AI Agents for Economic Research

Korinek, A., 2025

Francisco Iñaqui Irustia, Federico Ariel Lopez, Martina Palazzini

7 de octubre de 2025

Indice

Progreso de las habilidades de las IA

- **LLM Tradicionales** (2022-2023)
- **Modelos de Razonamiento** (2024)
- **Agentic Chatbots** (2025)

El Panorama de Servicios de IA

Servicios Premium

Table 5
PRICING TIERS FOR CHATBOT SUBSCRIPTIONS OF LEADING AI LABS (USD PER MONTH, JULY 2025)

Lab / Provider	Basic tier	$/mo	Most expensive tier	$/mo
OpenAI ChatGPT	Plus	$20	Pro	$200
Google DeepMind Gemini	AI Pro	$20	AI Ultra	$250
Anthropic Claude	Pro	$20	Max 20× usage	$200
xAI Grok	X Premium	$8	SuperGrok Heavy	$300
Microsoft Copilot	Copilot Pro	$20		

Otros Modelos y Modelos Abiertos

- Mistral
- Kimi-K2
- Qwen
- DeepSeek
- Minimax Locales:
- Llama
- GPT-OSS (120B y 20B)
- Gemma
- Phi3
- Otra opción: T3chat ('agregador')

... fijarse en Ollama y HuggingFace

LLMs Tradicionales

'System-1 thinking' (Kahneman)

- Sin acceso a información en tiempo real
- No puede seguir pasos o derivaciones matematicas
- Sigue siendo muy cercano al next-word predictor (GPT-1)

Comparación de LLMs: `https://lmarena.ai/leaderboard`

Recomendado: 'What Is ChatGPT Doing ... and Why Does It Work? (Stephen Wolfram)'

Modelos de Razonamiento

'System-2 thinking'

- Resolución deliberada y paso a paso de problemas
- Identificación y corrección de errores
- Siguen sin ser más que next word predictors

Table 3
Top Reasoning and Agentic Capabilities by Lab, GPQA & SWE-Bench-V Scores

Lab	Model	Last Updated	GPQA ◇ Score	SWE-Bench V
OpenAI	GPT-5	2025-08-07	89.4%	75%
xAI	Grok-4	2025-07-10	88.9%	72-75%*
Google DeepMind	Gemini 2.5 Pro	2025-06-17	84.0%	63.8%
Anthropic	Claude Opus 4.1	2025-08-05	83.3%	72.7%
DeepSeek	DeepSeek-R1	2025-03-24	71.5%	49.2%

Source: Compiled by author. Last update: Aug 20th, 2025; * marks preliminary data.

Agentes con IA

Realizan acciones + usan herramientas externas
Los modelos previos que generan respuestas 'con lo que ya tienen', los agentes tienen funcionalidad extra:

- Búsquedas web
- Interacción con bases de datos
- Ejecución de código
- Manipulación de archivos

Es 'agentic' porque deja a la IA 'tomar la decisión' de usar una herramienta a la que tiene acceso.

EXTRA: ¿Qué quieren decir con 'Agente'?

En computación los agentes perciben su ambiente mediante 'sensores' y actúan sobre el ambiente por 'actuators.'

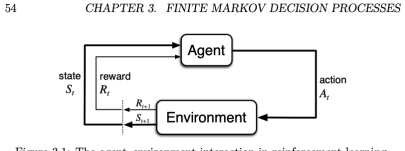

54 CHAPTER 3. FINITE MARKOV DECISION PROCESSES

Figure 3.1: The agent–environment interaction in reinforcement learning.

Sutton & Barto

Los agentes de IA combinan LLMs con herramientas externas (tools) y un orquestador que coordina el flujo: Think $\rightarrow$ Act $\rightarrow$ Observe.

Componentes:

- **Reasoning Engine**: LLM que decide qué hacer
- **Tools**: búsqueda web, ejecución de código, APIs
- **Memory**: contexto acumulado
- **Orchestrator**: coordina el ciclo

Alignment: Economía se centra en el problema de principal-agente; en computación el problema se le dice alignment (Hendrycks, 2025)

Arquitectura de un Agente de IA

Korinek: AI Agents for Economic Research

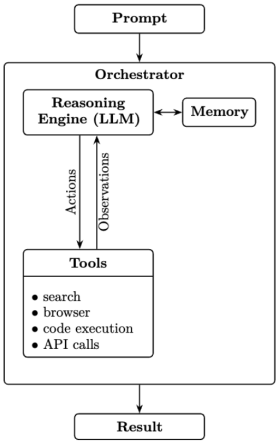


Figure 1. AI Agent Architecture

- Un *orchestrator* pasa el objetivo original (ej: prompt del usuario) y la lista de herramientas disponibles a un LLM de razonamiento.
- Este LLM decide qué herramientas externas llamar. Las herramientas proveen al sistema una interfaz con el mundo externo:
- Cada vez que el motor de razonamiento quiere llamar una herramienta, genera tokens que indican al orchestrator llamar la herramienta e insertar el resultado de vuelta antes de continuar.

Creado por Google DeepMind 2024

Table 6
OVERVIEW OF DEEP RESEARCH AGENTS

System	Availability & usage limits	Features	Time
Gemini Deep Research	Free to try; better & higher allowance on paid plans	Proposes a research plan that user can confirm/ modify	5-10 min
OpenAI Deep Research	Allowance of 25/200 reports per month for Plus/Pro plan	Asks follow-up questions to optimally target response	5-30 min
Claude Deep Research	Available subject to limits for paid plans	Runs immediately; can connect with workspace	5-15 min
Perplexity Deep Research	Free to try; unlimited under paid plan	Fast option covering lots of sources	2-4 min
xAI Grok DeepSearch	Available under X premium plan	Real-time info through close integration with X platform	<1-10 min

- Un agente líder es el orchestrator que recibe el query
- Arma subtasks y cada una la realiza alguno de los subagentes especializados
- Síntesis final de resultados de múltiples fuentes

Coding Agents

Terminal-based coding y 'Vibe Coding'

'Vibe coding' – crear proyectos de software completos basados en prompts en lenguaje natural. Ha hecho posible que usuarios sin experiencia en programación creen proyectos de software de principio a fin.

Full vibe coding:

- Claude Code
- Gemini CLI
- Codex + Open Codex

Herramientas intermedias:

- GitHub Copilot
- Cursor
- Windsurf
- Cline

Data Retrieval Agent (Primera Generación)

Lo que hace este código un 'agente' en lugar de un simple script es su comportamiento autónomo y dirigido a objetivos a través de múltiples pasos:

El Loop Think-Act-Observe:

1. **Think:** Piensa qué datos ayudarían a responder la pregunta
2. **Act:** Llama la herramienta externa apropiada (FRED API)
3. **Observe:** Analiza los datos retornados
4. **Respond:** Responde con una respuesta en lenguaje natural

Este loop Think-Act-Observe caracteriza sistemas agentic sofisticados.

LangGraph Agents

¿Y pasa si necesitamos capacidades más complejas?

- ¿Y si el agente necesita obtener múltiples series de datos?
- ¿Y si queremos que el agente pueda retroceder cuando se da cuenta de que está siguiendo un camino improductivo?

LangGraph provee un framework para construir agentes con:

- Grafos de estado complejos
- Capacidad de backtracking
- Memoria persistente
- Coordinación multi-agente

The Model Context Protocol (MCP) y Agent2Agent Protocol

Anthropic, 2024
Para economistas, los servidores MCP de uso general proveen a los agentes de IA acceso automatizado a:

- Sistema de archivos del usuario
- Sistemas de email (Gmail, Outlook)
- Bases de datos financieras
- Bases de datos espaciales
- GitHub
- Slack
- Wrappers de FRED
- FMI
- más... en pulse

Agent2Agent: (Abril 2025) Protocolo para la interacción entre agentes

EXTRA: AI Co-Scientist

Si todo lo puede hacer la IA ¿para qué servimos los economistas? ¿la investigación va a quedar obsoleta?

Por ahora, no.

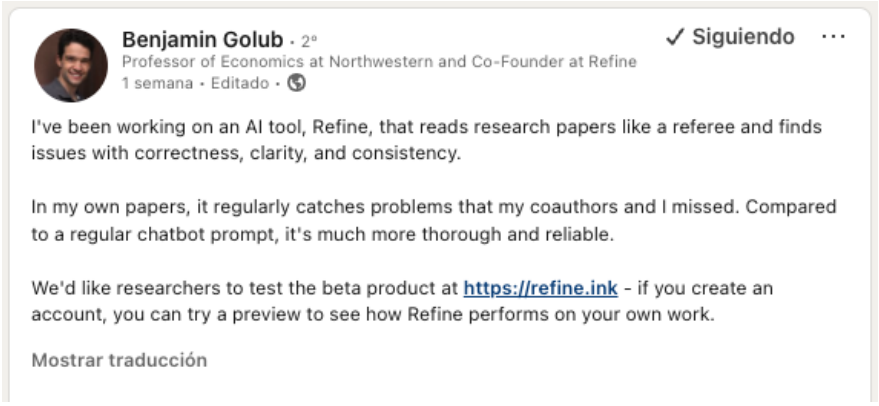
Mientras los agentes de IA sobresalen en síntesis e implementación,

- La identificación de preguntas de investigación novedosas
- La intuición conectando fenómenos dispares

siguen siendo contribuciones humanas.

Pero... hay esfuerzos en marcha que buscan automatizar estas contribuciones creativas usando sistemas de IA agentic — ver, por ejemplo, Google's AI Co-Scientist.

EXTRA: Refine.ink



Benjamin Golub · 2º
Professor of Economics at Northwestern and Co-Founder at Refine
1 semana · Editado ·

✓ Siguiendo ···

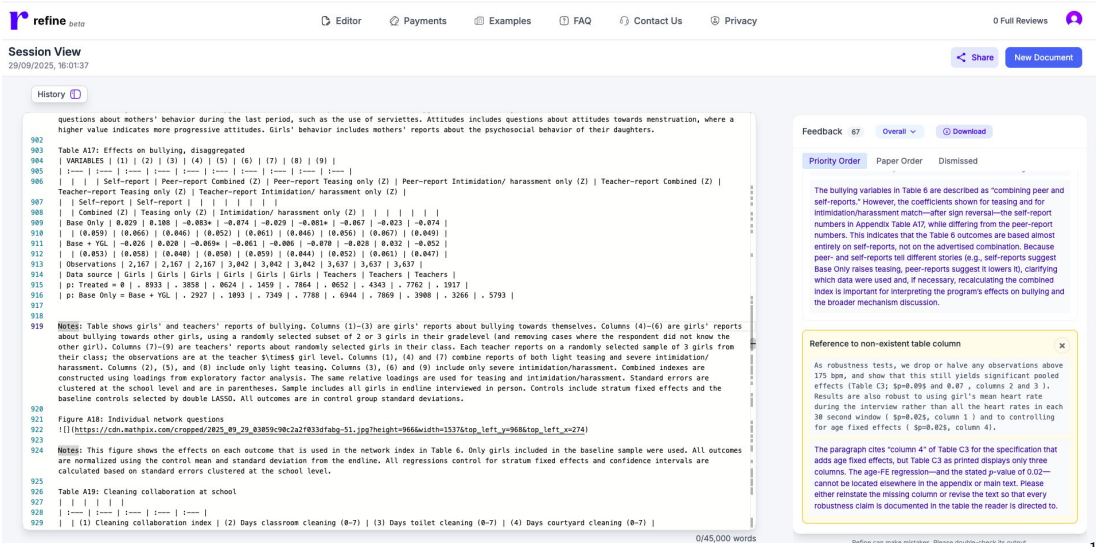
I've been working on an AI tool, Refine, that reads research papers like a referee and finds issues with correctness, clarity, and consistency.

In my own papers, it regularly catches problems that my coauthors and I missed. Compared to a regular chatbot prompt, it's much more thorough and reliable.

We'd like researchers to test the beta product at **https://refine.ink** - if you create an account, you can try a preview to see how Refine performs on your own work.

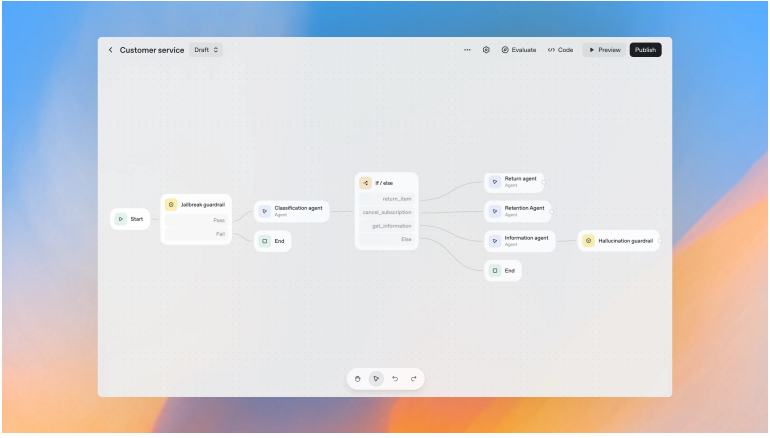
Mostrar traducción

EXTRA: Refine.ink



AgentKit (Ayer)

AgentKit



Conclusion

limitaciones:

- Alucinaciones y errores que se propagan sin detección
- Fragilidad ante variaciones en prompts
- Dificultades con razonamiento económico genuino

"Tratar a los agentes de IA como un profesor trataría a un equipo de asistentes de investigación: requieren planificación cuidadosa, supervisión durante la ejecución, y verificación detallada de resultados."