HAYSTACK

The Keystone - 2023

Haystack EU 2023
René Kriegler, OpenSource Connections
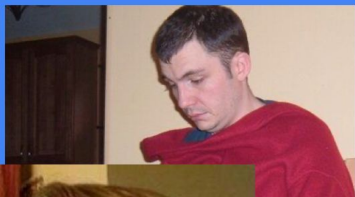
How and why it started

OpenSource Connections

The company behind Haystack

We help clients around the globe to get good search done.

Let's focus on Search Relevance!

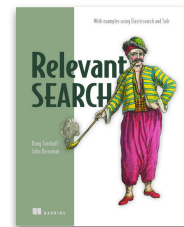

OpenSource Connections

pre 2015 life:

We want to focus our consulting on search relevance!

Nobody *gets* search relevance, they just want a pretty single-page search UI with some facets!
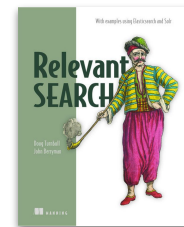
Doug Turnbull, Haystack Europe 2018 Keynote
https://haystackconf.com/europe2018/keynote/

Relevant Search
Book 2016

Let's focus on Search Relevance!

OpenSource Connections

Relevance talks (bbuzz)

The age of NoSQL & big data

Num Rel talks

1.5   2   5   8

2015   2016   2017   2018

Relevant Search
Book 2016

2018: Importance of Search Relevance was becoming increasingly recognised & we could help many teams with our expertise

The Keystone - 2018

Doug Turnbull's Keynote

A community driven by openness, sharing and open source technologies

OpenSource Connections

HAYSTACK

We need to commoditize repeat/busy work
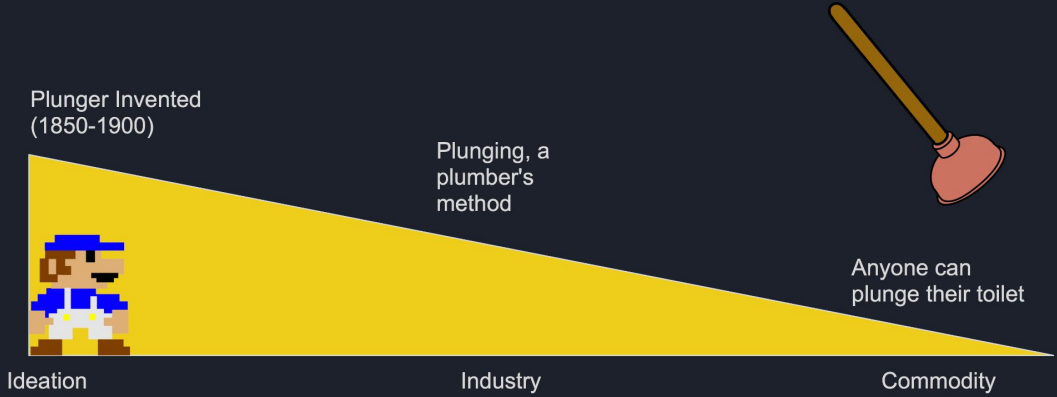
Plunger Invented
(1850-1900)

Plunging, a
plumber's
method

Anyone can
plunge their toilet

Ideation

Industry

Commodity

Doug Turnbull, Haystack 2018 Keynote https://haystackconf.com/2018/keystone/

#haystackconf
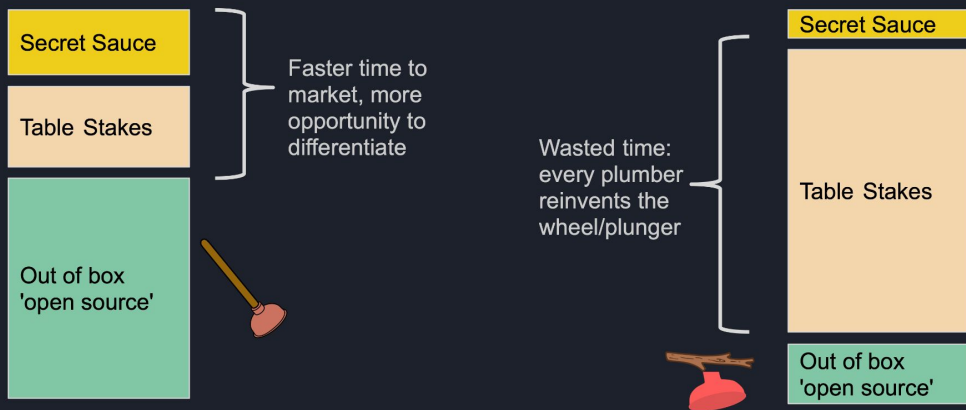
Doug Turnbull, Haystack 2018 Keynote https://haystackconf.com/2018/keystone/

#haystackconf

Let us build better open source tools for good search relevance!
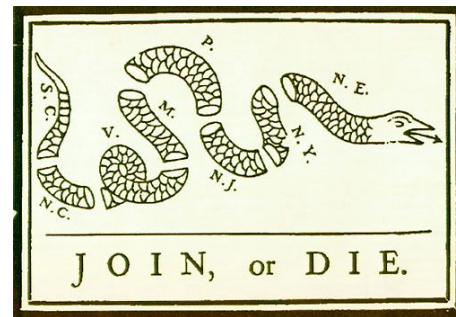
Doug Turnbull, Haystack 2018 Keynote https://haystackconf.com/2018/keystone/

And we meant it! ;-)

HAYSTACK - Five years on

10 major Haystack conferences

Haystack on Tour in Europe

Haystack Meetup - 1428 members
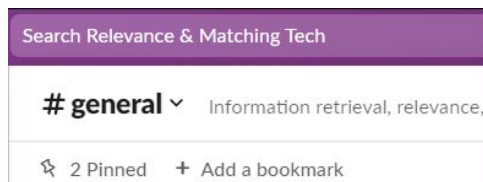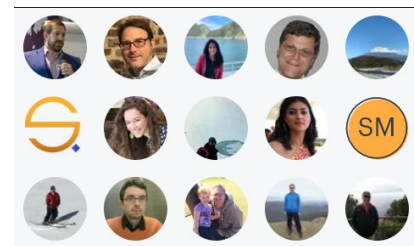
Relevance Slack - 4521 members

![HAYSTACK]

The focus is shifting
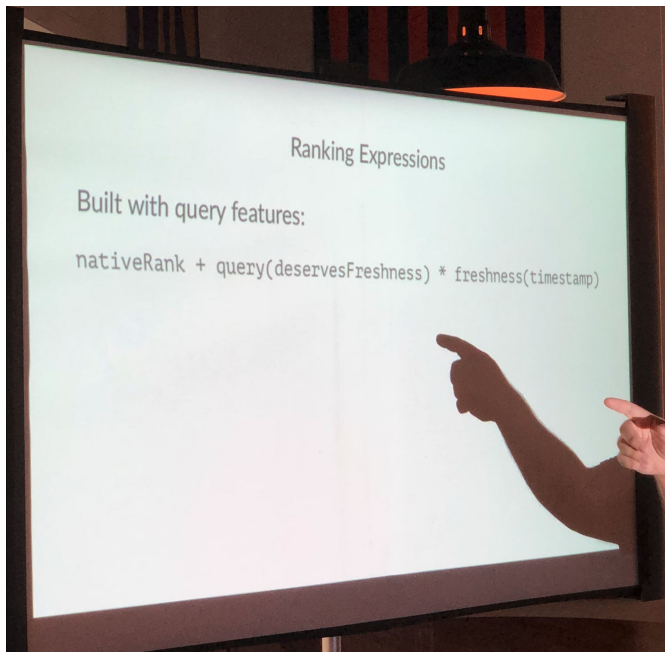
2018/19:
- Understanding search relevance
- Offline evaluation
- Click models
- Mastering relevance by tuning Solr, Elasticsearch
- Knowledge graphs
- NER
- LTR!
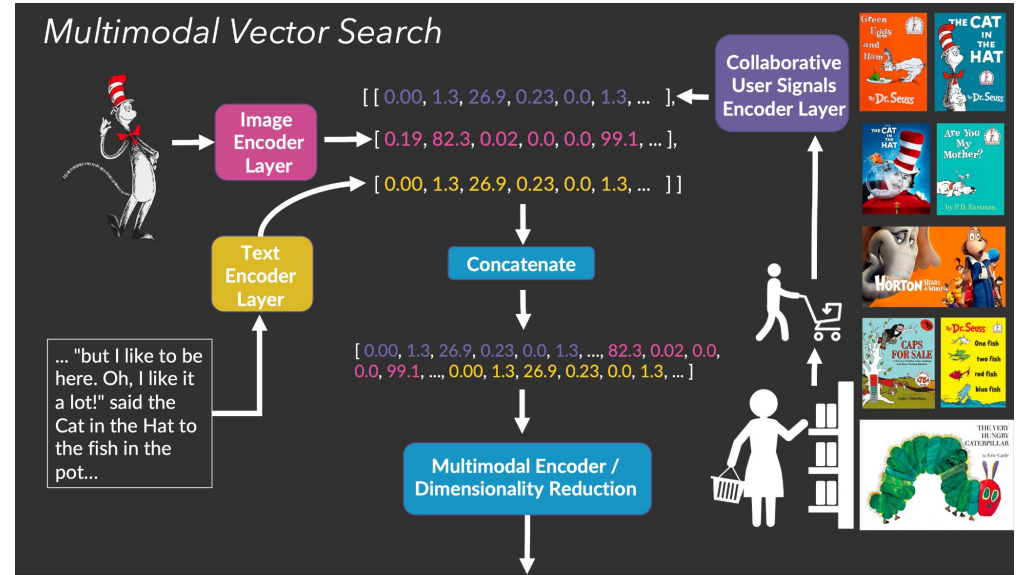
The focus is shifting

Ranking Expressions

Built with query features:

```
nativeRank + query(deservesFreshness) * freshness(timestamp)
```

Multimodal Vector Search

Trey Grainger. *Relevance in the Age of Generative Search*. Keynote - Haystack US 2023

HAYSTACK The focus is shifting

Shift towards AI
- Haystack US 2023: 8 out of 19 sessions on search-related AI, including keynote
- Haystack EU 2022: 4 out of 8, including keynote
- Key topic at "Haystack on Tour"events
- Talks reflecting
  - early adoption
  - fundamentals, new technologies & 'How to's

Trey Grainger. *Relevance in the Age of Generative Search*. Keynote - Haystack US 2023

HAYSTACK

The focus is shifting

Shift towards AI
- Haystack US 2023: 8 out of 19 sessions on search-related AI, including keynote
- Haystack EU 2022: 4 out of 8, including keynote
- Key topic at "Haystack on Tour"events
- Talks reflecting
  - early adoption
  - fundamentals, new technologies & 'How to's

Exciting times!
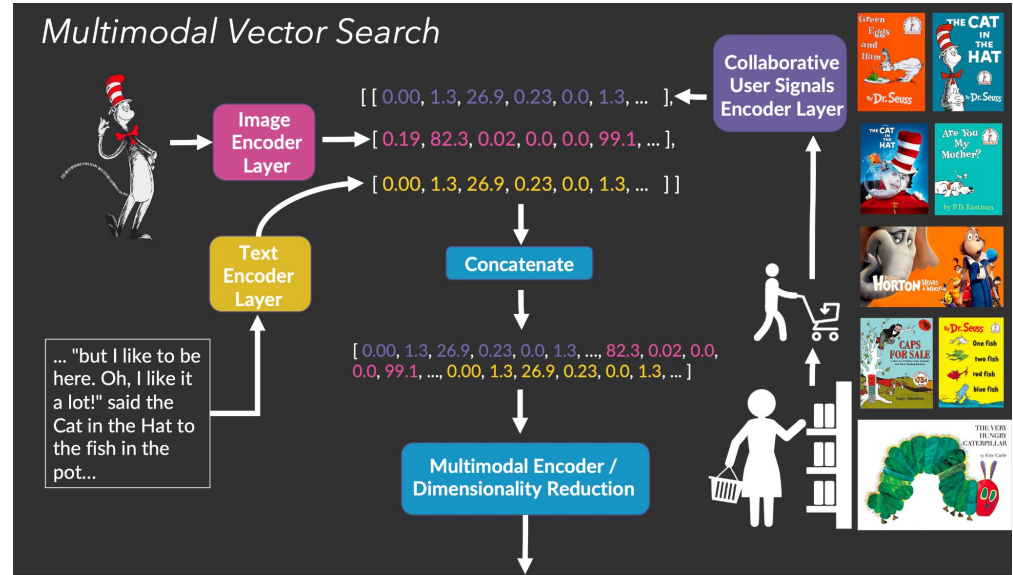
Trey Grainger. *Relevance in the Age of Generative Search*. Keynote - Haystack US 2023

AI for search is becoming ubiquitous!

Neural search - build
your own

We are witnesses and part of a paradigm shift in search

AI is becoming ubiquitous - at the verge of becoming a commodity

- Giant leap in AI solutions: conceptually, feasibility
- LLMs / generative AI have become available
- solutions for efficient ANN retrieval
- pretrained models, database/search platforms
- venture capital - evolving market

Neural search - build your own

AI & our community

HAYSTACK — Pushing the boundaries of 'search'

Search-related AI use cases (examples)

Semantic Retrieval	Vector search, result reranking
Question answering	Natural language understanding, fact retrieval
Query interpretation	Classify queries
Summarization	Summarize search result list, Summarize documents, passage highlighting
Retrieval Augmented Generation	Provide data to Conversational AI (Chatbots)
Content augmentation	Generate missing data, classify documents for search

HAYSTACK

Pushing the boundaries of 'search'

Search-related AI use cases (examples)

Semantic Retrieval

Question answering

Query interpretation

Summarization

Retrieval Augmented Generation

Content augmentation

Let's embrace these use cases in our community!

Just like 'traditional search' they enable our users to get a real-world thing done based on the information that we help them retrieve.
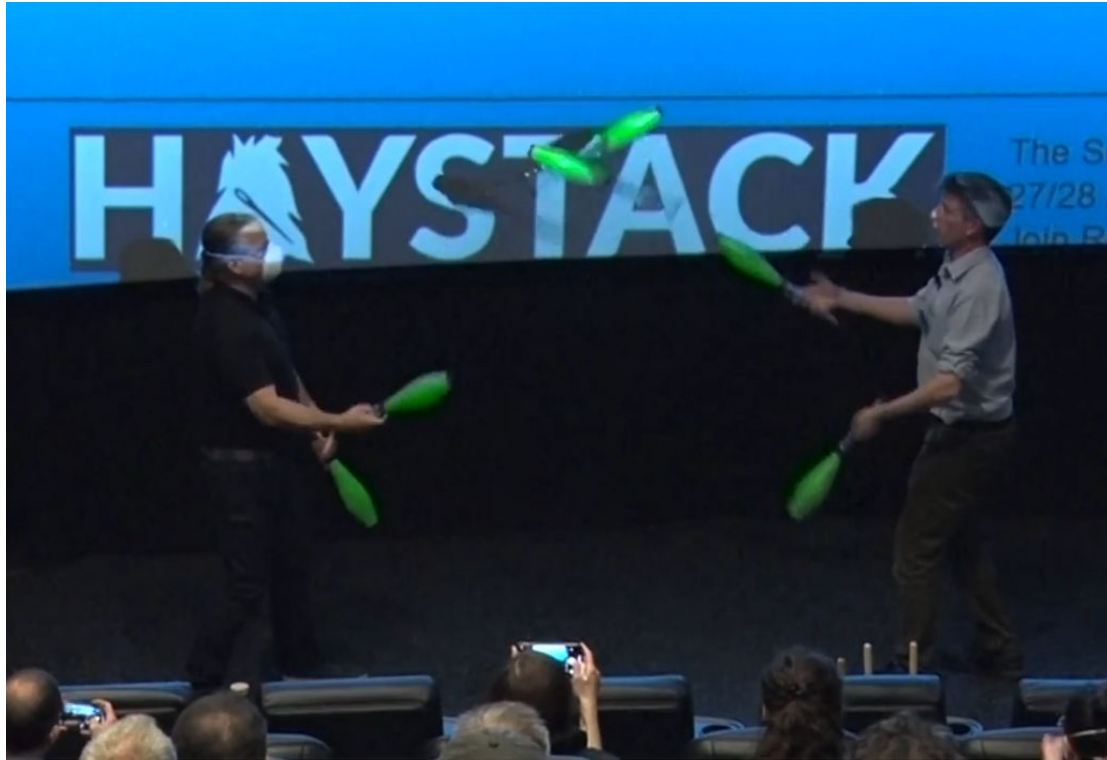
Let's be <u>jugglers, not magicians!</u>
A community built on openness and sharing
- even more important when conquering AI!

HAYSTACK Exciting times!

HAYSTACK Exciting times!

We are still in an early phase:

Gartner Hype Cycle for AI 2023: Foundation models & generative AI reaching Plateau in 5-10 years

But in search, companies invest as if we were here

expectations

| On the Rise | At the Peak | Sliding Into the Trough | Climbing the Slope | Entering the Plateau |

Activity beyond early adopters

Supplier proliferation

Negative press begins

Mass media hype begins

Supplier consolidation and failures

High-growth adoption phase starts: 20% to 30% of the potential audience has adopted the innovation

Early adopters investigate

Second/third rounds of venture capital funding

First-generation products, high price, lots of customization needed

Methodologies and best practices developing

Less than 5 percent of the potential audience has adopted fully

Startup companies first round of venture capital funding

Third-generation products, out of the box, product suites

R&D

Second-generation products, some services

| Technology Trigger | Peak of Inflated Expectations | Trough of Disillusionment | Slope of Enlightenment | Plateau of Productivity |

time

https://en.wikipedia.org/wiki/Gartner_hype_cycle

HAYSTACK Exciting times!

expectations

| On the Rise | At the Peak | Sliding Into the Trough | Climbing the Slope | Entering the Plateau |

Supplier proliferation

Activity beyond early adopters

Mass media hype begins

Negative press begins

Early adopters investigate

Supplier consolidation and failures

First-generation products, high price, lots of customization needed

Second/third rounds of venture capital funding

Startup companies first round of venture capital funding

Less than 5 percent of the potential audience has adopted fully

High-growth adoption phase starts: 20% to 30% of the potential audience has adopted the innovation

Methodologies and best practices developing

Third-generation products, out of the box, product suites

R&D

Second-generation products, some services

| Technology Trigger | Peak of Inflated Expectations | Trough of Disillusionment | Slope of Enlightenment | Plateau of Productivity |

time

https://en.wikipedia.org/wiki/Gartner_hype_cycle
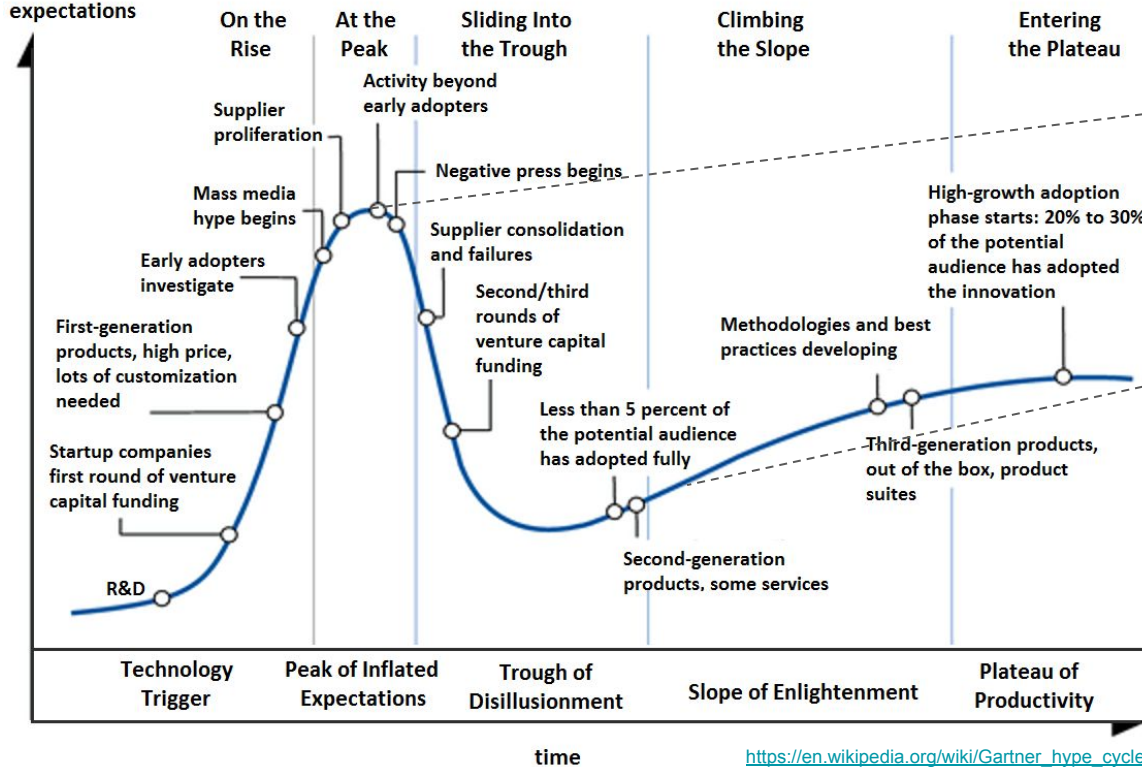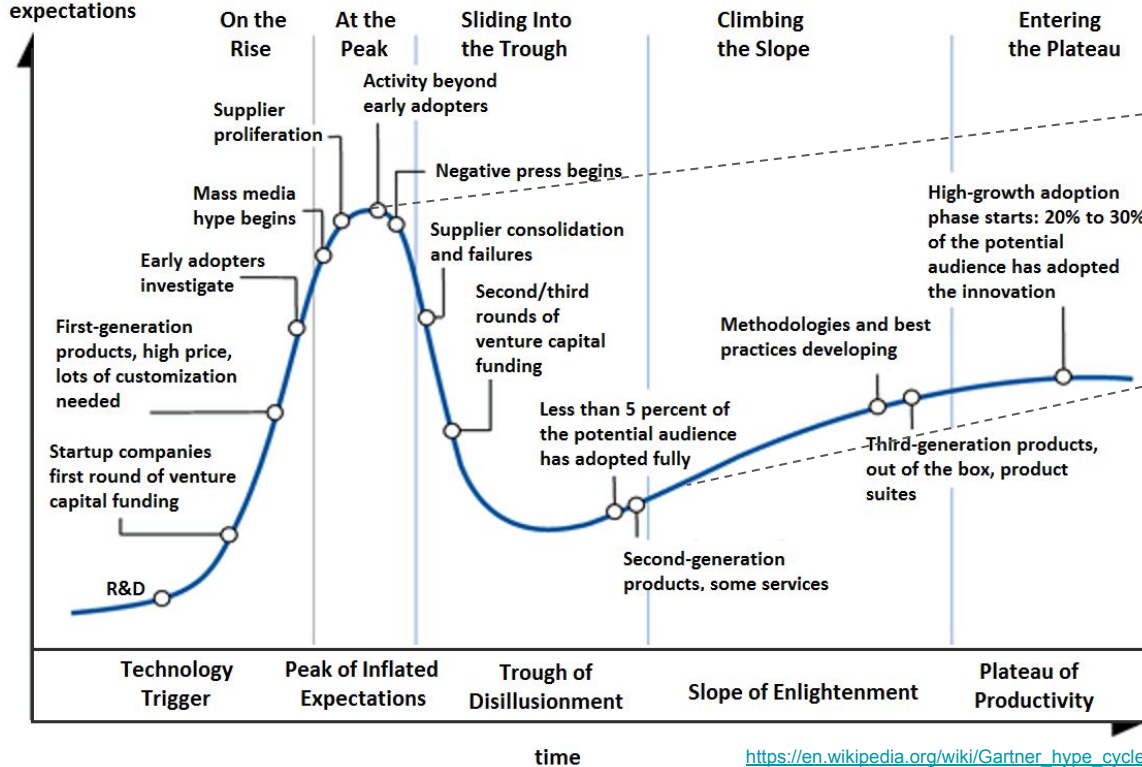
We are still in an early phase:

Gartner Hype Cycle for AI 2023: Foundation models & generative AI reaching Plateau in 5-10 years

But in search, companies invest as if we were here

Sharing insights within the community lets us all move faster

HAYSTACK

Conquering AI together

Creativity

Business

User Experience

Use Cases

Specifics of retrieval & generation task

Foundation models

Technology

There are many aspects of search-related AI that we need to learn and understand

HAYSTACK

Conquering AI together

Creativity

Business

User Experience

Use Cases

Specifics of retrieval & generation task

Foundation models

Technology

New databases (mainly Open Source!)

Vector Search in OpenSearch, Elasticsearch & Solr

Integration of Services & Platforms

Which one to use and how?

HAYSTACK

Conquering AI together

- Creativity
- Business
- User Experience
- Use Cases
- Specifics of retrieval & generation task
- Foundation models
- Technology

How to find the right pre-trained model

How to fine-tune? How to evaluate? How do they perform in production?

What's the process of their creation? - Open vs closed, ethics (e.g. biases and values), culture, license

HAYSTACK

Conquering AI together

Creativity

Business

User Experience

Use Cases

Specifics of retrieval & generation task

Foundation models

Technology

How to evaluate vector search, e.g. orthographic tolerance?

Do we need new metrics when search is used to provide context to a chatbot prompt?

How to measure whether the generated data was correct?

Shall we cut off nearest neighbour search at a certain vector search similarity threshold?

HAYSTACK

Conquering AI together

Creativity

Business

User Experience

Use Cases

Specifics of retrieval &
generation task

Foundation models

Technology

How to evaluate a chatbot across the conversational model
and the model for search?

Using vector search for long tail queries - do we really need
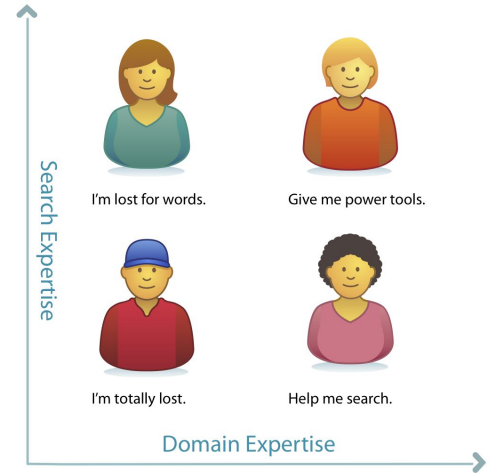to bring back precise results only?

HAYSTACK   Conquering AI together

Creativity

| Business |

| User Experience |

| Use Cases |

| Specifics of retrieval & generation task |

| Foundation models |

| Technology |
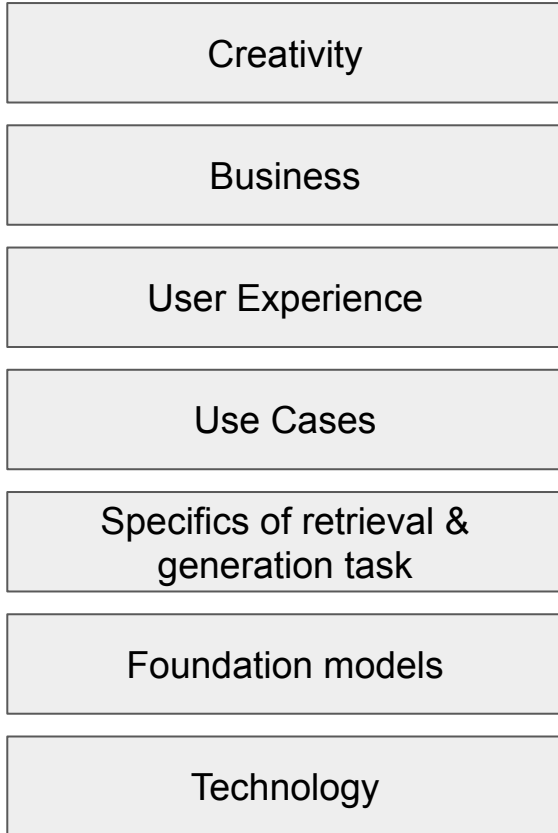
How can we build users' trust into our AI-based solution - and how can we avoid losing it?

When and why should we use AI?

How do we design for AI / principles of conversational UX

Search Expertise

I'm lost for words.   Give me power tools.

I'm totally lost.   Help me search.

Domain Expertise

Peter Morville, Expertise types, in: Search Patterns. 2010.

HAYSTACK

Conquering AI together

Creativity

Business

How does AI help business? - From customer experience to productivity

Could AI harm business? How could we avoid that?

User Experience

Use Cases

Specifics of retrieval & generation task

Foundation models

Technology

HAYSTACK

Conquering AI together

- Creativity
- Business
- User Experience
- Use Cases
- Specifics of retrieval & generation task
- Foundation models
- Technology

AI provides the tools to built entirely new solutions.

The expertise shared within the community empowers our creativity!

HAYSTACK The Keystone - 2023

The **scope has broadened** from 'traditional search' to **related use cases and AI**

We are moving **beyond search relevance** towards a more comprehensive **quality of results** in a variety of uses cases

Our community based on the **principles of openness and sharing** of insights will **guide us and help us conquer AI**!

HAYSTACK The Keystone - 2023

The **scope has broadened** from 'traditional search' to **related use cases and AI**

We are moving **beyond search relevance** towards a more comprehensive **quality of results** in a variety of uses cases

Our community based on the **principles of openness and sharing** of insights will **guide us and help us conquer AI**!

OpenSource Connections

And, BTW, at OpenSource Connections we are excited about this - providing guidance to our clients on search & AI is what we do!