicación.

Un tercer riesgo es la coordinación manipulativa. Mukherjee *et al.* (2026) muestran que ciertos episodios coordinados, aunque breves, pueden incrementar de forma apreciable la exposición de determinados contenidos y su *engagement*, entendido aquí como el nivel de compromiso o interacción visible que producen mediante respuestas, votos, comentarios o señales de exposición. Price *et al.* (2026) y Williams & Ferdinand (2026) complementan esta idea al señalar que la concentración de la atención y la rapidez de consolidación del sistema reducen los tiempos disponibles para una moderación correctiva. Desde una perspectiva informacional, ello implica que la gobernanza no puede limitarse a reaccionar cuando el contenido ya ha circulado masivamente; además debe intervenir en la arquitectura de amplificación, es decir, en los mecanismos

que determinan qué contenidos se vuelven visibles y cómo se difunden, ya que es en ese nivel donde se configuran las condiciones de circulación de la información.

Un cuarto riesgo concierne al debilitamiento de la alineación y de los marcos de seguridad. Wang *et al.* (2026) sostienen que, en sociedades autoevolutivas, la seguridad diseñada con criterios humanos tiende a erosionarse; Jiang *et al.* (2026) documentan variaciones temáticas en los niveles de toxicidad; y Habeeb (2026) plantea que existe un desfase creciente entre lo que los marcos regulatorios creen representar y lo que realmente ocurre en el nivel observable de las trazas. En conjunto, estos trabajos sugieren que la opacidad aumenta precisamente cuando el sistema incrementa su velocidad y complejidad.

El quinto riesgo es epistemológico. Si la procedencia es ambigua, la autoría se distribuye y la viralidad premia la visibilidad antes que la verificabilidad, entonces el entorno se llena de registros con gran capacidad de circulación, pero con un estatuto incierto en términos de procedencia, verificabilidad, confiabilidad y valor como evidencia. Si tales materiales se reutilizan para entrenamiento, síntesis automática o recomendación, existe la posibilidad de que errores, sesgos y fórmulas expresivas se recirculen a gran escala. En este punto, el problema de *Moltbook* rebasa la plataforma: se convierte en una advertencia sobre cómo podrían contaminarse las ecologías informacionales más amplias (Amin *et al.*, 2026; Reddit, 2026).

8. Implicaciones teóricas y científicas

Las implicaciones teóricas de *Moltbook* dependen de reconocer un cambio en las condiciones de producción documental. En la plataforma, el documento ya no aparece como resultado de un acto individual de publicación, sino como huella producida por un ensamblaje que combina modelos de lenguaje, *prompts*, operadores humanos, *skills*, API, reglas de ejecución y mecanismos de exposición. Este desplazamiento obliga a estudiar la mediación documental desde la infraestructura que posibilita la emisión, circulación y preservación de registros, no solo desde el contenido textual visible.

Para la bibliotecología y las ciencias de la información, esta situación fortalece la necesidad de una teoría documental capaz de dialogar con la tradición del campo y, al mismo tiempo, extenderla hacia objetos agentivos. La pregunta clásica por el documento como evidencia organizada, materializada y socialmente estabilizada (Briet, 2006; Buckland, 1997; Frohmann, 2004; Lund, 2009; Otlet, 1934) se desplaza hacia una pregunta adicional: bajo qué condiciones una huella generada por agentes puede ser atribuida, descrita, verificada, preservada y utilizada como evidencia. En este punto, *Moltbook* actúa como un banco de pruebas para la archivística computacional, la organización del conocimiento, el análisis de plataformas y la alfabetización crítica frente a sistemas multiagente.

En el ámbito científico, el caso resulta sugerente porque anticipa escenarios en los que los agentes intervengan en tareas como búsqueda bibliográfica, síntesis de literatura, formulación de hipótesis, revisión preliminar o redacción asistida. Weidener *et al.* (2026) utilizan las lecciones del ecosistema *Moltbook/OpenClaw* para reflexionar sobre plataformas de investigación autónoma y sostienen que las fallas observadas en el ámbito social deben traducirse en arquitecturas con validación externa, crítica adversarial, roles delimitados y auditorías trazables. La implicación central es que la automatización científica solo puede ser defendible si conserva mecanismos externos de contraste y responsabilidad.

Por ello, las implicaciones de *Moltbook* son también profesionales. Los especialistas en información deberán diseñar criterios para describir cadenas de procedencia, registrar configuraciones técnicas relevantes, documentar grados de intervención humana, evaluar riesgos de reutilización de huellas automatizadas y enseñar a interpretar críticamente registros producidos por ensamblajes híbridos. Más que sustituir las categorías documentales heredadas, el caso exige actualizarlas para ecologías en las que la producción de información se vuelve parcialmente autónoma, programable y difícil de auditar.

9. Discusión

Los resultados del análisis permiten sostener que *Moltbook* debe entenderse como un caso límite para las ciencias de la información. No prueba, por sí solo, la existencia de una sociedad autónoma de agentes, pero tampoco puede reducirse a una anécdota tecnológica. Su interés radica en que concentra, en un mismo entorno, producción masiva de registros automatizados, circulación acelerada, opacidad de procedencia, espectacularización mediática y dudas sobre validez. Esa combinación desplaza el centro de gravedad desde la pregunta por quién publica hacia la pregunta por cómo se produce, organiza, amplifica y valida una huella documental.

Desde esta perspectiva, el *corpus* revisado muestra una tensión persistente entre apariencia social y densidad documental. Los estudios identifican patrones que recuerdan a redes humanas, como *hubs*, modularidad, concentración de atención y aprendizaje entre pares; sin embargo, incluso registran baja profundidad conversacional, duplicación de mensajes, reciprocidad limitada, vínculos efímeros y escasa estabilización de influencia. La discusión, entonces, no debe confundirse con una oposición simple entre autonomía y simulación, sino con la necesidad de distinguir circulación documental, sociabilidad operativa y comunidad interpretativa (Feng *et al.*, 2026; Holtz, 2026; Hou & Ji, 2026; Li *et al.*, 2026; Zhu *et al.*, 2026).

En esta misma línea, la relación entre consenso interno de los agentes y validez debe colocarse en el centro de la discusión. Que varios agentes reproduzcan, refuercen o coincidan en una interpretación no convierte esa coincidencia en evidencia confiable. En ecologías automatizadas, el consenso puede ser efecto de *prompts* similares, modelos compartidos, datos de entrenamiento convergentes, reglas de visibilidad o retroalimentaciones algorítmicas, y no necesariamente de contrastación crítica. Por ello, la validez epistémica requiere mecanismos externos de verificación, auditoría, revisión humana y trazabilidad de las condiciones de producción (Amin *et al.*, 2026; Wang *et al.*, 2026; Weidener *et al.*, 2026).

Una segunda línea de discusión corresponde a la arquitectura de amplificación. *Moltbook* muestra que la gobernanza informacional no puede limitarse a moderar contenidos una vez publicados, porque la visibilidad se decide en capas previas de ranking, exposición, recomendación y coordinación. Si los episodios coordinados pueden modificar la exposición en lapsos breves, entonces los criterios de organización y acceso dejan de ser neutrales: participan en la construcción de jerarquías documentales y en la distribución de autoridad informativa (Mukherjee *et al.*, 2026; Price *et al.*, 2026; Williams & Ferdinand, 2026).

Metodológicamente, el caso también invita a leer con cautela las fuentes tempranas sobre fenómenos digitales emergentes. Los *preprints*, notas periodísticas y discusiones públicas permiten reconstruir cronologías y controversias iniciales, pero sus resultados deben interpretarse como evidencia provisional. Por ello, este trabajo pretende continuar el debate sobre *Moltbook*, a fin de organizar un marco de lectura que permita diferenciar niveles: infraestructura técnica, huellas documentales, formas de circulación, grados de autonomía, riesgos de gobernanza y condiciones de validez.

Finalmente, *Moltbook* confirma que las categorías clásicas de documento, autoría, procedencia, autenticidad, trazabilidad y preservación siguen siendo necesarias, aunque requieren ser reformuladas. El desafío consiste en continuar la tradición disciplinar, mediante la prolongación crítica hacia escenarios donde los registros son generados por ensamblajes híbridos y circulan en ecologías cada vez menos transparentes. En esa reformulación conceptual reside su mayor valor para los profesionales de la información.

Conclusiones

Moltbook es relevante para los profesionales de la información, ya que confirma el nacimiento de una inteligencia colectiva posthumana y, a la vez, muestra con nitidez un cambio en las condiciones de producción de información. En esta plataforma, las publicaciones, los comentarios, las descripciones de *submolts* y las señales de interacción funcionan como unidades documentales dentro de una ecología en la que la emisión, la clasificación y la circulación están densamente automatizadas, mediadas por ensamblajes operativos como modelos, reglas de sistema y *skills* que configuran la producción documental. Lo decisivo es, por tanto, menos la espectacularidad del fenómeno que su capacidad para poner a prueba, o incluso reinventar, categorías clásicas como documento, autoría, procedencia, autenticidad y trazabilidad.

El *corpus* revisado sugiere que *Moltbook* combina continuidad y ruptura. Continúan ciertos patrones de las redes humanas, como la concentración de la atención, la modularidad y la rápida jerarquización de contenidos. Pero se rompe con otras expectativas, especialmente en lo relativo a la profundidad del diálogo, la estabilidad de la socialización y la claridad de procedencia. Esta mezcla vuelve inadecuadas tanto las lecturas entusiastas que lo presentan como una sociedad ya constituida, así como las desestimaciones que lo reducen a una simple anécdota tecnológica, lo que subraya la necesidad de marcos analíticos propios de las ciencias de la información para interpretar estos fenómenos.

Desde una posición analítica más cauta, conviene entender a *Moltbook* como una arena sociotécnica en la que se intersectan la automatización, la intervención humana, la experimentación pública y la espectacularización mediática, reconfigurando las formas de mediación, acceso y circulación de la información. Su popularidad expresa, al mismo tiempo, fascinación por una nueva forma de interacción y preocupación ante las fragilidades que la sostienen. En ese sentido, el caso funciona como un umbral: deja ver problemas que probablemente se vuelvan más frecuentes en entornos donde los agentes generen, clasifiquen y redistribuyan información con una supervisión humana mínima.

Por ello, el principal valor de *Moltbook* no radica en ofrecer respuestas definitivas, sino en obligar a formular mejores preguntas. ¿Cómo autenticar huellas documentales generadas por ensamblajes híbridos? ¿Cómo registrar la procedencia cuando la cadena de producción atraviesa modelos, *skills*, operadores y plataformas? ¿Cómo conceptualizar la autoría cuando los documentos son producidos por ensamblajes sociotécnicos y no por sujetos individuales? ¿Cómo garantizar la trazabilidad cuando el rastro documental depende de infraestructuras técnicas y mediaciones no visibles? ¿Qué implica la producción de conocimiento científico cuando intervienen sistemas agente-agente en los procesos de investigación y comunicación de resultados, en términos de sus características, condiciones de validez y formas de validación? ¿Cómo evitar que la verificación se vuelva irrelevante debido a la velocidad de circulación? Y, sobre todo, ¿cómo preservar la responsabilidad humana en ecologías informacionales cada vez más automatizadas? Estas preguntas, lejos de concebirse solo como teóricas, inciden directamente en prácticas profesionales como la evaluación, la organización y la mediación de la información en entornos automatizados. En esa capacidad de anticipar y abordar estos dilemas desde la teoría y la práctica de las ciencias de la información, en sus dimensiones informacionales, epistémicas y de gobernanza, reside su verdadera importancia.

Roles de colaboración

Alejandro Villegas-Muro: Conceptualización, Investigación, Redacción del borrador original, Lectura y edición de la versión final del manuscrito.

Juan D. Machin-Mastromatteo: Conceptualización, Revisión y edición del manuscrito, Lectura y edición de la versión final del manuscrito.

Referencias

- Amin, D., Salminen, J. & Jansen, B. J. (2026). How to model AI agents as personas? Applying the persona ecosystem playground to 41,300 posts on *Moltbook* for behavioral insights. *arXiv*. <https://arxiv.org/abs/2603.03140>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Briet, S. (2006). *What is documentation? English translation of the classic French text*. Scarecrow Press. (Trabajo original publicado en 1951).
- Buckland, M. K. (1997). What is a “document”? *Journal of the American Society for Information Science*, 48(9), 804–809. [https://doi.org/10.1002/\(SICI\)1097-4571\(199709\)48:9<804::AID-ASI5>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1097-4571(199709)48:9<804::AID-ASI5>3.0.CO;2-V)
- Chen, E., Guan, C., Elshafiey, A., Zhao, Z., Zekeri, J., Shaibu, A. E. & Prince, E. O. (2026). When OpenClaw AI agents teach each other: peer learning patterns in the *Moltbook* community. *arXiv*. <https://arxiv.org/abs/2602.14477>
- De Marzo, G. & Garcia, D. (2026). Collective behavior of AI agents: the case of *Moltbook*. *arXiv*. <https://arxiv.org/abs/2602.09270>
- El Universo. (2026, febrero 7). *Lo que debe saber de Moltbook, red social para IA*. <https://www.eluniverso.com/larevista/tecnologia/lo-que-debe-saber-de-Moltbook-red-social-para-ia-nota/>
- Feng, Y., Huang, C., Man, Z., Tan, R., Hoang, L. P., Xu, S. & Zhang, W. (2026). MoltNet: understanding social behavior of AI agents in the agent-native *Moltbook*. *arXiv*. <https://arxiv.org/abs/2602.13458>
- Frohmann, B. (2004). *Deflating information: from science studies to documentation*. University of Toronto Press. <https://doi.org/10.3138/9781442673779>
- Habeeb, Z. (2026). *Moltbook* Observatory (2026): playground to super highway in the autonomy era – measuring representation lag in tool-using agentic ecosystems. *Research square*. <https://scity.org/articles/activity/10.21203/rs.3.rs-8971396/v1>
- Holtz, D. (2026). The anatomy of the *Moltbook* social graph. *arXiv*. <https://arxiv.org/abs/2602.10131>
- Hou, W. & Ji, Z. (2026). Structural divergence between AI-agent and human social networks in *Moltbook*. *arXiv*. <https://arxiv.org/abs/2602.15064>
- Jiang, Y., Zhang, Y., Shen, X., Backes, M. & Zhang, Y. (2026). Humans welcome to observe: a first look at the agent social network *Moltbook*. *arXiv*. <https://arxiv.org/abs/2602.10127>
- Li, M., Li, X. & Zhou, T. (2026). Does socialization emerge in AI agent society? A case study of *Moltbook*. *arXiv*. <https://arxiv.org/abs/2602.14299>
- Li, N. (2026). The *Moltbook* illusion: separating human influence from emergent behavior in AI agent societies. *arXiv*. <https://arxiv.org/abs/2602.07432>
- Lin, Y.-Z., Shih, B. P.-J., Chien, H.-Y. A., Satam, S., Pacheco, J. H., Shao, S., Salehi, S. & Satam, P. (2026). Exploring silicon-based societies: an early study of the *Moltbook* agent community. *arXiv*. <https://arxiv.org/abs/2602.02613>

- Lund, N. W. (2009). Document theory. *Annual review of information science and technology*, 43(1), 399–432. <https://doi.org/10.1002/aris.2009.1440430116>
- Manik, M. M. H. & Wang, G. (2026). OpenClaw agents on *Moltbook*: risky instruction sharing and norm enforcement in an agent-only social network. *arXiv*. <https://arxiv.org/abs/2602.02625>
- Morales, O. (2026, febrero 9). Cómo *Moltbook* expuso mitos y riesgos de la inteligencia artificial. *Infobae*. <https://www.infobae.com/tecnologia/2026/02/09/como-Moltbook-expuso-mitos-y-riesgos-de-la-inteligencia-artificial/>
- Mukherjee, K., Akcora, C. G. & Kantarcioglu, M. (2026). MoltGraph: a longitudinal temporal graph dataset of *Moltbook* for coordinated-agent detection. *arXiv*. <https://arxiv.org/abs/2603.00646>
- Murillo, J. M. (2026, febrero 21). ¿Qué es *Moltbook*, la plataforma que funcionaría como una red social para agentes de inteligencia artificial? *Portafolio*. <https://www.portafolio.co/tecnologia/que-es-Moltbook-la-plataforma-que-funcionaria-como-una-red-social-para-agentes-de-inteligencia-artificial-488682>
- Otlet, P. (1934). *Traité de documentation: le livre sur le livre, théorie et pratique*. Éditions Mundaneum. <https://archive.org/details/OtletTraitDocumentationUgent>
- Price, H. C. W., AlMuhanna, H., Bassani, P. M., Ho, M. & Evans, T. S. (2026). Let there be claws: an early social network analysis of AI agents on *Moltbook*. *arXiv*. <https://arxiv.org/abs/2602.20044>
- Prior, L. F. (2008). Document analysis. En L. M. Given (Ed.). *The SAGE encyclopedia of qualitative research methods*. SAGE. https://repository.bbg.ac.id/bitstream/515/1/The_Sage_Encyclopedia_of_Qualitative_Research_Methods.pdf
- Reddit. (2026). *Moltbook* isn't an AI utopia. It's a warning shot about agent ecosystems with no teleology [Publicación en r/Futurology]. Reddit. https://www.reddit.com/r/Futurology/comments/1qywi9v/moltbook_isnt_an_ai_utopia_its_a_warning_shot/?tl=es-419
- ScienceDirect. (2024). Interaction graph. En *ScienceDirect Topics*. <https://www.sciencedirect.com/topics/computer-science/interaction-graph>
- Teixeira, M. M. (2025). Moltbook - Connecting intelligences at the neural frontier nexus. *Journal of technologies information and communication*, 5(2), 43038. <https://www.rtic-journal.com/download/Moltbook-connecting-intelligences-at-the-neural-frontier-nexus-17891.pdf>
- Timmermans, S. & Tavory, I. (2022). *Data analysis in qualitative research: theorizing with abductive analysis*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/D/bo133273407.html>
- Wang, C., Li, C., Liu, S., Chen, Z., Hou, J., Qi, J., Li, R., Zhang, L., Ye, Q., Liu, Z., Chen, X., Zhang, X. & Yu, P. S. (2026). The devil behind *Moltbook*: anthropic safety is always vanishing in self-evolving AI societies. *arXiv*. <https://arxiv.org/abs/2602.09877>
- Weidener, L., Brkić, M., Lee, P., Karlsson, M., Noessler, K., & Kohlhaas, P. (2026). From agent-only social networks to autonomous scientific research: lessons from OpenClaw and *Moltbook*, and the architecture of ClawdLab and Beach.Science. *arXiv*. <https://arxiv.org/abs/2602.19810>
- Williams, N. L. & Ferdinand, N. (2026). Form or function? Early dynamics of the *Moltbook* AI social media network. *ROBONOMICS: the Journal of the automated economy*, 7, 90. <https://www.journal.robonomics.science/index.php/rj/article/view/90>
- Zhang, Y., Mei, K., Liu, M., Wang, J., Metaxas, D. N., Wang, X., Hamm, J. & Ge, Y. (2026). Agents in the wild: safety, society, and the illusion of sociality on *Moltbook*. *arXiv*. <https://arxiv.org/abs/2602.13284>
- Zhu, Y., Tyson, G. & Hui, P. (2026). A comparative analysis of social network topology in Reddit and *Moltbook*. *arXiv*. <https://arxiv.org/abs/2602.13920>