



Agent Builder: **Providing Relevant** **Context for Data Driven** **Agents**





**Kathleen
DeRusso**
Principal Software
Engineer,



Agenda

01

What are Agents?

02

Why do we need Context Engineering?

03

Elastic Agent Builder

04

Demo

05

Future Direction

06

Q&A

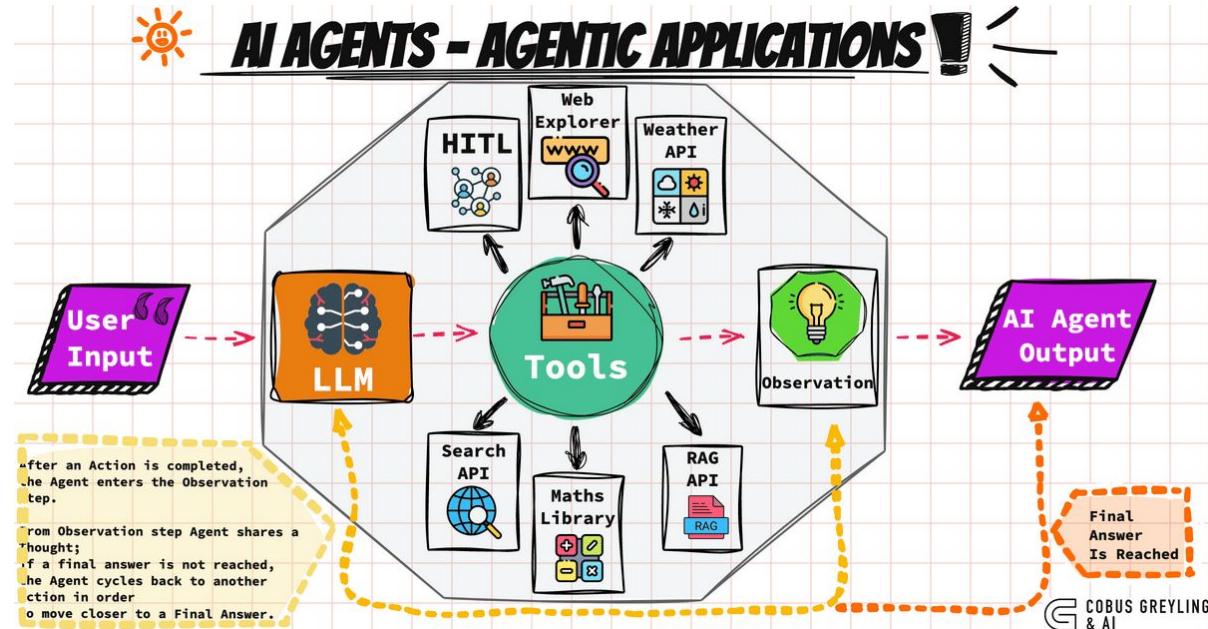
What are Agents?

Not a secret agent!



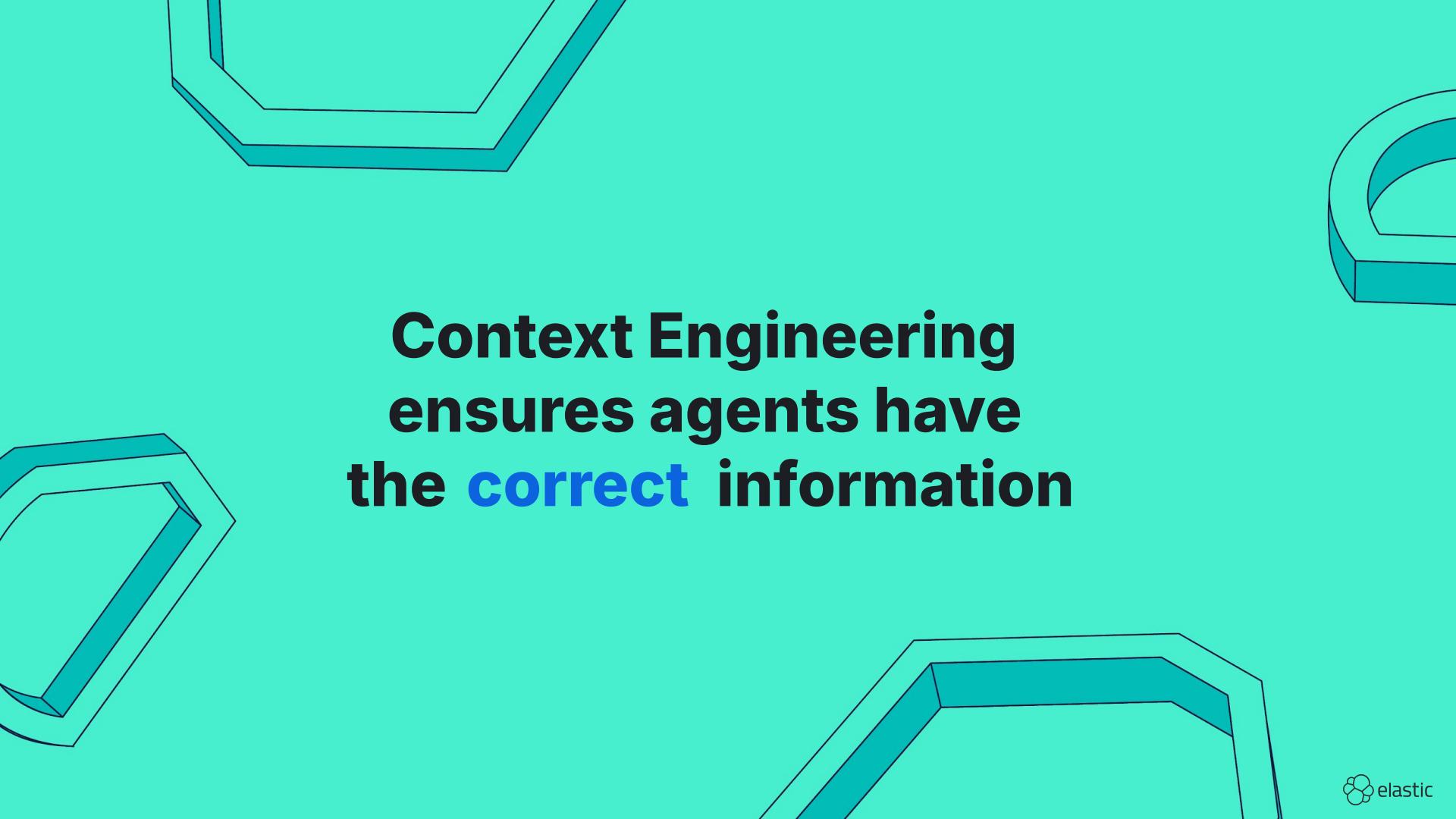
What is an Agent?

An AI agent is a software program that can autonomously perform complex tasks by making decisions, learning from its environment, and using tools to achieve goals



Types of AI Agents





**Context Engineering
ensures agents have
the **correct** information**

Context Window

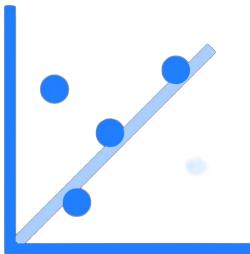
The maximum number of tokens an LLM can process at once.



Why do LLMs Hallucinate?



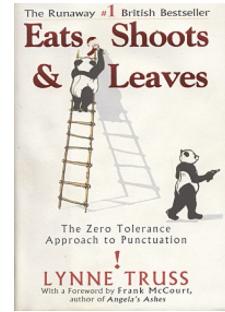
Frozen or Limited Knowledge



Model Overfitting



Dataset Biases



Language Ambiguity

Why Language Models Hallucinate

Adam Tauman Kalai*
OpenAI

Ofir Nachum
OpenAI

Santosh S. Vempala†
Georgia Tech

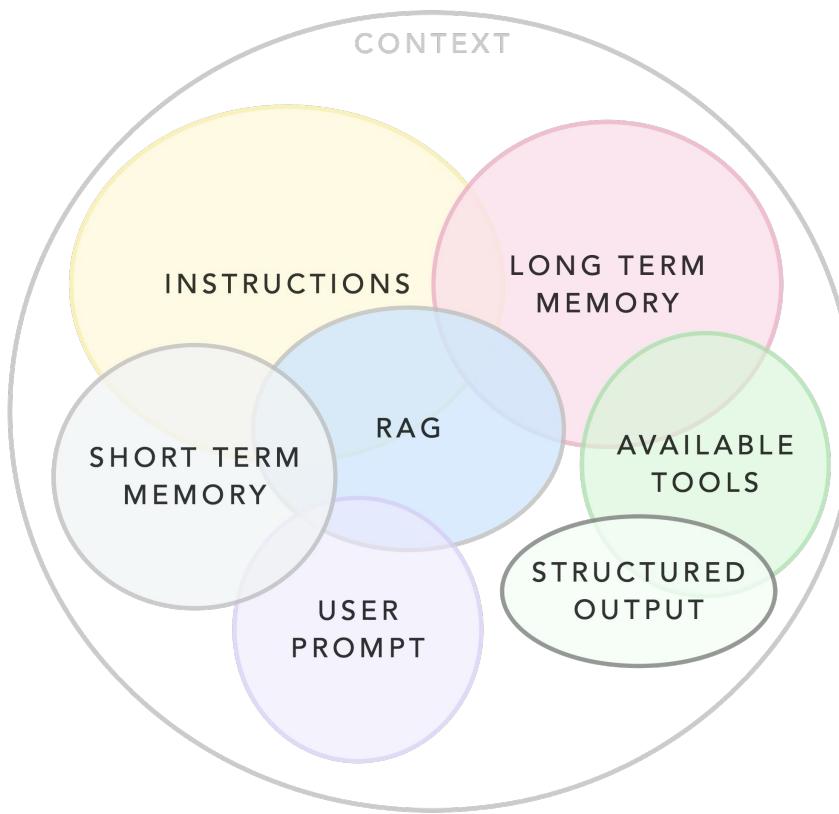
Edwin Zhang
OpenAI

September 4, 2025

Abstract

Like students facing hard exam questions, large language models sometimes guess when uncertain, producing plausible yet incorrect statements instead of admitting uncertainty. Such “hallucinations” persist even in state-of-the-art systems and undermine trust. We argue that language models hallucinate because the training and evaluation procedures reward guessing over acknowledging uncertainty and we analyze the statistical causes of hallucinations in the modern training pipeline. Hallucinations need not be mysterious—they originate simply as errors in binary classification. If incorrect statements cannot be distinguished from facts, then hallucinations in pretrained language models will arise through natural statistical pressures. We then argue that hallucinations persist due to the way most evaluations are graded—language models are optimized to be good test-takers, and guessing when uncertain improves test performance. This “epidemic” of penalizing uncertain responses can only be addressed through a socio-technical mitigation: modifying the scoring of existing benchmarks that are misaligned but dominate leaderboards, rather than introducing additional hallucination evaluations. This change may steer the field toward more trustworthy AI systems.

What is Context Engineering?

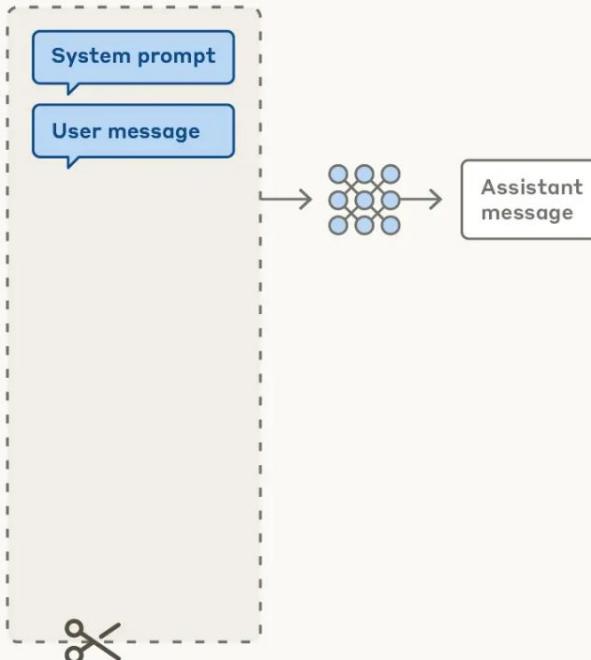


Credit: What is Context Engineering | Elasticsearch Labs

Prompt engineering vs. context engineering

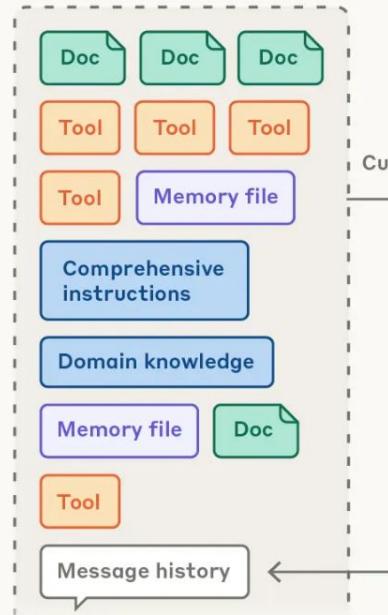
Prompt engineering for single turn queries

Context window



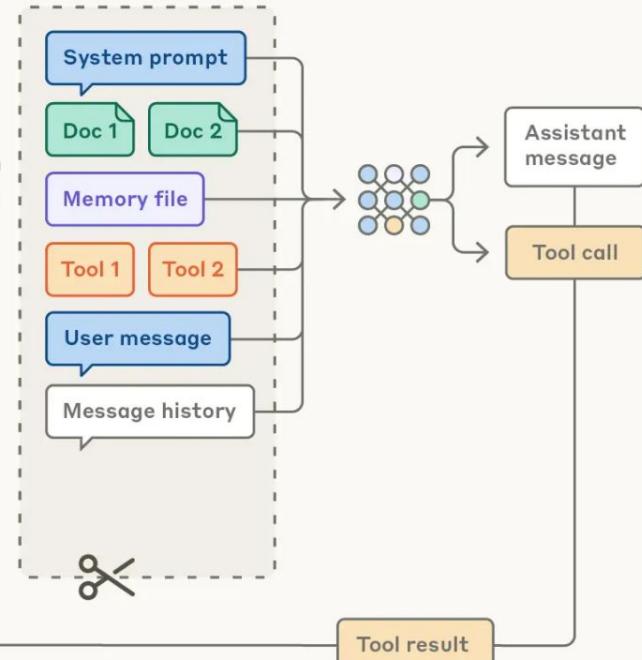
Context engineering for agents

Possible context to give model



Curation

Context window

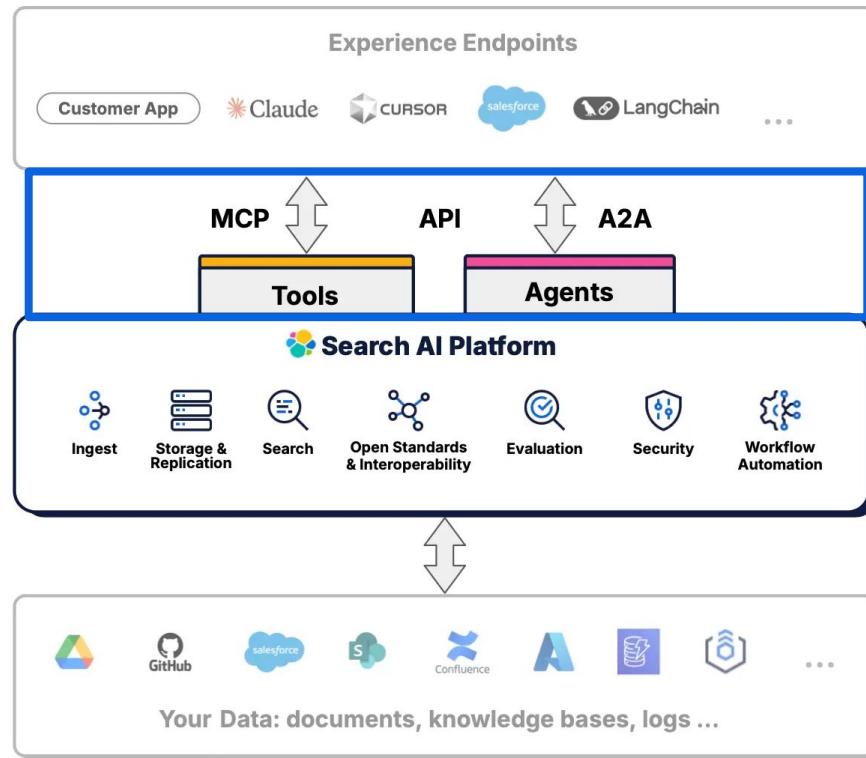


[Credit: Effective context engineering for AI agents | Anthropic](#)



**Agent Builder
allows building of
data-centric contextual
agents**

Where does Agent Builder fit?



Credit: Building AI Agentic workflows with Elasticsearch | Elasticsearch Labs

Capabilities

Elastic Agent Builder has the following features



Chat with your data using the native conversational agent and see full tracing of the steps and results



Leverage built-in tools to find relevant indices, understand data structure and translate native language into structured semantic, lexical or hybrid queries



Build custom tools using ES|QL for use in your own agents

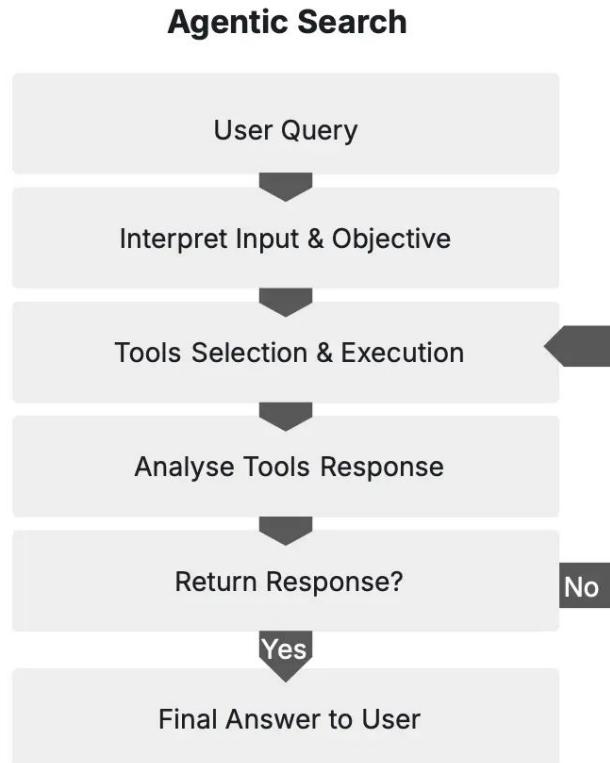


Built in visualizations for chat with editable chart types and the ability to add to Dashboards



Integrate with external agents and apps using MCP and A2A

How do Agents work in Elasticsearch?



In-built tools

- There are currently 11 tools available, prefixed by `platform.core`:
- Execute ES|QL queries
 - Generate ES|QL query from natural language
 - Get full document content based in ID
 - Retrieve mappings for one or more specified indices
 - List relevant indices, aliases and streams based on natural language query
 - List indices, aliases and datastreams in the cluster
 - Core search tool
 - ... and more!

The screenshot shows the Elasticsearch Agent Builder Tools page. On the left, there's a sidebar with navigation links: Elasticsearch, Discover, Dashboards, Agents, Machine Learning, Data management, and a gear icon. The main content area has a title "Tools" and a sub-section "Tools are modular, reusable Elasticsearch operations. Agents use them to search, retrieve, and analyze your data. Use our built-in tools for common operations, and create your own for custom use cases. [Learn more](#)". Below this is a search bar and a "Labels" dropdown. A table lists 11 tools, each with a checkbox, a name, a description, and a "Labels" column. The tools are:

	Tool Name	Description	Labels
<input type="checkbox"/>	<code>ID ↑</code>		
<input type="checkbox"/>	<code>platform.core.cases</code>	Retrieves cases from Elastic Security, Observability, or Stack Management. Supports three operation modes:	<code>cases</code>
<input type="checkbox"/>	<code>platform.core.execute_esql</code>	Execute an ES QL query and return the results in a tabular format.	
<input type="checkbox"/>	<code>platform.core.generate_esql</code>	Generate an ES QL query from a natural language query.	
<input type="checkbox"/>	<code>platform.core.get_document_by_id</code>	Retrieve the full content (source) of an Elasticsearch document based on its ID and index name.	
<input type="checkbox"/>	<code>platform.core.get_index_mapping</code>	Retrieve mappings for the specified index or indices.	
<input type="checkbox"/>	<code>platform.core.get_workflow_execution_status</code>	Retrieve the status of a workflow execution.	
<input type="checkbox"/>	<code>platform.core.index_explorer</code>	List relevant indices, aliases and datastreams based on a natural language query.	
<input type="checkbox"/>	<code>platform.core.integration_knowledge</code>	Search and retrieve knowledge from Fleet-installed integrations. This includes information on how to configure and use integrations for data ingestion into Elasticsearch.	<code>Integration</code> <code>knowledge-base</code> <code>fleet</code>
<input type="checkbox"/>	<code>platform.core.list_indices</code>	List the indices, aliases and datastreams from the Elasticsearch cluster.	
<input type="checkbox"/>	<code>platform.core.product_documentation</code>	Search and retrieve documentation about Elastic products (Kibana, Elasticsearch, Elastic Security, Elastic Observability).	
<input type="checkbox"/>	<code>platform.core.search</code>	A powerful tool for searching and analyzing data within your Elasticsearch cluster.	

At the bottom, there's a "Rows per page: 25" dropdown and a footer with the elastic logo.

Which LLM?

Default
LLMs provided via Elastic
Inference Service (EIS)

The screenshot shows the 'Connectors' page with a search bar and tabs for 'Connectors' and 'Logs'. A 'Create connector' button is at the top right. Below is a table listing connectors:

Name	Type	Compatibility	Action
Anthropic Claude Opus 4.5	AI Connector	Generative AI for Security Generative AI for Search Generative AI for Observability Workflows	PRECONFIGURED
Anthropic Claude Opus 4.6	AI Connector	Generative AI for Security Generative AI for Search Generative AI for Observability Workflows	PRECONFIGURED
Anthropic Claude Sonnet 3.7	AI Connector	Generative AI for Security Generative AI for Search Generative AI for Observability Workflows	PRECONFIGURED
Anthropic Claude Sonnet 4.5	AI Connector	Generative AI for Security Generative AI for Search Generative AI for Observability Workflows	PRECONFIGURED
Elastic-Cloud-SMTP	Email	Alerting Rules Security Solution Workflows	PRECONFIGURED
Google Gemini 2.5 Flash	AI Connector	Generative AI for Security Generative AI for Search Generative AI for Observability Workflows	PRECONFIGURED
Google Gemini 2.5 Pro	AI Connector	Generative AI for Security Generative AI for Search Generative AI for Observability Workflows	PRECONFIGURED
OpenAI GPT-4.1	AI Connector	Generative AI for Security Generative AI for Search Generative AI for Observability Workflows	PRECONFIGURED

Connector
Create a custom connector
using the AI Connector (tech
preview)

The screenshot shows the 'AI Connector' configuration dialog with a 'TECHNICAL PREVIEW' notice. It includes fields for 'Connector name' (set to 'my-connector'), 'Service' (set to 'OpenAI'), and 'Model ID' (set to 'gpt-4'). The 'Settings' section has a note about re-entering the API key. The 'Authentication' section shows an 'API Key' field with placeholder text and a note about generating OpenAI API keys.

AI Connector TECHNICAL PREVIEW

Send requests to AI providers such as Amazon Bedrock, OpenAI and more.

Compatibility: Generative AI for Security | Generative AI for Search | Generative AI for Observability

Connector name: my-connector

Service: OpenAI

Model ID: gpt-4

The name of the model to use for the inference task.

More options >

API Key: [REDACTED]

The OpenAI API authentication key. For more details about generating OpenAI API keys, refer to the <https://platform.openai.com/account/api-keys>.

You will need to re-enter your API Key each time you edit the inference endpoint

Back

Demo

Workflows

- Technical preview
- Declarative, composable workflows defined in yaml
- Composed of triggers, inputs and steps
- `ai.agent` steps invoke agents from within a workflow
- Workflows can be exposed to Agent Builder as tools
- Technical blog:
<https://www.elastic.co/search-labs/blog/elastic-workflows-automation>



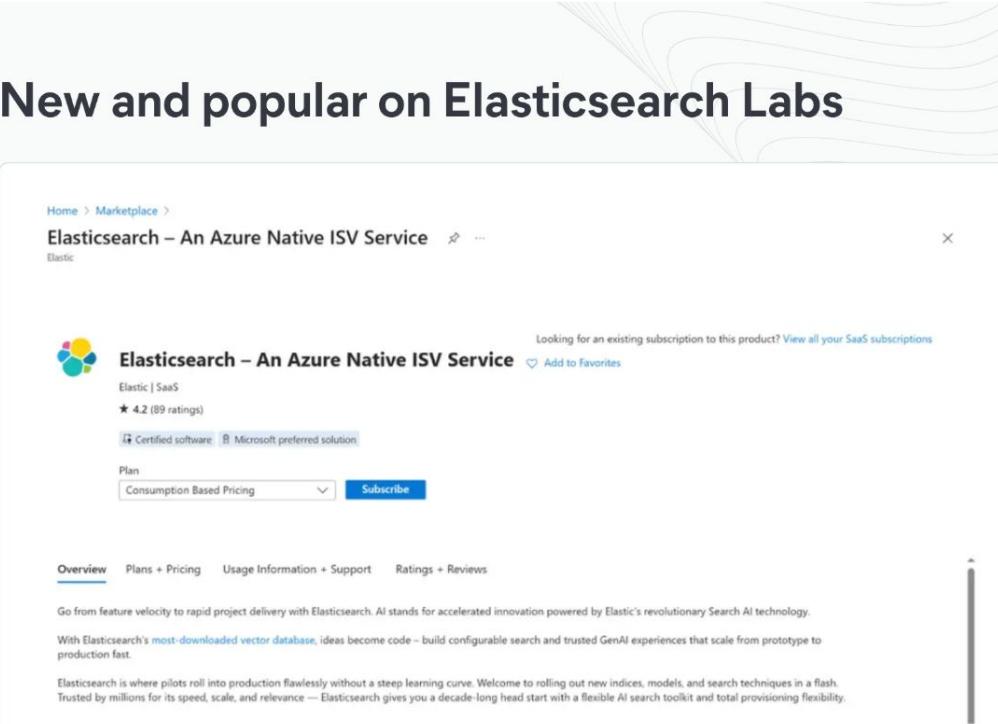
The screenshot shows a workflow named "Invoke an Agent" in the "Workflow" section of the Elastic Stack UI. The workflow has just been saved. The configuration is as follows:

```
1 version: "1"
2 name: 🛡 Invoke an Agent
3 enabled: true
4 triggers:
5   - type: manual
6 steps:
7   - name: Call the custom alert AI agent
8     type: kibana.request
9     with:
10       method: "POST"
11       path: "/api/agent_builder/converse"
12       headers:
13         kbn-xsrf: "true"
14       body:
15         agent_id: alert.agent
16         input:
17           | Summarize the alerts in last 24hrs
18       timeout: 10m
19       on-failure:
20         retry:
21           max-attempts: 3
```

Resources

- [Introducing Elastic's Agent Builder](#)
- [Building AI Agentic workflows with Elasticsearch](#)
- [Your first Elastic Agent: From a single query to an AI-powered chat](#)
- [Connecting Elastic Agents to Gemini Enterprise via A2A protocol](#)
- ... and many more!
- Search labs tag [Agentic AI](#)

New and popular on Elasticsearch Labs



The screenshot shows a product listing for "Elasticsearch – An Azure Native ISV Service" by Elastic. The page includes the product title, a green circular icon, the developer name "Elastic | SaaS", a rating of "★ 4.2 (89 ratings)", and badges for "Certified software" and "Microsoft preferred solution". Below the product card, there are tabs for "Overview", "Plans + Pricing", "Usage Information + Support", and "Ratings + Reviews". The "Overview" tab is selected. The content under "Overview" discusses the rapid project delivery and accelerated innovation provided by Elasticsearch's AI technology. It also mentions the product's speed, scale, and relevance, noting it is trusted by millions. At the bottom, there is a "Basics" button and the date "October 29, 2025".

Home > Marketplace >
Elasticsearch – An Azure Native ISV Service ⚙️ ...
Elastic

Elasticsearch – An Azure Native ISV Service [Add to Favorites](#)

Elastic | SaaS
★ 4.2 (89 ratings)
Certified software Microsoft preferred solution

Plan [Consumption Based Pricing](#) [Subscribe](#)

Overview Plans + Pricing Usage Information + Support Ratings + Reviews

Go from feature velocity to rapid project delivery with Elasticsearch. AI stands for accelerated innovation powered by Elastic's revolutionary Search AI technology.

With Elasticsearch's most-downloaded vector database, ideas become code — build configurable search and trusted GenAI experiences that scale from prototype to production fast.

Elasticsearch is where pilots roll into production flawlessly without a steep learning curve. Welcome to rolling out new indices, models, and search techniques in a flash. Trusted by millions for its speed, scale, and relevance — Elasticsearch gives you a decade-long head start with a flexible AI search toolkit and total provisioning flexibility.

Basics October 29, 2025

How to deploy Elasticsearch on an Azure Virtual Machine

Questions?

Thank you!

