



SEPT 2024

Knowledge Is the New Uranium



Nick Larew

Sr Engineer, Education AI, MongoDB



“Data is the new oil.”



Clive Humby

Mathematician & Data Scientist

2006



Knowledge is the new Uranium



Speaker



Nick Larew

SR ENGINEER – EDUCATION AI – MONGODB

MongoDB AI Chatbot

mongodb.com/docs

MongoDB Chatbot Framework

mongodb.github.io/chatbot

Evaluation & Training Data Sets

huggingface.co/mongodb-eai

AI Content Authoring Tools

Agenda

Data vs Knowledge

Use Cases

Nukes and Meltdowns

A Brighter Future

LLMs Run On Knowledge



Unsupervised Training

Get web scale knowledge from scraping, public APIs, and synthetic data

Retrieval-Augmented Generation

Enhance AI with domain-specific knowledge

Fine-tuning

Update models to match a style guide and use the latest information

Evaluation

Define and monitor optimal outcomes

AI Agents

Learn from knowledge and take action



Data vs Knowledge

```
{
  location: "San Francisco, CA, USA",
  forecast: [
    { day: "2024-09-10",
      temperature: { unit: "F", high: 68, low: 54 } },
    { day: "2024-09-11",
      temperature: { unit: "F", high: 70, low: 55 },
      precipitation: { unit: "in", value: 0.5 } },
    { day: "2024-09-12",
      temperature: { unit: "F", high: 76, low: 54 } },
    { day: "2024-09-13",
      temperature: { unit: "F", high: 71, low: 56 } }
  ]
}
```

“The weather for the rest of the week will range from 68°F to 76°F during the day, which is **generally mild**.”

At night, the temperature will be in the mid-50s and **it may be foggy**.”

You might want to **wear a jacket**, especially since there will be rain on Wednesday.”

Low Quality Knowledge

Low information/token density

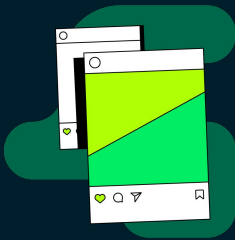
Irregular or illogical format

Spotty coverage

From unvetted sources



Chat Logs



Social Media Posts



SEO Spam

High Quality Knowledge

High information/token density

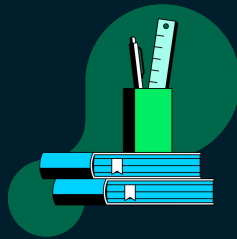
Logically formatted

Comprehensive coverage

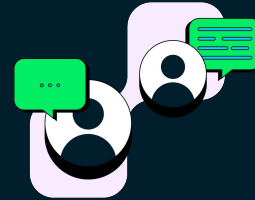
From trusted sources



Documentation

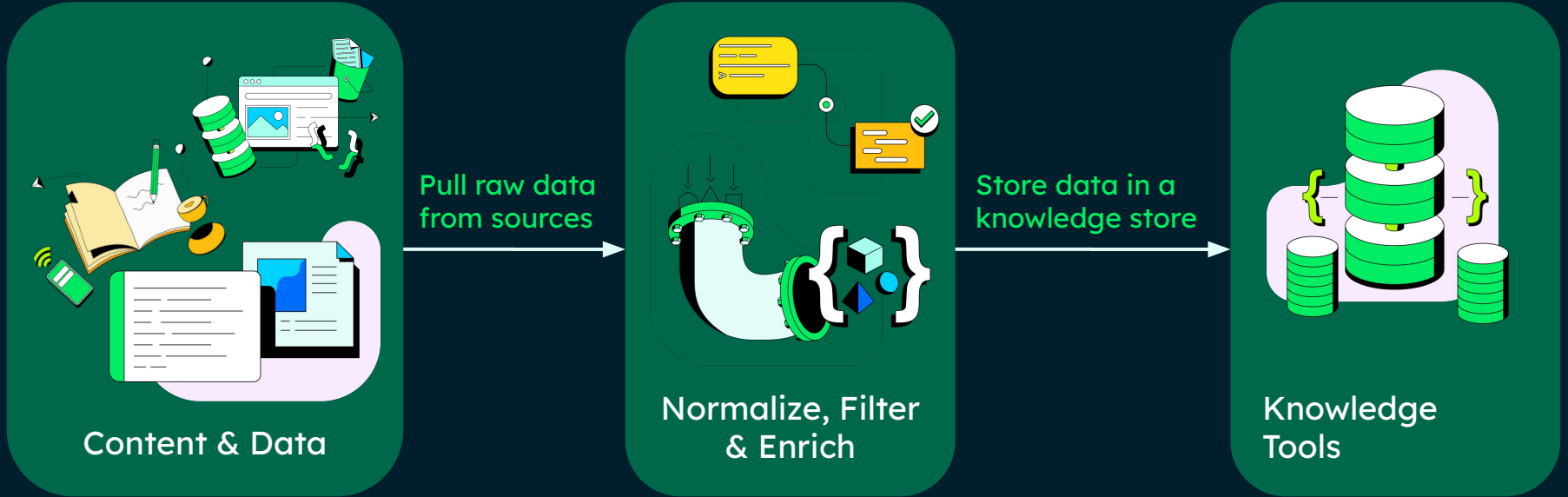


Books



FAQs

Mine & Refine





How We Use Knowledge



Chatbots

MongoDB Products Resources Solutions Company Pricing Support Sign In Try Free

Docs Home MongoDB Atlas Database Manual Migrations, Tools, and Connectors Client Libraries C C++ C# Go Java Kotlin Node.js PHP Python Ruby Rust Scala Swift TypeScript

EXPERIMENTAL

Why should I use Atlas Search?

Answer Verified By MongoDB Staff On May 13, 2024

MongoDB Atlas Search allows you to quickly build fine-grained search functionality into your applications. Consider using Atlas Search for the following reasons:

1. Relevance-based search: Atlas Search uses advanced search algorithms to provide highly relevant search results based on the search query and the data in your collections. This functionality ensures that users can easily find the information they are looking for.
2. Easy integration: Atlas Search seamlessly integrates with your existing MongoDB collections, allowing you to leverage the same MongoDB Query API that you are already familiar with. There's no need to maintain a separate search engine or synchronize data between different systems.
3. Flexible search capabilities: With Atlas Search, you can perform complex searches

Ask a Question Enter 0 / 300

This is an experimental generative AI chatbot. All information should be verified prior to use.

Migrations, Tools, and Connectors → Explore tools and integrations for MongoDB, from data visualization and development to migration and management.

Common First AI App

- High value, low effort
- Can be a standalone project

Ecosystem of Tools

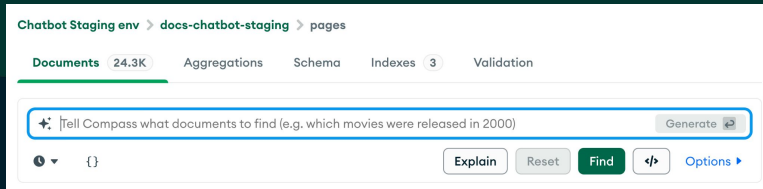
- Custom GPT, Credal.ai, etc
- MongoDB Chatbot Framework

Examples at MongoDB

- Docs AI Chatbot
- Sales Enablement Chatbot
- Content Drafting Tools



Product Features



An interface for unstructured requests

- Users can describe what they want without crafting domain specific queries
- Knowledge about the problem domain helps AI add structure around it

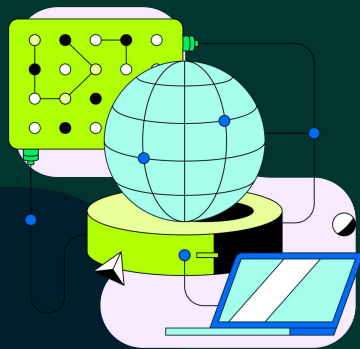
Highly personalizable

- Enrich user data & inputs with contextual knowledge
- Customize operations to a user's data, resources, etc.

Examples at MongoDB

- NLP query builder in Compass
- NLP chart builder in Atlas

Generative Content Tools



Apply AI to solve focused content problems

- Can solve a one-off issue or be used for iterative content pipelines

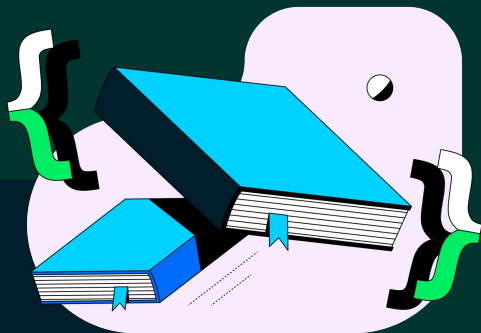
Compose complex workflows

- Perform multiple operations
- Chain together results
- Controlled structure, input, and output

Examples at MongoDB

- MongoDB Artifact Generator
 - Docs meta descriptions
 - Code translation
- GitHub Copilot

Training Datasets



Refined knowledge for training and fine-tuning models

- Aggregate, process, and synthesize data for consumption
- Users include model providers and external applications

Examples at MongoDB

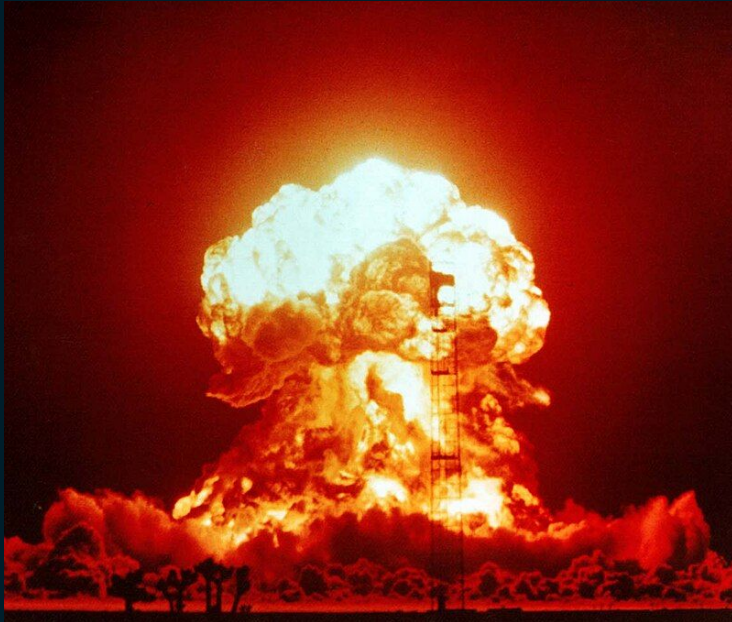
- Education content snapshots
- LLM-generated prompts for official code snippets
- Mock certification exams



Knowledge Nukes & Meltdowns



Nukes



Malicious uses of knowledge resources

- Exfiltrate confidential information
- Create disinformation

High potential for damage, but we can defend against this with robust and rigorous security measures.



Securing Knowledge

Avoid ingesting sensitive information in the first place

Enforce strong access controls on sensitive information you must ingest

Consider siloing sensitive information separately from public knowledge



Meltdowns

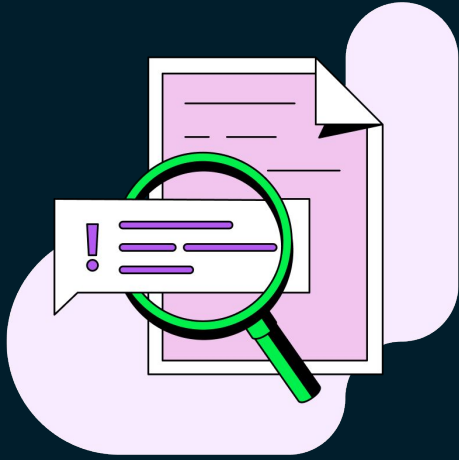


Unintended negative effects of knowledge resources

Accidental leaks of sensitive information

- In chatbot responses
- In a training dataset
- In a generated query

Can be harder to spot than malicious abuse and can have even higher impact

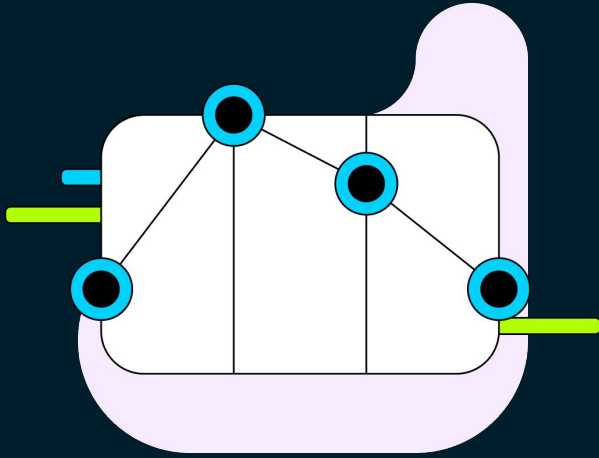


Evaluation & Observability

Evaluate the output of complex AI systems regularly

Apply runtime guardrails to prevent leaks

Monitor user feedback



External Benchmarks

Standardized, consistent evaluations of systems you don't directly control

Understand how different AI models and applications handle your use case

Analyze results to find and fill content gaps

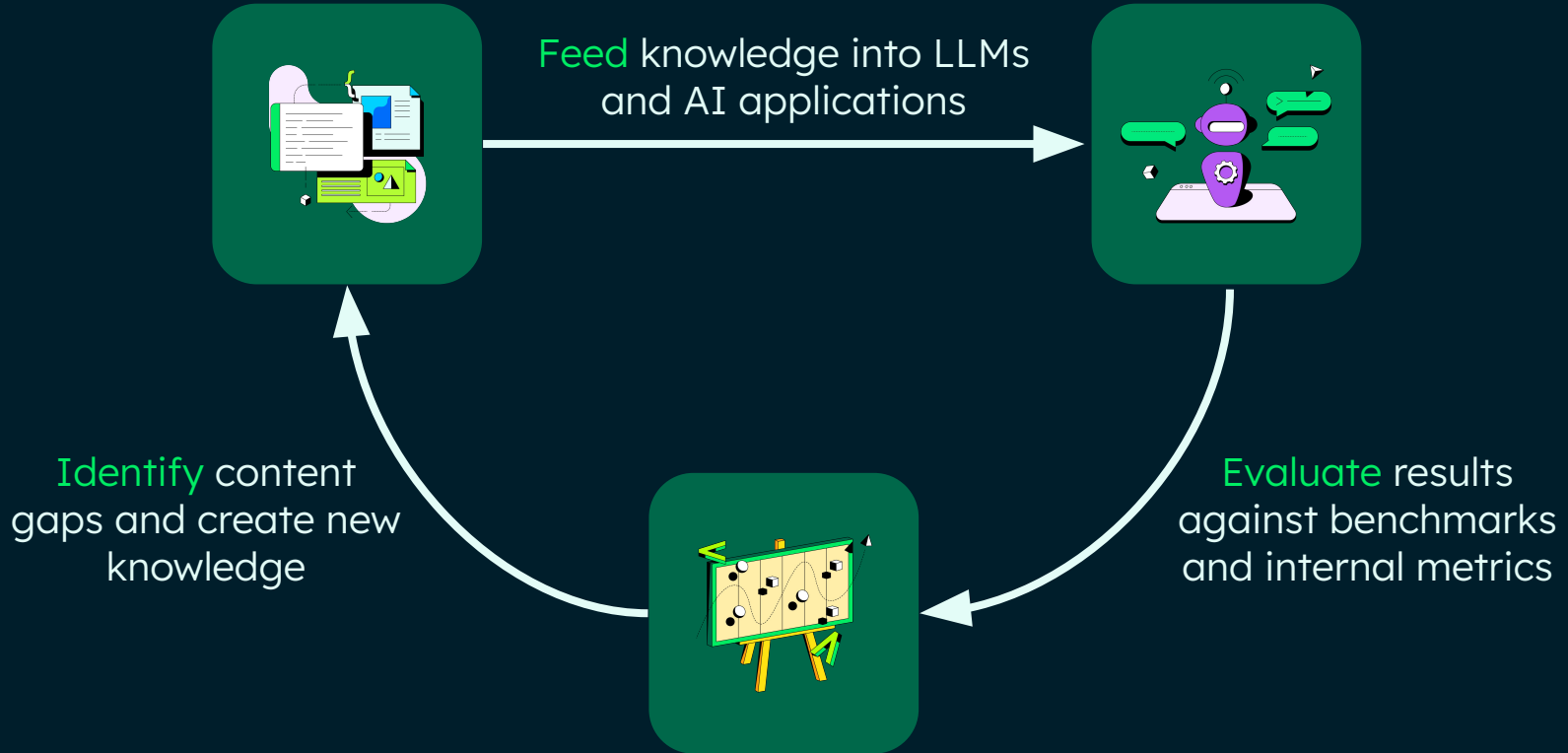
Examples from MongoDB

- Multiple choice MongoDB quiz questions
- Top MongoDB search queries
- MongoDB mentions in general queries



Knowledge
for a
Brighter
Future

The Knowledge Cycle



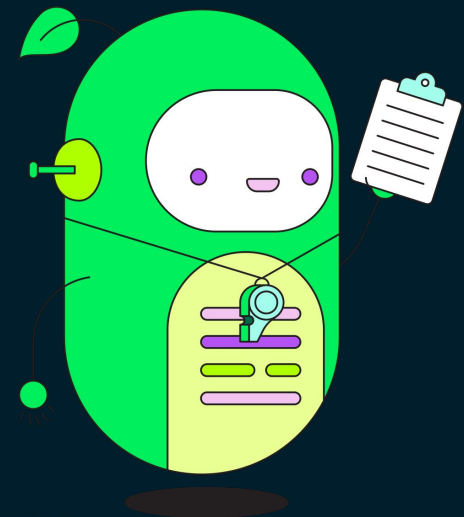


Human Builders

Internal and external developers

Create applications with traditional tools, AI, and domain specific knowledge

Often use 3rd party tools and plugins (e.g. custom GPTs, Github Copilot integration)



AI Agents

AI uses your knowledge resources for complex or long running tasks

Examples

- Internal HR assistant
- External agent builders

The Knowledge Service



API for programmatic access to knowledge

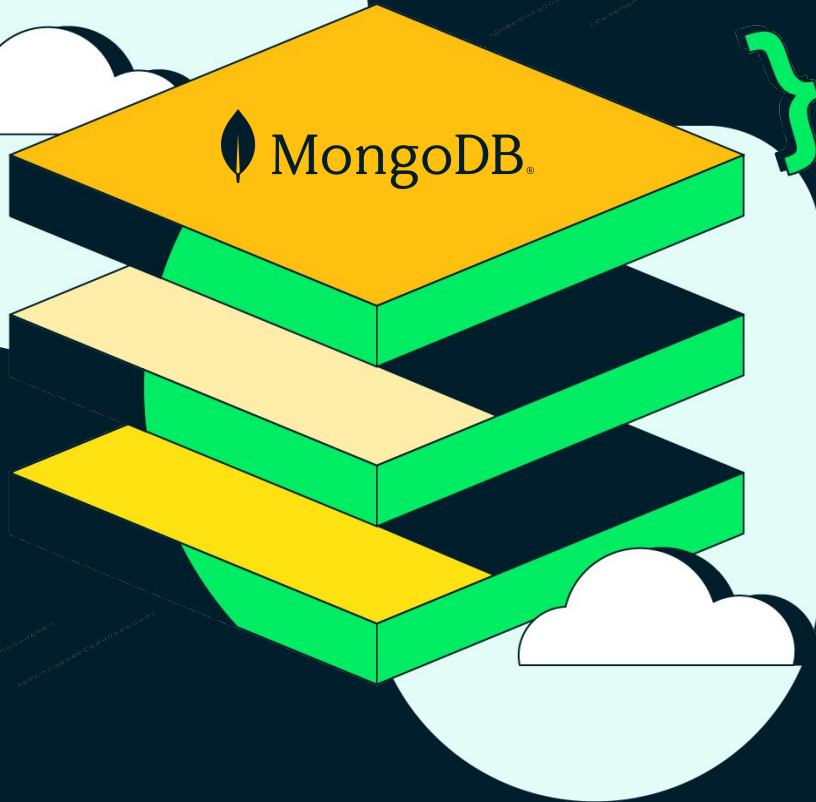
Use in chatbots, product features, dataset creation, agentic applications, and more

Interact with content according to your use case

- Get raw content
- Search knowledge base
- Chat with documents



Data Platform for Knowledge Resources



MongoDB Atlas as one platform to manage knowledge lifecycle

MongoDB Atlas

- Operational data store
- Atlas Vector Search
- Flexible document model
- Charts, Aggregation, and more!