All of us need you!
Speak at a future meetup

elastic | The Search AI Company

How to get involved

- Interested in presenting? We would love to hear from you. We welcome lightning talks, deep dives and everything in between.
- Help us grow the user group! Bring a friend or colleague to our next meetup.
- Is your company interested in hosting? We love working with local community members and their companies on joint events. Hosting a meetup is also a great recruiting opportunity. And as always, Elastic will cover the pizza, beer and swag!

- Do you know of an affinity meetup chapter (a group for developers interested in various technology) that we could partner with for a joint event? Let us know. We love collaborating with local tech groups.
- Something else? Feel free to reach out to us to chat more: meetups@elastic.co

Speak at a future meetup

https://ela.st/chicago-meetups

elastic | The Search AI Company

February 2026 Chicago Meetup

ACT with Your Data!

(continuation from Nov 2025..."Talk to your Data")

Terry Dupureur (pronounced 'doo-pure') | Solution Architect
terry.dupureur@elastic.co

elastic | The Search AI Company

Elasticsearch is famous for `/_search`

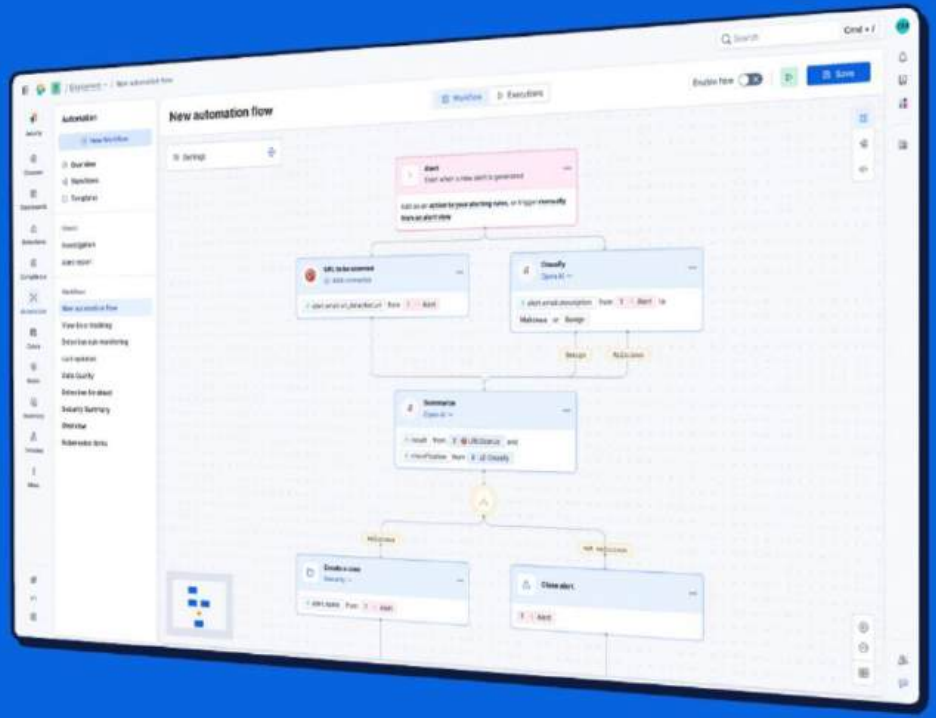
~~What if~~ WE (indeed) made it easy to `/_chat`

What if we made it easy to `/_workflow`

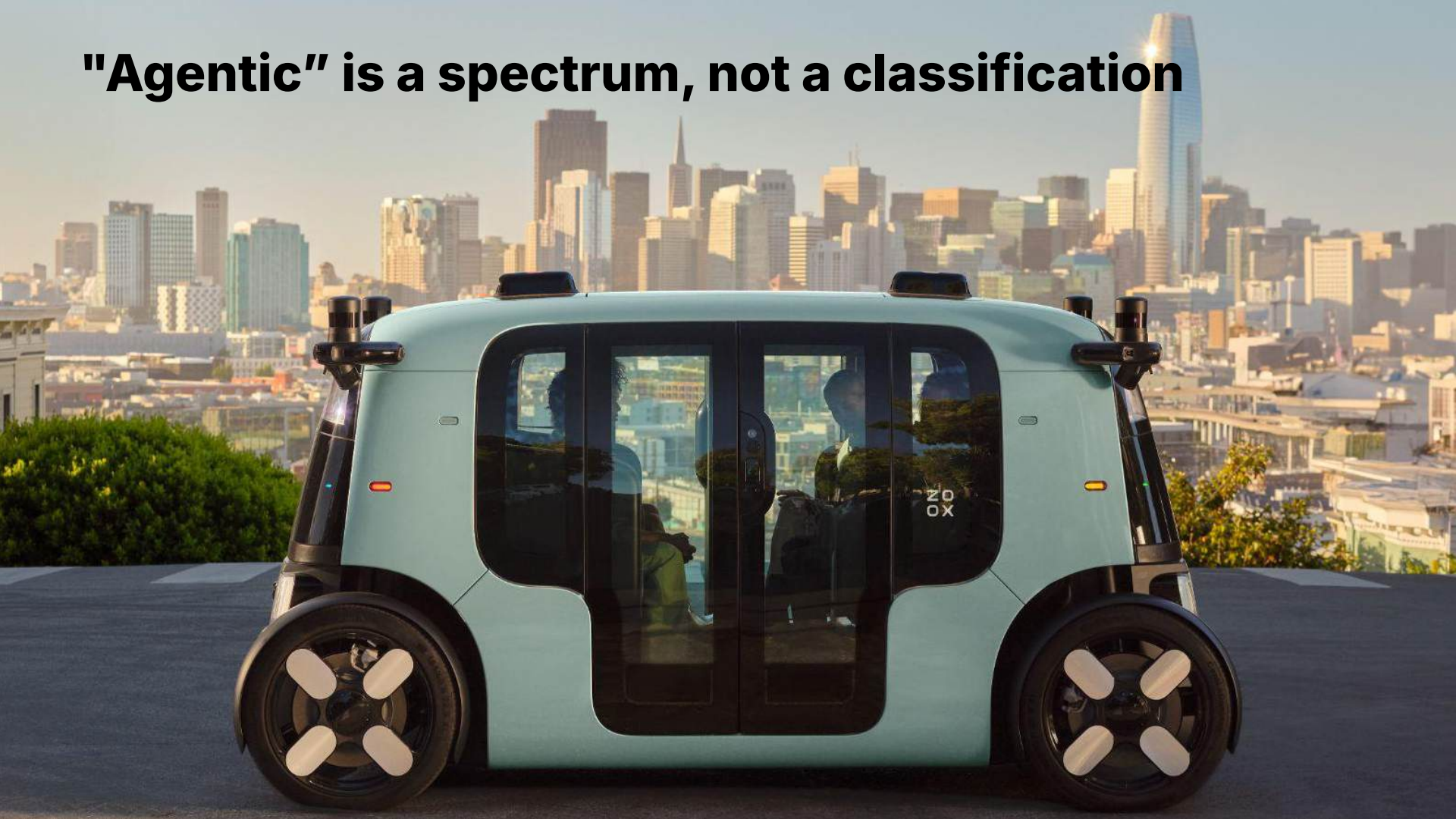
NEW

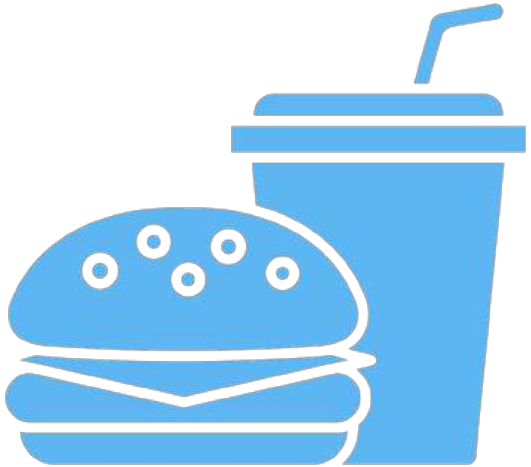
Introducing **Elastic Workflows**

- Automate actions directly where your data lives
- Both rules-based and agentic automation



"Agentic" is a spectrum, not a classification





Workflows

are about **predictable results** and **speed**, at **scale**.

Agents

are about **decisions**, **actions**, and **outcomes**.

elastic

Workflows and Agents

Workflows are systems where LLMs and tools are orchestrated through predefined code paths (including **RAG**).

Agents are systems where LLMs:

- Dynamically direct their own processes
- Dynamically manage their own tool usage
- Maintain control over how they accomplish tasks.

Demo

Today’s presentation:

01	02	03	04	05
Understanding the landscape	What Agents are and aren't	Where we fit	Where we shine	Call to action

Today’s presentation:

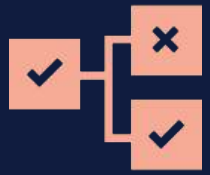
01

Understanding
the landscape

An AI timeline

Symbolic

1950s



Expert Systems

1970s

Neural Networks

1980s

Machine Learning

~~1990s~~ 2010s

Generative

2018

RAG & Agentic

2020s

NLP is the undergirding of all of these technologies!

Core AI technologies

NLP

Understand and process language.

ML

Learn patterns from data.

GenAI / RAG

Generate content from a prompt, with or without context.

Agentic AI

Plan, reason, and act towards a goal.

All of these are native to the Elastic Stack.

Today’s presentation:

01 Understanding the landscape

02 What Agents are and aren't

What is Agentic AI?

Agentic AI describes systems that are designed to:

1. **Autonomously reason** and make decisions.
2. **Discover** tools, resources, and data in their environment
3. **Independently plan** and **take actions** based on goals
4. **Revise plans** and adapt based on new information
5. **Pursue** complex goals with limited supervision

What are the components of Agentic AI?

Agents

What goals should be accomplished

Tools

What actions are available to accomplish the goals

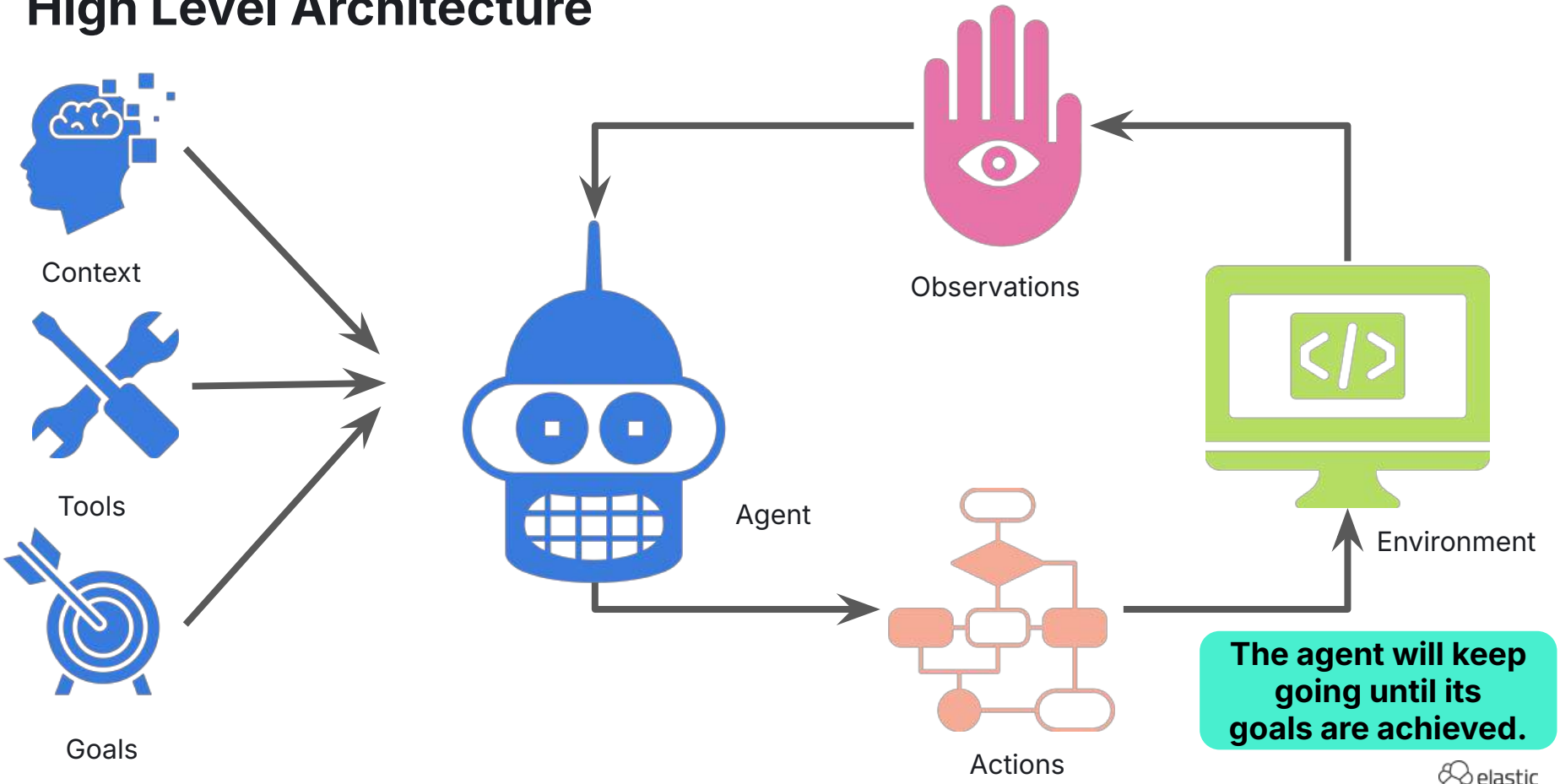
Resources

The assets that can be used to accomplish the goals

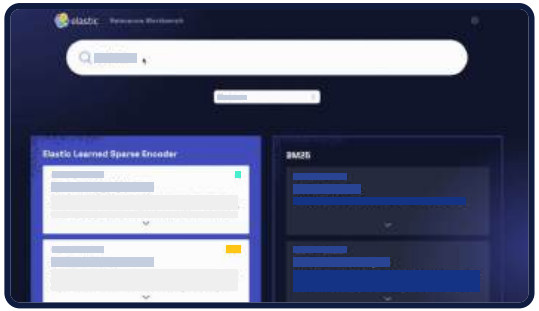
Context

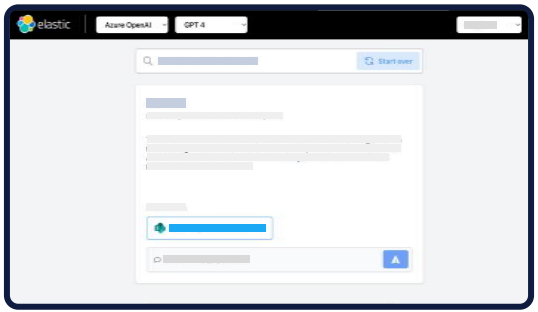
When, where, and why actions should be taken.

High Level Architecture



Search AI is the foundation for building **customer experiences of the future**





AI Assisted

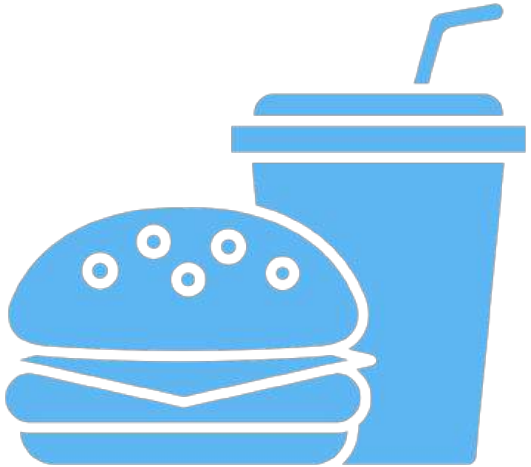
Semantic search applications

Human Assisted

RAG applications

Fully Agentic

Autonomous agents



Workflows

are about **predictable results** and **speed**, at **scale**.

Agents

are about **decisions**, **actions**, and **outcomes**.

elastic

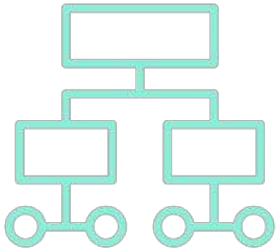
Workflows and Agents

Workflows are systems where LLMs and tools are orchestrated through predefined code paths (including **RAG**).

Agents are systems where LLMs:

- Dynamically direct their own processes
- Dynamically manage their own tool usage
- Maintain control over how they accomplish tasks.

Should I build a workflow or an agent?



Complexity

Tasks with a **defined decision tree** should be workflows, not agents.

Cost

Tasks with a low **transaction cost** (less than $1) should be workflows, not agents

Understanding

Tasks where all **steps are understood** should be workflows, not agents.

Risks

Tasks with a **high cost of errors** should be workflows (or include humans), not agents.

elastic

Today’s presentation:

01	02	03
Understanding the landscape	What Agents are and aren't	Where we fit

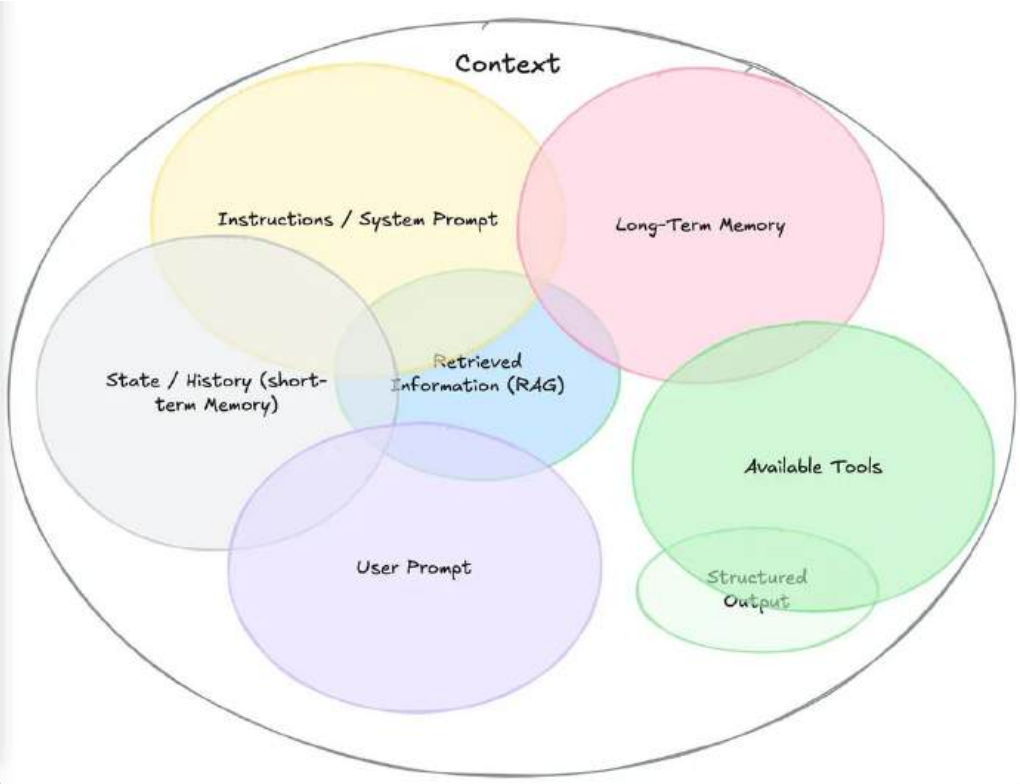
Context

Context is the information that shapes how an AI understands a task and produces its answer.

Context engineering is the process of giving AI the right information, in the right way, so it behaves accurately and helpfully.

Elastic is **a complete, differentiated platform for Context Engineering**.

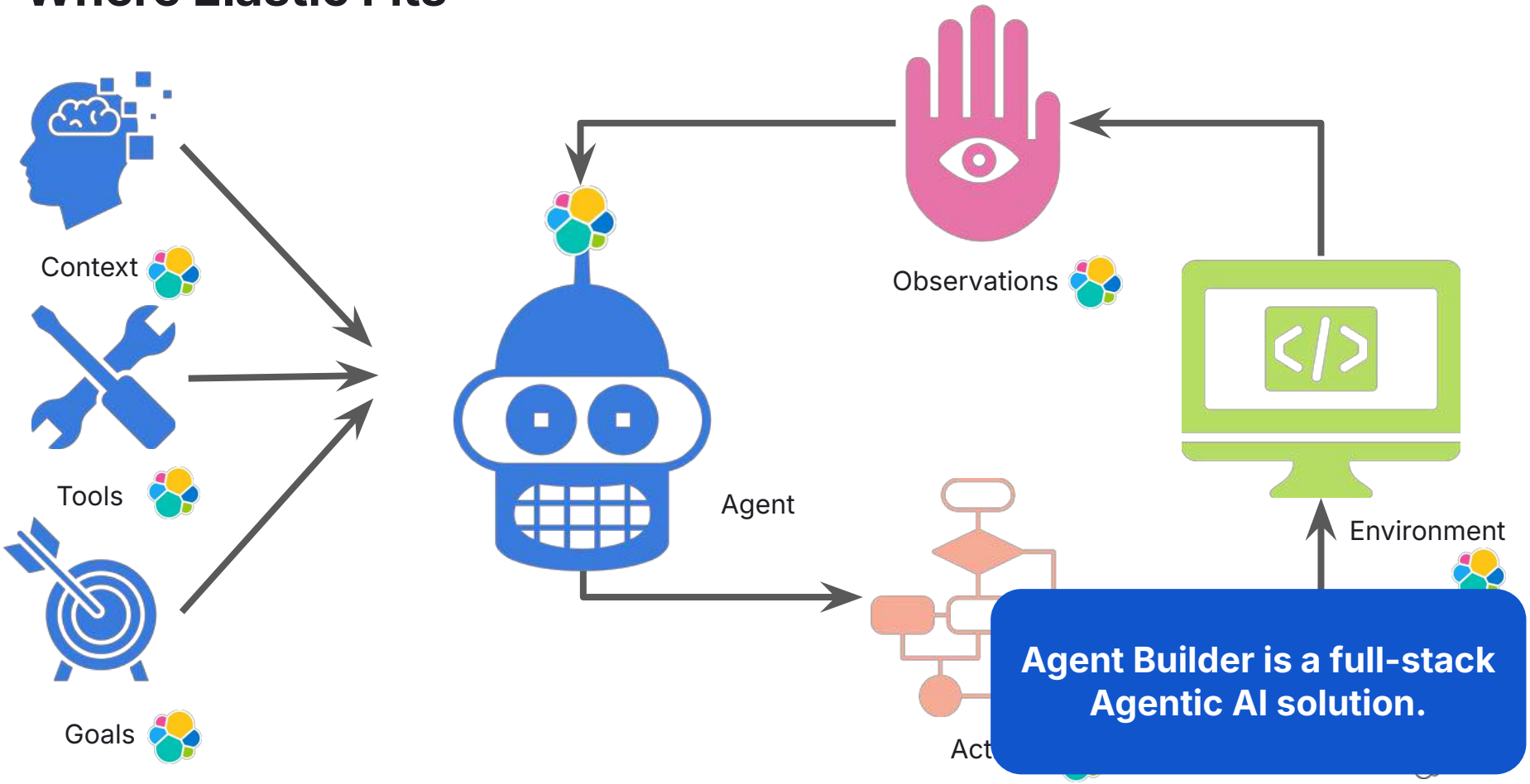
Components of Context Engineering

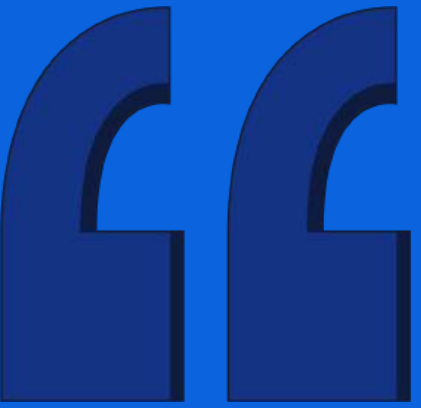


Elastic Blog: [What is context engineering?]

elastic

Where Elastic Fits





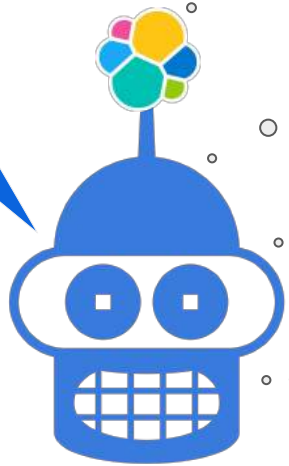
I need something warm but lightweight for a fall trip to Seattle.

How do we use the *context* effectively?

I need something warm but lightweight for a fall trip to Seattle.

Seattle's fall weather is notoriously unpredictable ...

Layering is absolutely essential!



What is the intent of the question?

- We need clothing.
- Fall means cooler weather.
- Travel means packable.
- Seattle has wet weather.

What context do I have or need?

- Consider previous conversations.
- I don't have weather data.
- Avoid out-of-stock items.
- I don't know if this is part of another order.
- Educate the customer, if possible.
- Are there any offers or discounts?

How do I get what I need?

- Products
- Inventory
- Offers
- Weather

What tools can I use?

- index_explorer
- search_offers
- search_conversations
- get_document
- search_catalog
- get_weather
- generate_query
- search_ratings

Can I give an answer with what I have?


Ratings


Offers


Items


Conversations


Weather

Today’s presentation:

01	02	03	04
Understanding the landscape	What Agents are and aren't	Where we fit	Where we shine

Context Rot

Context Rot is a challenge with **Large Language Models** where they struggle with more text.

An **LLM** “thinks about” every token in its context window before generating a new token. This means more opportunity for inaccuracies.

At 32K tokens, "11 models drop below 50% of their strong short-length baselines."

Context Rot is even more pronounced for complex, multi-step reasoning.

Giving an LLM more text can make it less accurate, slower, and more expensive.

[NOLIMA: Long-Context Evaluation Beyond Literal Matching]

elastic

More text *IS NOT*

the path to better AI.

Accurate search

is the path to better AI.

Recall

Recall measures the share of how many items within the top K position are relevant. It is **scored between 0 and 1** (higher is better).

Recall answers the question: Out of all the relevant items in the dataset, how many could you include in the top K?

Higher recall means less data is needed to get the results the users want. This means fewer tokens per operation and **lower cost of AI**.

Recall@5 = 4/7 = .57

K = 5

- “The Spice Girls”
- “Five-spice powder”
- “Scary Spice”
- “Spice World”
- “Ginger Spice”
- “Sporty Spice”
- “Ice Spice”
- “Romeo Beckham”
- “Christian Horner”
- “SpiceJet”

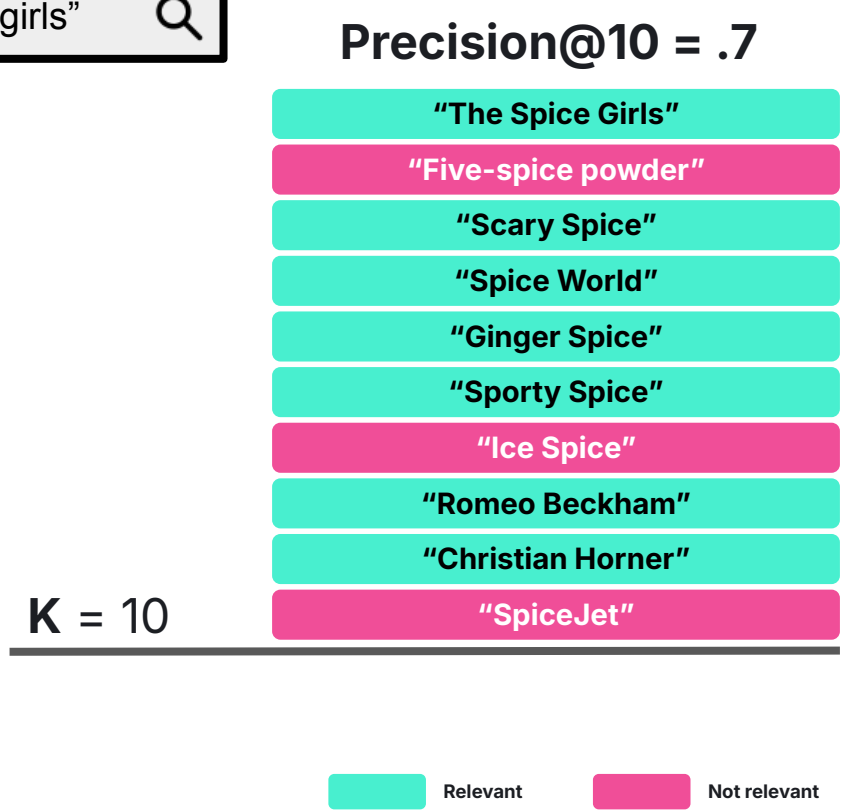
Relevant Not relevant

Precision

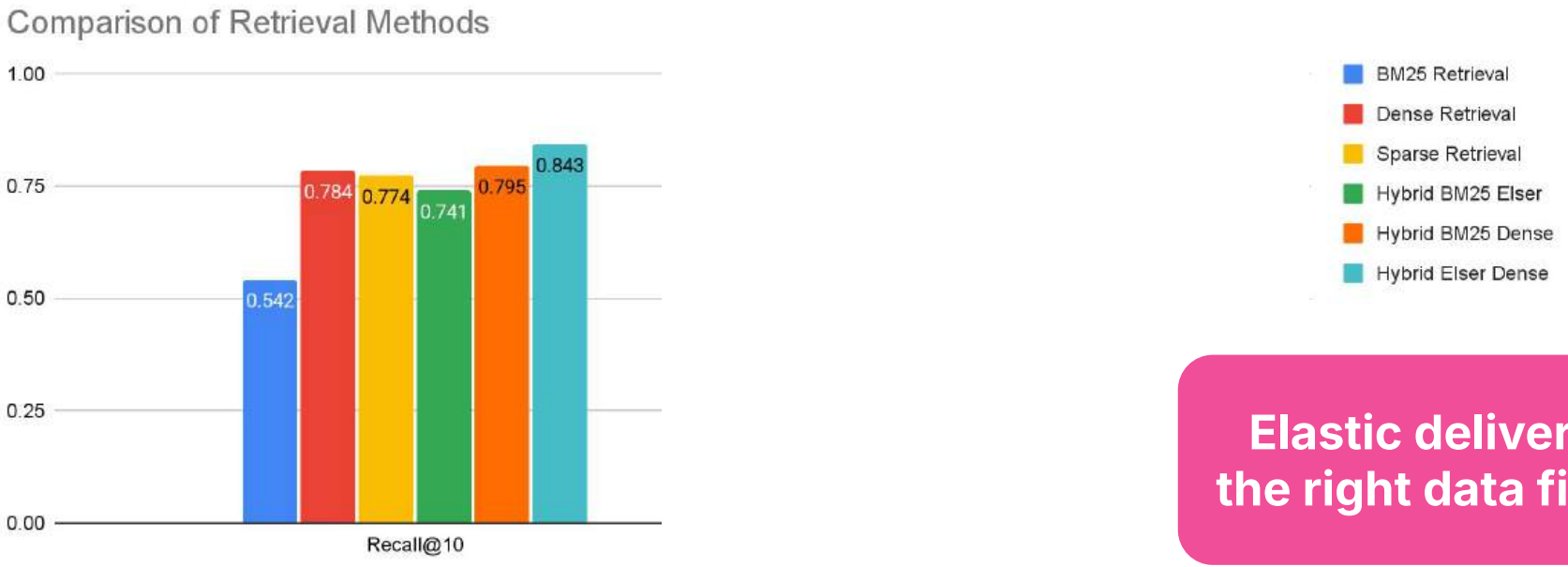
Precision is measured by the number of relevant items returned within a list of length K. It is **scored between 0 and 1** (higher is better).

Precision answers the question: Out of the top-k items suggested, how many are actually relevant to the user?

Higher precision means more accuracy, and less data for the same results. This means fewer tokens per operation and **lower cost of AI**.



Real World Performance



Elastic delivers the right data first

Elastic Blog: [The impact of relevance in context engineering for AI agents]

elastic

Why This Matters

1. Hybrid retrieval outperforms other strategies with **Recall of 84.3%** and **MRR of 0.53**. This means that Elasticsearch delivers the most accurate context engineering out-of the box.

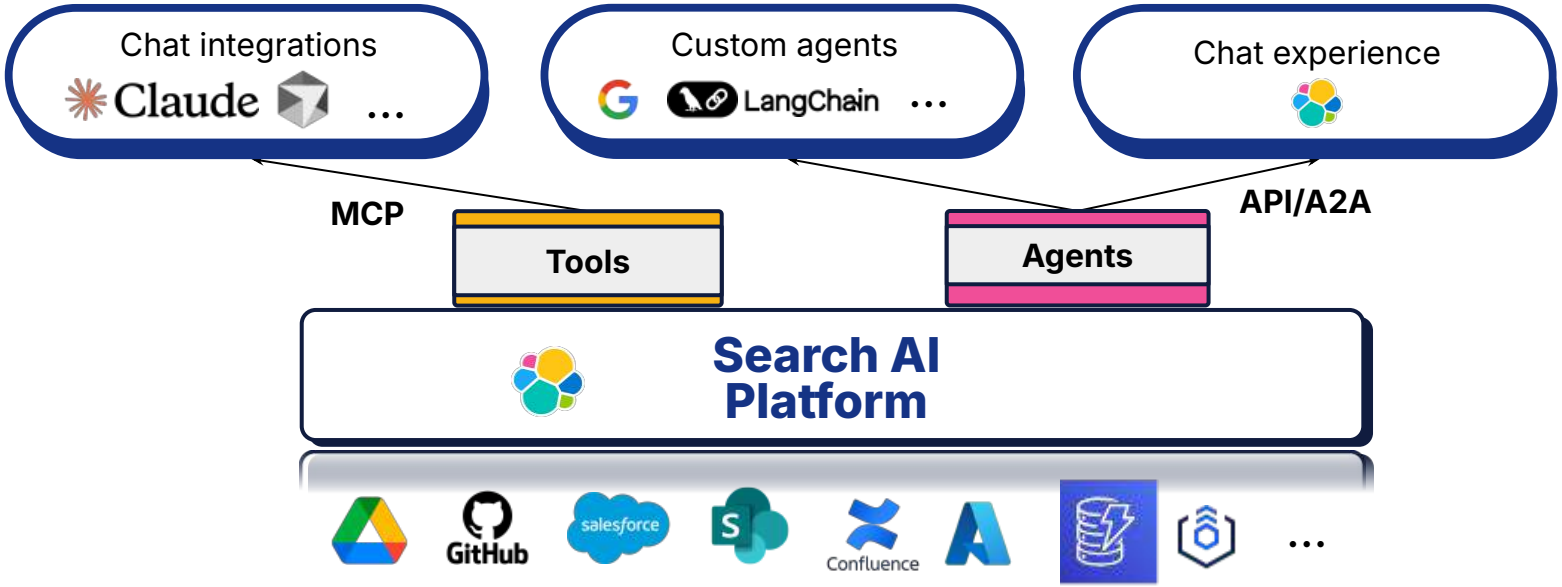
2. Semantic chunking had **a 93.3% hit rate** which **reduced context by 40%**. This means that leveraging some simple Elasticsearch features can significantly lower the cost of AI application.

3. Structured retrieval with ES|QL showed **a 100% success rate on agentic tasks.** This means that Agents built on Elasticsearch will be more accurate and have lower latency.

Elasticsearch makes AI smaller, faster, more accurate, and less expensive.

Elastic Blog: [The impact of relevance in contex](#)

Elastic Agent Builder Makes it Simple

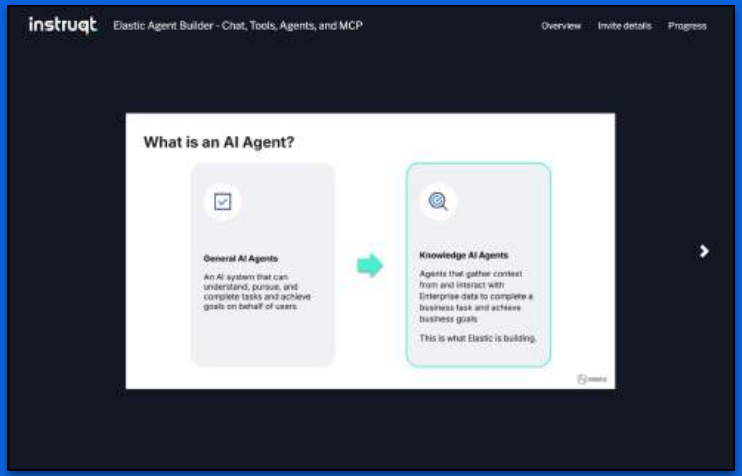


Today’s presentation:

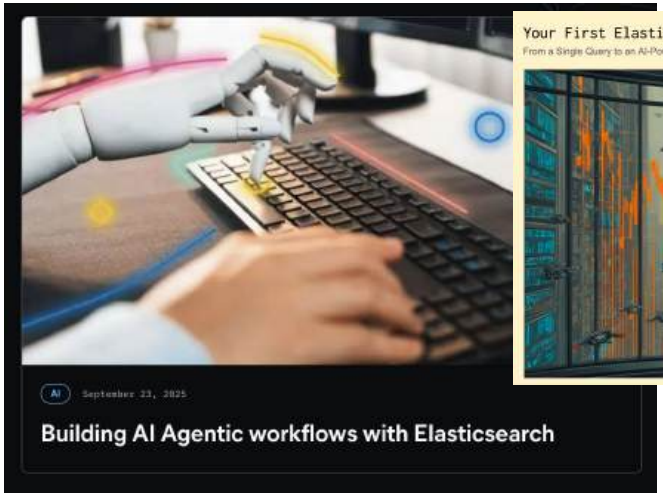
01	02	03	04	05
Understanding the landscape	What Agents are and aren't	Where we fit	Where we shine	Call to action

More Resources

[Elastic Agent Builder Workshop]



[ElasticON Slides]

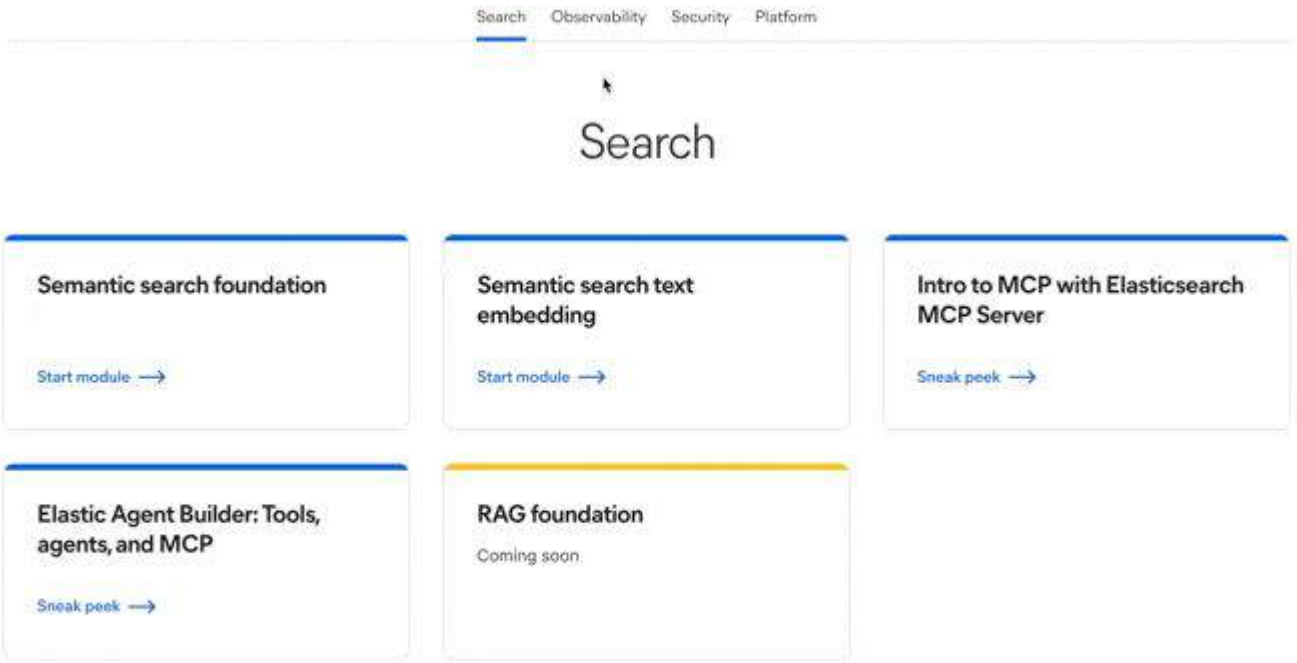


elastic

Training

www.elastic.co/training

It's Free



elastic

Speak at a future meetup

https://ela.st/chicago-meetups

elastic | The Search AI Company

Thank you!