Is a Large Language Model all you need for e-commerce Search?

ISIR-eCom 2023

Keld Lundgaard

About me

~5yrs experience in AI for e-commerce search @ Salesforce

Personalized/Semantic/Generative search

Assumptions

General knowledge about

Large Language Models (LLMs) & e-commerce search

Warning

Quickly evolving field!

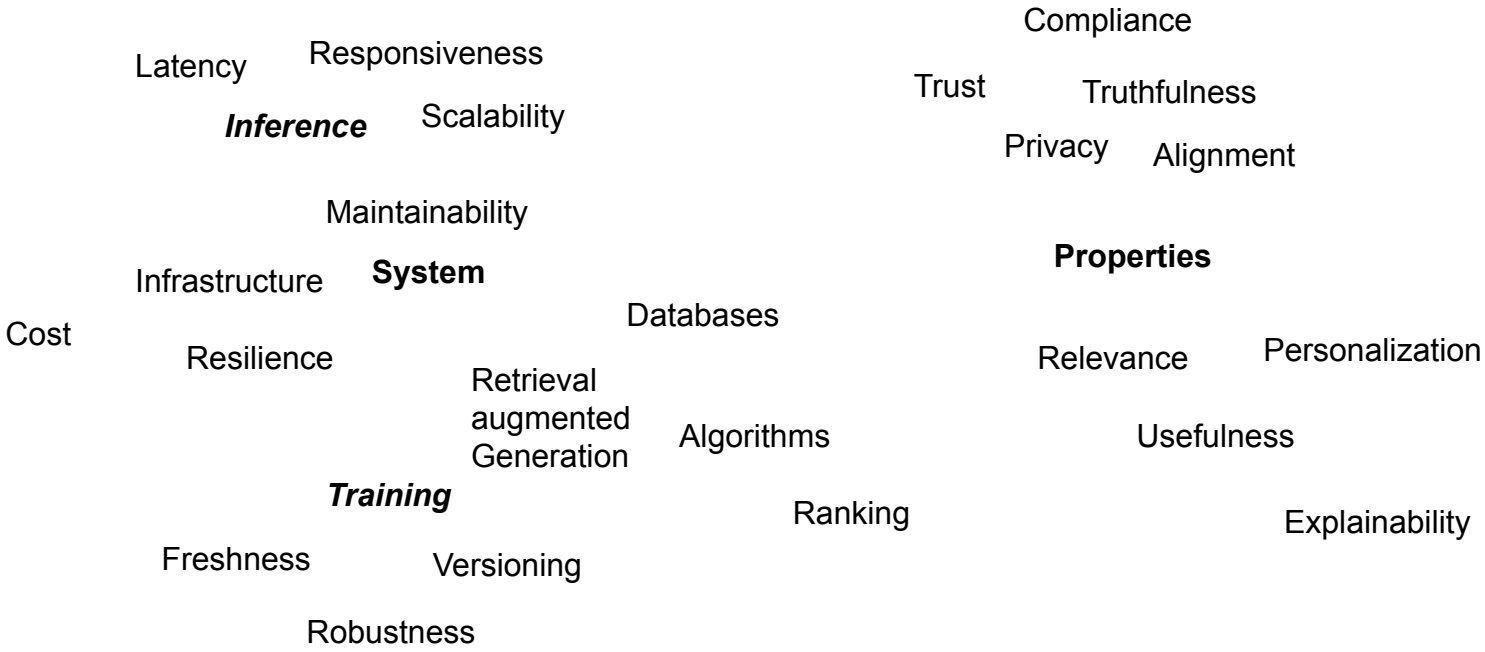
This talk offers questions, hopefully useful ones

Personal talk

Opinions are my own

Not a reflection of Salesforce

What will LLMs affect in search system?



JTBD: Job To Be Done

Shopper

- Find item → Solve problem → fulfill aspirations
  (dress Shoe) → (what to wear for event) → (feel respected and appreciated)

Store / search system

- Help shopper find stuff → sell stuff → maximize corporation’s long term profit

https://hbr.org/2016/09/know-your-customers-jobs-to-be-done

Signals

Product catalog (SKUs)
- Facts about products
Content
Business goals

Product images, name, tags,
Suggestions, AutoComplete,
Refinements, Recommendations,
…

Store

Shopper

Search queries (text, image, …)
Filters/sliders/toggles
Feedback: clicks/hover/scroll/…
Likes/dislikes, addToCart, Purchase
…

Shopper needs/intents
Budget, interests,
knowledge, fears,
aspirations, emotional
state

Difference between e-commerce and web search

E-commerce search

- Control over catalog representation
- Optimizing for conversion on site
- Works with rest of store

Web search

- Little/No-control over web catalog
- Optimize for Ads (Relevance when no Ads)
- Wants everything to stay in search (unless paid)

Towards generated search results

What you see was made for you

Static → **Generated**

Static	Semantic Search	Generated
Word/token matching	Deep learning embedding similarity	Generated output
Sparse TFIDF Matching	Two tower model: Query embedding || product embeddings	LLMs used to generate
BM25	Learning from relevant datasets pre-trained model → fine-tuning → distillation)	X-GPT models
Tuned using dictionaries: Synonyms, stopwords, …		Foundational model + tuning (RLHF)
		Learning from example behavior

What to generate?

Search suggestions: search query completion

Product tiles: name, image, tags

Explanation: For X problem, try these products, because Y reasons

Refinements/Dialog: create refinement option, follow up questions

Augmented reality: Here’s how product will work (E.g. couch in my living room)

Offline or online generation?

Offline → **Hybrid** → **Online**

Generated name/description/content

Per segment generation

Ability to use human-in-the-loop

(Cheaper/Safer)

Partly generated - only non-sensitive parts

Differentiation between generated and lookup?

*Toolformer

(compromise?)

Explanations for results

Dialog driven search

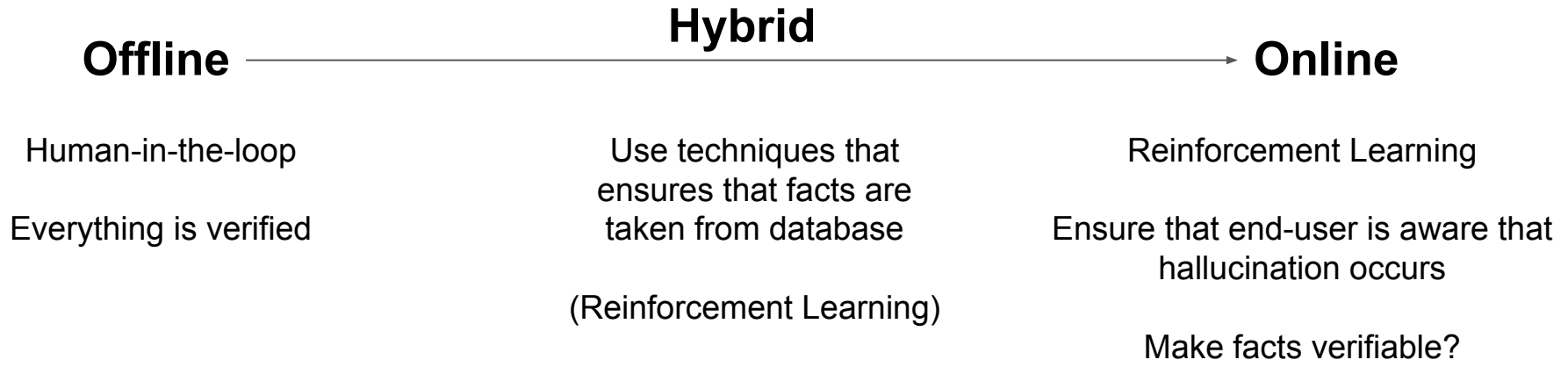
Fully personalized results

**Retrieval-Augmented Generation (RAG)

(most flexible and powerful)

*https://arxiv.org/abs/2302.04761 **https://arxiv.org/abs/2005.11401

How to prevent/deal with hallucination?



*https://arxiv.org/abs/2302.04761 **https://arxiv.org/abs/2005.11401

Matching → Semantic → Generated

Potential for results that are more

- Relevant
- Powerful
- Useful
- Converting

But also more

- complexity
- opaqueness
- risk of hallucination
- costly system

Truthfulness and alignment: Shopper vs Store

Store want to

- sell now instead of later
- have shopper buy more expensive option
- capture sale in store and not lose to competitor
- empathize product properties that makes shopper likely to buy than not

Like a sales representative on commission

Truthfulness and alignment: Shopper vs Store

Store also want to

- Keep the shopper coming back later - maximizing long term value
- Ensure that the brand value is not damaged
- Not get into any legal problems

So will want to make sure that the generation is truthful enough

Ethics of shopping assistant

Is it OK for it to

- Express empathy?
- Talking about its “own” experiences with X product?
- Inquire into sensitive information from the shopper?

When does personalization go too far?

We need to design for ethical and proper behavior.

Danger of optimizing system end to end

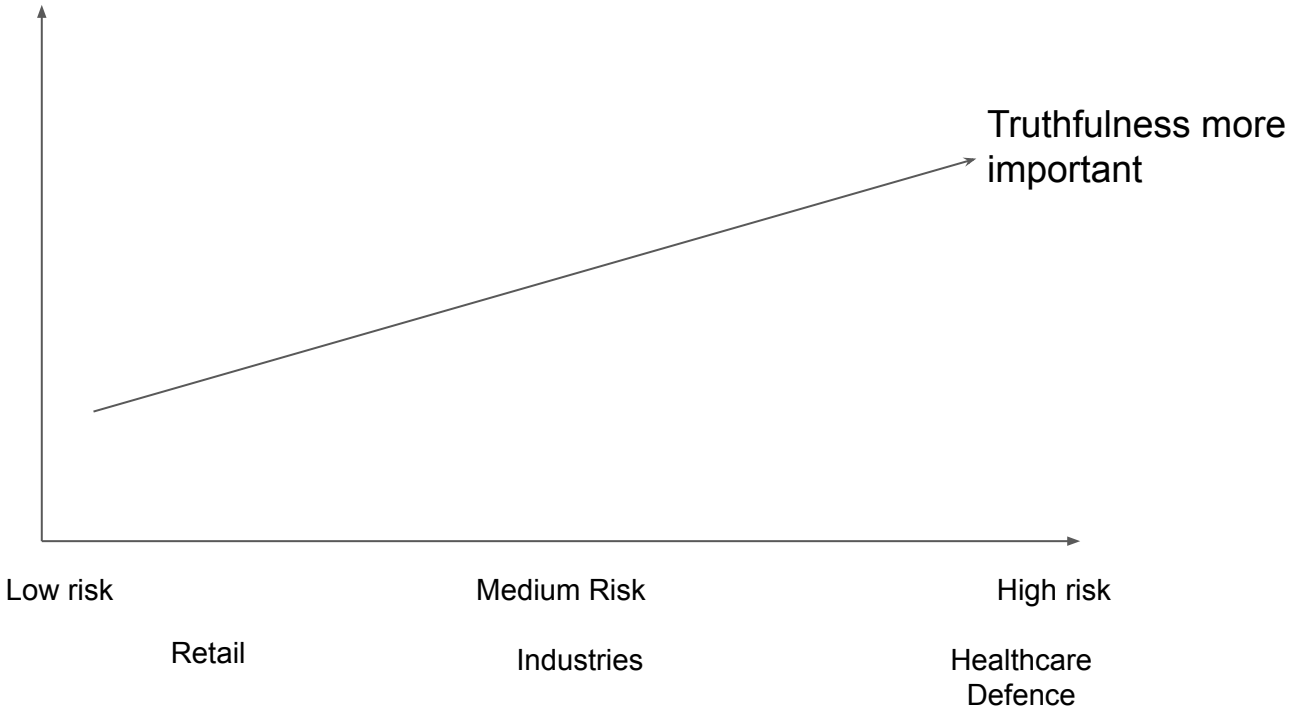
Feedback to system:

Clicks → purchases → returns? → reviews → satisfaction X years later

How to optimize for long term profitability and not short term reward

- Start with only restricted generation?
- Ensure able supervision of generation?
- Restrict how to learn from end-to-end feedback?

Importance of Truthfulness



Solution? Separation of facts from generation

Two systems:

- Fact (what is in the catalog)

- Content that represent facts: dialog, product grids, …

Allow shopper to determine what is fact and what is potentially fiction!

Or alternatively ensure that all facts can be verified.

How will LLMs change search in e-commerce?

Evolution: extension of current search

- Deeper context understanding
- More useful search results
- Done within the same interfaces as current systems

Revolution: changes the paradigm of search

- Changes the shopping experience fundamentally online
- Changes how/where e-commerce is being done
- Creates new shopping aggregators

Are LLMs all we need for Search?

IMHO: “Yes”, as long as we

- Ensure that facts about products are kept trustworthy and/or verifiable
- That the search system stays ethical
- System stays safe
- Focus on creating long-term value (not extracting)

Other considerations omitted

- Cost, complexity, architectures, algorithms, in-house vs utilizing APIs, sustainability

Thanks