
Point-and-Shoot AI: Solving Enterprise's Unstructured Data Problem

You might not have heard of the expression “point-and-shoot” outside of choosing a camera - and that’s a good thing. In the enterprise AI world, [it's been a long road](#) from concept to useful technology. Even now, when AI is making a splash in [pharma](#), [finance](#), and [consulting](#), you most often hear specifically about deep learning or the open source approach, which work well for some data types and poorly for others.

At Coseer, we’ve spent years carefully taking the best elements of different AI paradigms to create Calibrated Quantum Mesh, our NLS-based algorithm. But we evaluated other methods first, and we learned a lot on our way to point-and-shoot capability. So how did we get here?

Deep learning doesn't work for unstructured data

Deep learning is a machine learning technique that uses deep or multilayer neural networks. To train, it needs 100s of thousands or even millions of data points where inputs and outputs are clearly tagged. This is great for click streams, images, IoT analytics, health records, etc.

But businesses don't work exclusively with this type of structured data. The trouble arises when we come to things that are inherently unstructured, like natural language. Think emails, scientific papers, meeting notes and customer service transcripts - all of the data that contains insight businesses are hoping to leverage, but can't get to in an efficient way. Trying to tackle this problem using deep learning means that the training data needs to be put into some kind of structure, often manually. This is a big stumbling block for using deep learning, especially in enterprise environments. Consider IBM Watson which uses deep learning-like algorithms. Training the engine takes a ton of expensive human-annotated data. And after months of data prep and upfront cost, there is still no guarantee that it will work. MD Anderson Cancer Center, an early adopter of Watson Oncology, [famously cancelled its partnership with IBM](#) after three years and nearly \$60 million in data preparation costs. Additionally, deep learning behaves like a “black box”- inputs go in, answers come out - but there is no traceability as to where decisions are made. It's easy to trick, and very difficult to trace, which can be prohibitive in an industry valuing transparency.

AI is well-suited for the challenge of unstructured data, but deep learning is impractical at best, potentially disastrous at worst. What else is there?

Open source solutions don't deliver on accuracy promises

Most easy to automate workflows have one thing in common - a single set of rules and parameters can cover 100% of possible scenarios. Unfortunately, many corporate processes aren't so black and white. Natural language processing itself is a classic probabilistic problem.

Why bring it up? In our quest to achieve point-and-shoot, we tested both [Stanford NLP](#) and [Ling-](#)

[Pipe](#), two of the leading open source software packages in Natural Language Processing (NLP). Both claim 98% accuracy, but if you ask them to process real-world data, the probabilistic nature of language means they only agree 70% of the time. [We actually tested this](#); asking both Stanford and Ling-Pipe to extract the noun phrases from a corpus consisting of one week's content posted by a popular Wall Street Twitter account.

These disappointing results led us back to the drawing board. Armed with data, we took another look at the landscape. Open source solutions do not deliver on the accuracy modern business needs, and deep learning fails on unstructured data even after a steep bill. Yet, a recent [MindMetre Research survey](#) of nearly 400 senior information management professionals in Europe and the US showed that 89% see insights from unstructured information as essential to gaining a competitive edge. Surprisingly, their biggest challenge is not data volume, it's the scatter. 71% report say that the information needed is scattered across different BUs and databases, and stored in different formats.

To recap: an ideal solution would be:

1. Highly accurate
2. Relatively quick and easy to deploy - minimal to no data prep needed, regardless of whether input is structured or unstructured
3. Able to integrate well into existing IT infrastructure, and handle disparate file formats with ease

Calibrated Quantum Mesh is developed to point-and-shoot

After much trial and error, Coseer developed a proprietary algorithm, called CQM (Calibrated Quantum Mesh). This NLS-based algo emulates human cognitive thought process, specializing in natural language, taking the best elements from neural networks and probabilistic graphical models.

The true power here is in its ability to correlate different ideas just like a human would. For example, after reading through a financial glossary it can say that "EBITDA" and "Profit" are related. These associations can be at various levels of abstraction. For example, reading through some earnings calls, the machine may conclude that "Profit" is associated with positive sentiment, so EBITDA must be positive, too. This way, a set of documents converts to concepts and the ideas contained in it. CQM keeps evolving as the system is used for ingesting new documents or for finding answers, so if someone searches and accepts the third answer instead of the first answer, Coseer will tweak the strengths of various edges in the mesh to learn from such behavior. This ability to use abstraction as a tool for real understanding is what differentiates it so thoroughly from existing AI, but there are logistical benefits as well.

Training CQM is easy; with various APIs and an infrastructure in place to ingest [virtually any file type](#), there is no need for data annotation, or even file conversion before the real work can begin. Because training is so quick, multiple iterations can be completed within days or weeks rather than months, leading to a much better end product at the end of deployment. An often

overlooked benefit of Coseer's engine is in security and explainability. Each Coseer deployment is 100% unique; there is no central repository continually being fed by our clients' aggregate data. Your data stays safe at all times, behind your firewall, and all decision points are logged for full traceability. This would be impossible with a deep learning algorithm.

With years of trial and effort to show for it, we are proud to have brought point-and-shoot AI to life. No more data annotation, no more suspect answers and low accuracy - just instant insights. Check out our [point-and-shoot page](#) to read further about how point-and-shoot is being deployed at Fortune 500s, or [set up a call](#) to learn how it can benefit your team.

What is Next-Generation Enterprise Search?

Coseer's search solutions are transforming industries from healthcare to finance. Our point-and-shoot AI trains finds answers and insights with 95%+ accuracy within 4-12 weeks - all of this in 100% security. The reason? We founded Coseer on the principle that computers should take care of the boring stuff so that humans can focus on creativity and judgment. To that end, we've built enterprise search solutions to complete complex workflows just as humans would in a fraction of the time. Fortune 500 leaders are using Coseer to speed up and automate their most complex work.

We follow a tactical approach to enterprise search:

- We deliver 95-98% accurate solutions within 4-12 weeks.
- Our solutions deploy entirely behind your own firewall for 100% security, and every decision point is logged for full transparency.
- You add the finishing touches, but our point-and-shoot AI practically trains itself. No more huge training data sets or time wasted annotating and tagging.

Visit our [website](#) for in-depth case studies, ROI breakdowns per industry, and other insight.