BA INSIGHT

AI & Search

Two Great Tastes That Taste Great Together

by
Sid Probstein

Introduction

It seems that every enterprise search vendor is now claiming to have AI or cognitive capabilities. What exactly do those things mean? And why should you care? This white paper aims to define these frequently misunderstood terms, explain how to apply them to search, and then talk about some of the specific applications the combination can enhance or even automate.

> *"Rita stared at her monitor. She had been searching for background on two companies involved in a minor M&A transaction. And nothing was coming up. To be honest, she wasn't sure she had the names right. There appeared to have been a lot of consolidation in the space. But mapping out the relationships was hard. Ultimately, she needed something like what she'd seen in a movie, a system called Palantir that showed all the company relationships in a map. But there was no way to get that across the dozen financial news portals she had access to. She sighed and began printing out the documents. She'd have to grab an intern and map it out by hand to figure out where to focus…"*

For Starters, What is AI?

We can see what we would like it to be in Science Fiction … apparently autonomous, intelligent robots, in various sizes and shapes, real and virtual – that can help us. R2D2. C3P0. The Star Trek Computer.

Maybe not HAL.

Closer into our reality - we see bots that win quiz-shows, listening devices that map text to hand-written executions using cloud APIs, and numerous devices that can understand speech. But mostly, if you peel back even one onion layer, they are heavily human modelled, starting with a rule set, a list of watch words, a hard-wired set of inferences that make question answering across a sparse matrix of possible answers relatively simple. Are these AI?

From Wikipedia

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

> ***Inferences are steps in reasoning, moving from premises to logical consequences; etymologically, the word infer means to "carry forward". Inference is theoretically traditionally divided into deduction and induction, a distinction that in Europe dates at***

Privileged and confidential | © Copyright 2019 BA Insight. All rights reserved.

> ***least to Aristotle (300s BCE). Deduction is inference deriving logical conclusions from premises known or assumed to be true, with the laws of valid inference being studied in logic. Induction is inference from particular premises to a universal conclusion…***

From the search engine perspective:

> ***An inference system's job is to extend a knowledge base automatically. The knowledge base (KB) is a set of propositions that represent what the system knows about the world.***
>
> ***Inferencing is key element of AI - IBM's Watson used inferencing to excel at Jeopardy by determining what the target of each formulaic question was. For example, it correctly answered the question "the son of Italian immigrants he served in the House of Representatives before becoming NYC's mayor in 1934" correctly, partly by inferring that the target was a person who was mayor, not the son of the mayor, some other member of the House, or an immigrant.***

The reality is that they are AI 1.0. Or maybe 0.2, depending on how long a view you take. The key is that the combination of several technologies – machine learning (ML) and inferencing, at the core – plus modelled or observed behavior, frequently in the form of rules – are starting to solve problems that haven't been solved before, in ways that are powerfully different.

It's not that just we are creating AI that beats humans, today. What we are doing is creating AI that augments human capability; the combination is unbeatable. AI can link data and find patterns across quantities beyond human capacity. Then it can surface those patterns with examples, and show the impact of possible choices, to help the human make decisions.

> ***Today, AI absolutely can beat humans on path dependent problems like Chess and Go. But this is because the rules of those games are clear and quantifiable. In the enterprise, the rules are less clear. As Gary Kasparov said, "For machines to beat humans, it doesn't have to be perfect it just has to make less mistakes than a human. Humans eventually get less vigilant and make mistakes from tiredness. Machines don't."***
>
> More:
> https://www.theregister.co.uk/2018/05/10/heres_what_garry_kasparov_an_old_world_chess_champion_thinks_of_ai/

AI is already helpful in our consumer lives: helping us be on-time, reducing the effort it takes to write short replies in email, formatting our pasted content, identifying issues in our layers of formulae.

Privileged and confidential | © Copyright 2019 BA Insight. All rights reserved.

There's a bit of a drop-off when we go to work though, especially in the large enterprise. Apart from reminders of all kinds - we are largely on our own to get our jobs done. If we are in sales, customer acquisition or service – tracking oriented progressions – we might very well receive a steady stream of insight. But for internal purposes, trying to find training material, the latest version of a presentation, a contract or design document… there are few if any answers.

AI and Enterprise Search

Enterprise search may be claiming AI, but it sure doesn't feel like it when we still can't find a current report, a well-remembered email message or an article written by a colleague working on the same problem, a decade past. Demos of Alexa or chat bots returning the same results you'd get from.

The intranet has never been more impenetrable. Today's hot new results pages are packed with navigation options – source, topic, type, language, sentiment, key phrases, quotes, people, company names, query suggestions, more like this. These options help knowledge workers slide into increasingly large and complex result sets. They leave the problem of relevancy to the human.

Is this AI? Of course not. But it is a pre-requisite: creating paths for the user to select. Each selection, and the time spent on the resulting results, is a judgement that can be used to train a machine learning system and improve results.

Collecting the right data is key to training AI. From InfoWorld:

> ***If someone asked you for your customer and prospect data, would you give it to him or her? Your response would likely be, "Absolutely not." Your data is the crown jewel of your organization. It includes valuable information on key targets, their preferences, and motivations. What if you are trying to conduct predictive analytics to determine the likelihood of customers purchasing goods or services in the next six months based on their historical and real-time product usage, or to find why your quarterly sales figures missed the mark? Generic data just won't do. To get the answers to these types of company-focused questions, you really need the data that is most relevant to you—your own constantly updated data.***
>
> More: https://www.infoworld.com/article/3296044/data-is-the-lifeblood-of-ai-but-how-do-you-collect-it.html

Privileged and confidential | © Copyright 2019 BA Insight. All rights reserved.

Write a Better Query

One of the best applications for AI in search is to help the user write a better query. This can be developed quickly using session analysis; simply finding the trend in query refinement by looking at consumption pattern or rating can allow a learner to suggest the end of the chase, saving the user valuable time.

Bringing the user's role into the equation can then help the AI learn patterns for that specific function – suggesting a query with more or less recall, for example, or focused on a specific silo or type of data.

Relevancy, Relevancy, Relevancy

In search, relevancy is the equivalent of location in real estate. It sells itself, and users love it. They won't care what you call it.

Relevancy is relatively hard for a large group of users with a broad set of search missions. For that, you need to be able to infer the user's intent – for example, if it is an attempt at navigation, finding something previously seen, looking for the latest/greatest information, or exhaustively researching a topic. This requires a lot of data; but if you can predict it, you can adjust the approach used for relevancy and save time and effort.

Examples of good AI relevancy tend to be about picking the right fields for a single schema. And getting data into this schema is the key starting point. From there, the prevalent AI approach is to study the user query patterns and what they consume. If possible, the AI is trained from human rated examples; if not, time spent on a page must be observed. With that data, a learner can be trained to infer relevancy, adjusting the weights on fields and other search factors (like freshness) automatically.

Using machine learning you can get beyond document-centric scores and see what other, similar users consumed after running this query. Using metrics like time spent, and especially direct ratings like thumbs-up/thumbs-down or 1-5 scales, re-ranking results becomes a very real possibility.

One of the key points about AI is that it has to improve over time as it "learns" from feedback. Using ratings and observations of user behavior is one of the most important ways to do this. Capturing data around these areas is an essential step to developing more sophisticated AI in future.

> ***Ever wonder why R2D2 outwits apparently more powerful and capable droids throughout the entire Star Wars series? Many Sci-Fi fans believe this is because he never had a "memory wipe". This is an entirely logical inference based on current practice. Learning systems with the most data, and especially the most deep, historical data, tend to produce the most accurate results.***

Privileged and confidential | © Copyright 2019 BA Insight. All rights reserved.

Join the discussion here: https://scifi.stackexchange.com/questions/2089/why-was-r2-d2-smarter-than-other-similar-droids

Bots & Conversational Search

AI and search tend to converge when users need so-called conversational search. Very often this is a customer self-service touchpoint. The user's question is translated into a search query – often using simple "slotting" templates (e.g. Watson). For example, "Who is the X of Y?" from the user is re-written into the search query [person:X AND title:Y].

Collecting questions and answers, and how they were received, is another route to AI. Training a learner to infer the template, or identify entities like people and titles without templates, ultimately provides the sort of automation required for more "unstructured" situations – like customer service questions regarding complex products or legal issues.

Entities, Relationships & Links

AI is also excellent at inferring relationships, given some kind of structure in which to use it. An ontology of companies might provide a core way of relating owners and subsidiaries in various ways – depending on local laws.

Other data might provide AI with raw information to populate such a structure, thus giving it an understanding that company X is really a subsidiary of company Y which is owned by company Z. This could be given data, perhaps from a registry, or derived from the co-occurrences of company names in a large corpus of data – financial news, for example. This is just the sort of thing C3P0 might excel at.

Another valuable inference AI can provide to search are indirect linkages in language, such as synonyms, acronyms, portmanteaus and even slang. Sufficient training data and suitable NLP pre-processing can automate a lot of this – and, as an investment, leaves dictionary management approaches in the dust.

Why Search at All?

Users often ask – in the era of AI, why do I have to search at all? Why can't the search engine keep track of what I need to know?

Privileged and confidential | © Copyright 2019 BA Insight. All rights reserved.

It's the ultimate ask, of course. Consumers receive exactly this type of "intelligent" prioritization all the time. Facebook and LinkedIn are two excellent examples. And the power of such capability can't be overstated.

> ***The campaign was running 40,000 to 50,000 variants of its ads, testing how they performed in different formats, with subtitles and without, and static versus video, among other small differences. On the day of the third debate in October, the team ran 175,000 variations. "A/B testing on steroids." The more variations the team was able to produce… the higher the likelihood that its ads would actually be served to Facebook users.***
>
> More: https://www.wired.com/2016/11/facebook-won-trump-election-not-just-fake-news/

AI in the enterprise is starting to get traction with these use cases as well. Employees who keep their calendars or CRM systems up-to-date can be sent pre-built briefings consisting of primarily public data prior to sales meetings. (And Facebook, for the record, now offers an enterprise offering that reportedly includes the infamous "News Feed". More: https://www.facebook.com/workplace)

Today's enterprise search users expect the ability to save any query as an alert and notify them when something interesting and relevant is available for them to review. This requires a number of AI approaches to do well, including discerning the document's date, relevancy to the original query, and importance. Calculating surprise factor, as well as complex duplicate detection, underpin high quality implementations. (Users hate little more than incorrect notifications.)

Tag, You're It

Buyers of search have long known that clean content is essential to most search applications. AI equally depends on clean, analyzable data. But data in the enterprise tends to be dirty; minimally tagged, with metadata that is frequently unreliable - or, worse, always some corporate default.

One way to jump-start AI efforts is to focus on tagging content. Even a folksooomy – without a formal taxonomy behind it -Using a variety of techniques including NLP and what other users have tagged similar content, it's easy to build a system that suggests tags – and then improves as users either accept or reject and replace those suggestions.

Is a learner that you trust enough to tag your content without human review in 99% of cases AI? Users may not care, especially if they felt involved in training it.

This is another key for AI adoption: the best ones are built from user feedback; not despite it.

Privileged and confidential | © Copyright 2019 BA Insight. All rights reserved.

New Query Types

Perhaps the most important frontier for AI and Search are new, emerging query types. Today more than 60% of consumer queries are issued on mobile phones. (Source: https://www.statista.com/topics/2479/mobile-search/). These queries include location, and Google, for example, makes excellent use of them.

One particularly infamous enterprise search problem concerns a query for “lunch menu”. The company intranet includes a menu page for each cafeteria in the company’s dozen worldwide locations. As one might expect, the pre-AI search engine tended to return the most popular lunch menu page, that of the corporate headquarters with 5x the employees compared to location. The search engine vendor suggested applying the worker’s assigned location from their HR system. After a few month’s effort, this was implemented and improved results for that query… for users who did not travel.

Search engines already need AI to make sense of audio, image and video content. In the future, AI will help sort out and search through complex, disparate data sets by inferring links between them using concepts, not just entities (as noted above). They will use visual and haptic feedback from users to interactively refine search results. And they will present results in new, immersive formats like AR and VR.

Let’s start by collecting the data. And then we can live in a world that even more resembles science fiction!

> *“A few years later, Rita is relaxing in her backyard with her new iPhone. She’s spending the day preparing for an M&A review in the morning. She reflects on how much easier due-diligence has become since the new AI search engine was installed. Powered by their last few years of manual review and linkage, the engine seamlessly maps out company relationships – without human effort – and provides superb visualizations that let the team quickly determine what each link means and resolve it using an integrated case management system. It now takes a few days, instead of months and months. They now acquire companies at a much faster rate as a result. Even though they’re busier than ever, they have more time for decisions. ‘It’s a new world’, she thinks…”*

Further Reading

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

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

https://www.theringer.com/tech/2018/11/8/18069092/chess-alphazero-alphago-go-stockfish-artificial-intelligence-future

Privileged and confidential | © Copyright 2019 BA Insight. All rights reserved.

https://www.theregister.co.uk/2018/05/10/heres_what_garry_kasparov_an_old_world_chess_champion_thinks_of_ai/

https://www.forbes.com/sites/jeannemeister/2018/01/11/ai-plus-human-intelligence-is-the-future-of-work/#1ba593272bba

https://www.infoworld.com/article/3296044/data-is-the-lifeblood-of-ai-but-how-do-you-collect-it.html

https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

https://www.aol.com/2011/02/17/the-watson-supercomputer-isnt-always-perfect-you-say-tomato/

https://ark-invest.com/research/ibm-watson

https://www.wired.com/2016/11/facebook-won-trump-election-not-just-fake-news/

Privileged and confidential | © Copyright 2019 BA Insight. All rights reserved.