Singularity Systems White Paper

How to efficiently implement AI projects at scale to address real world business challenges?

Yingchao "YZ" Zhang, PhD

How to efficiently implement AI projects at scale to address real world business challenges?

It is said that data is the oil of 21st century. This is because data has hidden value to propel a business to the next level in global competition. About 80% of enterprise data is unstructured and thus not consumable by machines; like oil, however, these data can be refined to become more valuable. To capture that hidden value and leverage the data for business analytics, enterprises must develop the ability to clean their unstructured data to make them consumable. AI is exactly the right technology to process unstructured data: to drive business decisions and be consumed by downstream systems.

AI Projects Hard to Scale

Traditionally, AI projects are hard to scale and maintain. Usually a large amount of data must be labeled before training can start, which requires a large amount of manpower for a long period of time. If an enterprise is trying to implement multiple AI use cases in parallel, an army of data scientists, data science engineers and business people who know how to label the data will work on them together for months.

After the AI models are finally trained and deployed in production environments, they are hard to maintain. This is because business data is often dynamic and changes over time. The AI models and predictions will not be in synch with the latest trends of your data and will become even less accurate as more time passes and the data continues to evolve. The team of data scientists, engineers and SMEs must re-assemble periodically to retrain the models with yet again a large amount of labeled data.

The latest developments in deep learning, for example BERT for NLP & deep convolution networks for image and vision use cases, along with transfer learning and online learning can help address these challenges. It takes **three steps** for training and maintenance to occur in a more systemic and scalable fashion.

How a Human Learns

Imagine how a Subject Matter Expert (SME) learns and works on document processing. For the **first step**, s/he needs to understand the language in which the document is written as well as the domain knowledge for the particular industry. For the **second step**, s/he must

understand the requirements of the business, so s/he can learn quickly how to process the documents with a small amount of annotated document samples. For the **third step**, if and when new variance of data comes into the system, s/he can learn quickly how to deal with them based on input from the business and begin to process them correctly.

How a Machine Learns

Now imagine recreating these three steps using the latest approaches in Machine Learning. The **first step** is to understand the language in which the document is written and combine that with domain expertise. To achieve that we feed all available data into a backbone model using a technique called Bidirectional Encoder Representations from Transformers (BERT).

Transformers are a type of neural network architecture for neural machine translation that represents a great leap forward for Natural Language Processing. The platform knows the language because it starts with backbone models that have been trained on massive amounts of data (think of all of the text available online and every published document ever written.). Advances in computational power make it possible to leverage these powerful neural networks in a cost-effective and time-efficient manner.

> WE ACHIEVE THE DOMAIN KNOWLEDGE BY FEEDING THE PLATFORM INDUSTRY-SPECIFIC DOCUMENTS… RESULTING IN SUBSTANTIAL ACCURACY IMPROVEMENTS COMPARED TO TRAINING THESE DATASETS FROM SCRATCH.

We achieve the domain knowledge by feeding the platform industry-specific documents, like safety data sheets and leases for Oil & Gas. This learning is unsupervised, which means a human does not have to label any data. Through this process we build a foundation of domain-specific knowledge that makes it much easier to fine tune models to address specific requirements, resulting in substantial accuracy improvements compared to training these datasets from scratch.

Transfer Learning

Once we have established this backbone model, trained on industry-specific knowledge, it is ready to apply this knowledge to particular use cases. This is the **second step**, where transfer learning kicks in: where a human will provide a small amount of labeled data, "gold" data, to define the requirements of the particular use case. To stay with our Oil & Gas industry example, the human will define exactly what information needs to be extracted from the contracts. A model can be fine- tuned rather quickly to address the given use case accurately. Many use cases based on the same backbone model can be worked on in parallel, easily and quickly, to make it scalable for the enterprise to work on multiple AI projects concurrently across its business.

The backbone models enable transfer learning to make the fine tuning for a specific use case or document schema much more efficient. But the learning from each use case feeds back to improve the backbone model. As a company develops more and more use cases, not only does it deliver more value to the business but it increases the power and accuracy of the backbone models, which in turn makes the creation of additional document schemas even more efficient in a continuous virtuous cycle.

MANY USE CASES BASED ON THE SAME BACKBONE MODEL CAN BE WORKED ON IN PARALLEL, EASILY AND QUICKLY, TO MAKE IT SCALABLE FOR THE ENTERPRISE TO WORK ON MULTIPLE AI PROJECTS CONCURRENTLY ACROSS ITS BUSINESS.

The **third step** is to allow the model to keep up with changes in production data. Our patent-pending Real-Time AI model accelerates both the transfer learning in step 2, when a human worker defines the use case requirements, and the model maintenance in step 3, when a human-in-the-loop continues to interact in real time with the model to verify results.

Data evolves because business is dynamic. Data models that do not have the ability adapt to new data will become less accurate over time, eventually requiring you to build a new model from scratch. Real-Time AI incorporates all of the human in the loop interactions that occur

during the traditional data verification step to keep the models adaptive to the changes in the data over time.

The real-time or near real-time feedback system from humans to AI models minimize the time it takes to finish the labeling-training-inference full cycle. This further reduce the time-to-create-model and time-to-improve-model to make Singularity Systems' SinguAI platform even more practical, scalable and cost effective.

What Now?

If you are interested in learning how to implement AI solutions to solve real-world challenges in your business, then visit Singularity Systems at https://singularitysystems.com/. We exist to provide the most advanced technologies available for Intelligent Document Processing, Intelligent Image Processing, and Predictive Analytics.