



Let AI Agents Do the Work

Extract and Enrich Data from Text and
Images with Python

September 5, 2025
Marcel Pauly

bit.ly/scicar-agents

What are Agents?

- LLMs that can reason, plan ...
- ... and **use tools** (web search, local resources, code execution, APIs)
- Return **structured outputs** rather than free text

Why Pydantic AI?

- Works across providers (OpenAI, Anthropic, Google, ...)
- Typed outputs with Pydantic → safer, verifiable results
- Common tools and MCP integration
- Alternatives: LangChain, CrewAI, ..., LLM providers' libraries

Which Model Should I Use?

- Check models' capabilities (tool-use and structured-output support)
- Start with the cheapest one that meets your requirements
- Scale up if a larger model delivers better results

As of Sept 3, 2025

Provider	Model	Input	Output	Builtin web search	Custom tools
OpenAI	gpt-5-nano-2025-08-07	\$0.05/MTok	\$0.40/MTok	?	✓
OpenAI	gpt-5-mini-2025-08-07	\$0.25/MTok	\$2.00/MTok	✓	✓
Anthropic	claude-3-5-haiku-latest	\$0.80/MTok	\$4.00/MTok	✓	✓
Google	gemini-2.5-flash-lite	\$0.10/MTok	\$0.40/MTok	✓	✗

Tools

- **Built-in tools:** Web Search, Code Execution
- **Common tools:** DuckDuckGo Search, Tavily Search
- **Your own tools:** wrap any Python function; call external APIs, ...
- **MCP:** open protocol for exposing tools/resources to LLMs

MCP

- Find servers: [MCP Server Directory](#), [MCP Archive](#)
- Examples: [Wikipedia MCP Server](#), [Genesis MCP Server](#) (unofficial!)
- Only use MCP servers from people and organisations you trust!
- Build your own: [using Python or TypeScript \(Node.js\)](#)
- Use in Pydantic AI: connect MCP servers as [toolsets](#)

Batch Mode

- 50% cost savings
- Worth considering for very large datasets
- **Non-urgent tasks only:** jobs can take up to 24 hours
- **Not yet supported** in Pydantic AI → use provider batch APIs
- Alternative: pseudo-batch in a single prompt

The Golden Hammer?

- LLMs and agents are **not always the best tool**
- **Simple methods** (e.g., regex) often suffice
- Sometimes **specialized ML models** outperform general LLMs ([Hugging Face](#))
- Often you can apply tools yourself on the LLM's output
- Small facts can simply go into the prompt
- Keep **trade-offs** in mind: cost, latency, limited reproducibility, hallucinations, environmental impact

Thank you!