White paper

Getting started with enterprise AI

Understanding AI, finding use cases, and calculating ROI

Indico Data Solutions Inc.
Indico Data c/o Venture Lane
55 Court Street Floor 2
Boston, MA 02108

www.indicodata.ai

Introduction

Motivation for this document

Since 2013 Indico Data has helped companies around the world – from two-person start-ups to the Global 100 – implement and leverage modern machine learning techniques. Over the past five years we've lived through radical ebbs and flows in public perception, watching ML go from an obscure set of academic papers to the dominant trend in today's enterprise zeitgeist. Living through this transition has been thrilling but has also presented us with a unique set of challenges. As machine learning and related technologies have moved through the hype cycle there has been an explosion in marketing material meant to supplant old academic papers and make these technologies more palatable and approachable to the average business user.

Unfortunately, as often is the case, the marketing material has generally been developed by those without a strong understanding of the underlying technology, and in many cases with strong incentives to obfuscate the truth from the public. The net result is confusion. An explosion in crude low-quality marketing material that leaves average business users with more aversion to machine learning than ever. The number of terms being used interchangeably without regard for accuracy is truly staggering. To start understanding Enterprise AI, several definitions are of particular importance:

- ***Artificial intelligence*** – Any computer program which automates a process typically assumed to require human intelligence. This may be achieved through any number of tools including, but not limited to, machine learning and deep learning.

- ***Data science*** – A generic set of skills including machine learning, deep learning, and transfer learning used to produce enterprise value from data through understanding, automation, and optimization.

- ***Machine learning*** – A field of computer science that focuses on "teaching" machines to make decisions and determinations based on data rather than relying on explicit programming.

- ***Deep learning*** – A set of machine learning algorithm that have become increasing popular in recent years due to their near-human levels of performance for tasks involved unstructured data – primarily text, image, and audio data.

- ***Transfer learning*** – A particular set of machine learning techniques aimed at solving problems more robustly with less data by "transferring" knowledge across tasks. Specifically, in deep learning transfer learning has shown particular promise in drastically reducing the required size of datasets.

This is intended for non-technical individuals within the enterprise with little to no existing machine learning expertise. By reading this whitepaper you should expect to come away with the following:

- An examination of trends that compel enterprise adoption of machine learning.
- Key building blocks for deploying Enterprise AI.
- A system for identifying and valuing new potential machine learning use cases.

Why now?

GPUs and the deep learning Renaissance

For the past 30 years, the field of Data Science has been largely dominated by classic machine learning regression-style approaches, which have been particularly successful in analyzing tabular, structured data. However, attempting to use these same techniques to analyze unstructured data has led to extremely cumbersome, brittle systems. While neural networks have existed for decades, up until approximately five years ago, network models had failed to show serious promise. Deep learning was believed to be an uninteresting and unpromising field of research, with only a small handful of universities centered in Canada and Europe continuing to research this esoteric field. The consensus today is reversed, which can generally be attributed to two things.

Over the course of the next few years deep learning transformed from an unpopular, abstruse area of research into the clear frontrunner for machine learning tasks from computer vision to natural language processing.

1. Advancement in and wide availability of powerful algorithms: Today we generally credit three people with this progress – Yoshua Bengio from uMontreal, Geoffrey Hinton from uToronto, and Yan Lecun from NYU – the "grandfathers" of deep learning.

2. Ubiquitous GPU computing. A GPU - a graphics processing unit – is a specialized piece of compute hardware built to rapidly create and display images. Originally used primarily for video game applications, during the early 21st century GPUs were progressively made to handle more general and sophisticated computation tasks, and by the early 2010s a $500 GPU was as effective at running deep learning models as the world's largest supercomputers (commanding nine-figure price tags) just a few years earlier.

These two advances came together in the early 2010s to lead the deep learning renaissance. In retrospect it became obvious that the historic failings of network models had more to do with limited compute power and unrefined algorithms than anything intrinsic to deep learning itself.

So what?

Deep learning is a significant advancement over traditional techniques. However, technical advancement alone is not sufficient to drive enterprise value. Technical breakthroughs happen every day, but very few of them deliver real value into the enterprise. The question is then exactly what about deep learning is interesting for the enterprise, and how can it effectively be used?

Near human performance for simple tasks

When thinking about Enterprise AI, a good analogy is mechanization during the industrial revolution. These machines were a far cry from the capabilities of skilled artisans. They're not capable of designing new products, and without a deep understanding of the problem they're solving and effective engineering they would be incapable of preforming even the most basic tasks. What mechanization did bring is the ability to repeatedly and reliably execute a well-defined task at scale.

So why is Deep Learning's arrival so profound? Deep Learning's application of neural networks is an attempt to model (although crudely), the way in which the human mind problem solves. Essentially neural networks allow the computer to "deconstruct" a given problem down to essential "features", and then leverage those features to solve new problem it is given. That said, Deep Learning is only effective when the goal of the task is clearly defined. Deep Learning's ability to solve "fuzzy" problems is the real breakthrough here. Understanding images and text are specific strengths of Deep Learning and represent a real breakthrough when compared to the previous "brittle" approach from the past such as expert systems, rule-based systems, and dictionary-based systems.

Unlocking the next wave of enterprise productivity

The next wave of corporate productivity and profitability will be driven by Enterprise AI.

From a business perspective these advances come at a particularly prudent time. Through the late 90s and early 2000s, the rapid uptake of modern computing and high-speed internet drove extensive improvements to productivity throughout the enterprise. Generally speaking though, as we moved into the 2010s the returns on digitization have been rapidly dwindling. In addition, the cost cutting of the last 30 years has delivered most if not all of the profitability gains that can be achieved.

Automation versus augmentation

This has very real impacts for thinking about and calculating ROI. Most enterprises don't have a good way of tracking the capacity of their workforce or thinking about the implications of adding work beyond their capacity to support it. We encourage customers to think of deep learning systems as augmentation systems. That is – these tools allow workers to do more work than before, to turn around tasks more quickly than before, and allow individuals to spend less time on rote tasks that impede their process, and more time on highly-leveraged work that directly adds value to the business.

Bringing AI into your organization

Machine learning strengths

The key difficulty in implementing any new technology lies primarily in giving business users appropriate technical context to drive reasonable high-level directives, and secondarily in giving technical implementers appropriate business context to drive implementation decisions. For cutting edge technology like machine learning this process can be particularly difficult. We often find ourselves in meetings where business and technical stakeholders are mutually unclear on what the best path forward is. Appropriate business perspective is very context-dependent, but for machine learning the technical context is somewhat universal. Essentially Machine Learning's strengths can be listed as follows:

1. ***Classification and regression***
   The first, and simplest machine learning element is classification and regression. To put it simply, these are tasks where the goal is to take input or a document and use it to predict either a specific category or a number. For instance, looking at a tweet and determining sentiment (classification) or a 1-5-star rating (regression). Classification and regression problems are algorithmically simple compared to the other categories.

2. ***Unsupervised discovery***
   Let's say for a moment that we don't have a problem that is well-scoped as a simple classification problem. You may not exactly understand what is in your data and may not have a particularly good understanding of what you want to do with it. Often people are looking to sift through enterprise systems of record (i.e. customer support logs) and discover interesting signals within that data. These use cases come with their own issues, specifically that it is impossible to determine ahead of time whether the unsupervised discovery is working effectively or not. There is a lack of quantitative rigor around these techniques because the desired output is fundamentally not known.

   As a result, most evaluation of unsupervised discovery must happen qualitatively. Do the classes that have been discovered in your data feel right or not? In some cases, the exploration is truly open-ended, you have no idea what is or is not in your dataset and so any kind of reasonable taxonomy would be considered successful. In other cases, the exploration is only partially open-ended. You might not know exactly what is in the dataset, but you have some sense and so are largely looking to either confirm or deny these suspicions. This second case is problematic and fundamentally means that pure unsupervised discovery is not a perfect fit. Instead you should think about iterating between unsupervised discovery tasks and classification and regression tasks to come to a satisfying conclusion.

3. ***Comparison***

In the first three tasks the specific applications were different, but generally speaking we were referring to an individual document. Broadly speaking, comparison is a tool for both obtaining an intuitive “distance” between documents, as well as making a classification based on multiple documents. For instance, if you were looking to compare a 10-Q document released in Q2 to one released in Q3 and then make a determination as to whether any of the changes are significant or meaningful, you would be dealing with a comparison task. Similarly, if you were looking to recommend articles to a user based on their similarity to the document they were current reading, you would be dealing with a comparison task.

From an evaluation perspective comparison tasks blend a bit of both of the first two tasks. In many comparison cases you will be comparing two documents and attempting to make a decision based on the content in both of those documents. This means that all of the evaluation metrics for classification still apply. However, if you are looking for a similarity metric without a high-degree of focus on the specific outcomes your use case starts to look more like unsupervised discovery. In the case of comparison, it is relatively easy to provide a limited number of ground truth similarities though, which makes evaluation more straightforward than in the case of unsupervised discovery.

4. ***Extraction***

Classification and regression are very helpful when every piece of content has some kind of value. In customer support for instance every ticket must be handled by someone, the only question is who. There is a similar, but separate class of problems we typically dub extraction. This addresses the question of what to do when the majority of information is not relevant. These “needle in a haystack” problems require a different approach than simple classification.

It’s related to the traditional problem of named-entity-recognition or part-of-speech tagging. While the metrics used to benchmark these solutions are less intuitive than those used for classification they are still quite rigorous, and any search or extraction problem can effectively be quantitatively validated.

5. ***Sequence generation***

In each of the preceding four tasks the output is structured. The final goal has been either a set of classes, or a concrete action. These are well-established machine learning problems; however, they are not all-encompassing. There is a fifth class called sequence generation uses a source document to create another sequence of text. Translation is a good example of a sequence generation task. Sequence generation is a fascinating field that will doubtless see massive application in the future. However, the current difficulties in implementation combined with the massive size of the required datasets make sequence generation a challenging endeavor.

The five classes above cleanly dictate exactly what capabilities machine learning can bring into new applications. Think of these as the tools in your toolbox that should be able to address any use cases that are particularly good candidates for machine learning augmentation. One particularly important point to remember is that machine learning does not solve data access. Invariably you will find use cases that are helpful for your organization but are built on top of data that you do not own. Machine learning may be a good fit, but you must address the issue of data access before tackling the actual machine learning implementation.

Assessing & implementing enterprise AI use cases

Once you've identified a use case it's time to move onto implementation. What – at a minimum – is required to implement a machine learning solution? Coarsely it can be broken down into four key components: compute, data, expertise, and ROI.

Compute

Any modern deep learning solution relies on GPUs in order to run at a reasonable rate. While it may technically be possible to run deep learning algorithms on CPU it is generally too slow and expensive to be feasible. Luckily all major cloud providers have GPU support. If your organization is comfortable using a multi-tenet SaaS platform then compute also becomes a non-issue. However, in on premise environments, which may be required when dealing with sensitive data, production-grade GPUs must be acquired as most legacy data centers still lack them.

As important as compute hardware is, compute software is even more critical. This specifically refers to deep learning algorithms coupled with the knowledge required to create, train, and deploy them. Organizations should assume that building a solution entirely from scratch is infeasible, unless they are willing and able to compete tooth-and-nail with tech giants to acquire talent. There are many vendors today that specialize in making the machine learning process more approachable. Our recommendation is that in order to implement a machine learning solution quickly, organizations should evaluate multiple vendors and choose the vendor that best fits their corporate needs and culture.

Data

As mentioned above, machine learning systems can be thought of as parrots. This means that the algorithm requires sufficient training data to operate. Importantly this isn't just any data. You cannot simply give a machine learning algorithm news articles and expect it to predict the weather. You must have a well-formed set of inputs and outputs of sufficient volume to make even the most basic machine learning problems tractable. Fundamentally, this data is the framing of the problem and its quality and volume together will determine your chances for success in applying machine learning to your business.

For example, let us say that we wish to develop a machine learning model for sentiment analysis. Stated another way you are looking to give your algorithm a piece of text – say a tweet – and have it product a score representing whether the given tweet is positive or negative. In order to create a machine learning model, you then need a "corpus" of tweets with their sentiment labeled. How many tweets do we need? It's very hard to answer this question definitively. Simpler tasks require less data, and text tasks generally require more data than image tasks, but these are only approximate rules of thumb.

If applying standard Deep Learning approaches, a standard dataset size for training would be between ten thousand and one hundred thousand labeled examples. This does not mean that simply providing on hundred thousand tweets is sufficient. Importantly they must be labeled. Additionally, you cannot generate this data by using heuristics. It must be labeled by human hands if you want your machine learning algorithm to function. This is typically where Enterprises "hit the wall" when trying to generate results from an AI approach. As defined earlier, the newer approach of transfer learning, which specializes in dramatically reducing the amount of training data required for creating machine learning models, can make Enterprise AI tractable. In some cases, these transfer learning models can operate on just a few dozen examples, but implementation of transfer learning approaches requires access to a large generic base training model.

Expertise

As one would expect, some level of data science expertise is needed to define and implement a machine learning use case. However, what is commonly missed is the importance of having the business subject matter expert intimately involved in the development of the use case. The business SME can define the task to be augmented or automated, provided valuable input in the form of training data, and help define the ROI. While implementing a deep learning algorithm today is drastically easier than the deep learning algorithms of yesteryear, it is still a developing science. You should expect that it will take multiple iterations to frame the problem appropriately and to arrive at a solution that will deliver real value to your organization. It is thus important to ensure that any relevant subject matter experts (SMEs) that are familiar with the process are available for feedback during the development process.

Outcomes, goals & ROI

The fourth and perhaps most vital component for a successful use case is an understanding and definition of the desired outcome. A clear goal statement enables the project team to work backwards in terms of identifying the steps that can be augmented, enhanced or automated, the data available, and a set of previously identifiable outcomes that can be used for training the models.

Another key factor is defining a use case is determining how accurate the process must be in order to generate a successful outcome. A common misconception is that deploying AI is a binary decision-it will either work or won't work. The reality is somewhere in the middle. Depending on the quality and amount of data, the homogeneity of the process, and other factors, Enterprise AI solutions will deliver a range of "accuracy". The cost of mitigating errors in the process need to be considered when calculating ROI.

The use case ROI is driven by an understanding of the goals of the use case. These typically take of the following forms:

1. ***Automating and accelerating existing transactional processes and workflows***
   In the simplest case, the task being automated already has a transactional nature and thus a specific dollar amount tied to it in the form of hours of labor or process "cycle time". Both of these can be translated into a set of hard costs or opportunity costs. The ROI from process automation can be calculated by a goal set at the start which might take the form of "if we can reduce the hours of labor or cycle time by 80%, we will save $X of labor costs or increase revenue opportunities by $Y. The ROI obviously is the dollars of accrued benefit divided by the cost of building and operating an AI-based process. Some key factors that play in here are the robustness of the AI-based solution. It is rare that any AI-based approach will be able to completely automate an existing process, but rather is likely to augment a large portion of it. The cost of the final human review of the process needs to be factored in to the ROI analysis.

2. ***Increasing capacity for overburdened processes***
   A good example is customer service. There is often a significant investment in the triage of inbound customer service requests and other forms of inbound communications. Enterprise AI is particularly good at classifying and "routing" this kind of content. The ROI here in addition to expanding the capacity of the current process (one agent can now handle X% more volume), another aspect is the impact on customer satisfaction from having requests handled more rapidly and more accurately.

3. ***Enhancing existing products/creation of new products***
   Another common application of Enterprise AI is the enhancement or creation of the enterprise's products and services. This is often the most challenging implementation to perform an ROI calculation for, because at a fundamental level this is a speculative product improvement that was previously infeasible in the past due to its large labor cost and therefore lacks a good comparison. Another approach to justify the ROI in cases where net new functionality is being offered is to calculate what the price of such a process would be if it were implemented completely manually. While it's disingenuous to claim any improvement here as true ROI, it can be very helpful in determining the amount of leverage your firm is deriving from machine learning in this context. A product offering that would take $10m annually to replicate is likely to have more intrinsic value than one that would take only $1m annually to replicate.

Conclusion

In conclusion, the last five years have witnessed an amazing acceleration in the availability of these powerful new technologies. With a framework for evaluating these opportunities, enterprise today can generate ROI in a tangible and practical way.

About Indico Data

The future competitive landscape is likely to be driven by those who can most effectively adopt and deploy the benefits of Enterprise AI.

Indico Data transforms unstructured data into actionable insights. With the Indico Unstructured Data PlatformTM, enterprises ofall sizes can automate, analyze, and apply unstructured data -- documents, emails, images, videos and more -- to a wide range of enterprise workflows. This enables them to gain rich insight and maximize the value of their existing software investments, including RPA, CRM, ERP, BI, by enabling these systems to work with unstructured data. Visit **www.indicodata.ai** to learn more.

Contact Us

Indico Data Solutions, Inc.
Indico Data c/o Venture Lane
55 Court Street Floor 2
Boston, MA 02108

Email: contact@indicodata.ai
Web: www.indicodata.ai
Twitter: @indicodata