**The Digital Transformation**

**What you need to know.**

Dec. 2017  
Andrew Ferlitsch

Your company has been exploring bringing in artificial intelligence (AI) and machine learning (ML) into its services and products it provides, streamlining its process and reducing human knowledge workers with intelligent ‘learning’ agents (IA).

There are some key questions you will need to explore:

1. What technologies are available?
2. What can we do?
3. Where do we start?
4. What is the roadmap?

***Change is Awaiting***

Unlike the traditional Software Development Lifecycle (SDLC), the development and deployment of ML/IA that is emerging is substantially different. These changes to the development and deployment process will move through the entire industry. It will be comparable, if not greater, than when the industry moved from Waterfall Development to Agile Process. When the transition to the Agile Process occurred, many companies and their software developers were well entrenched in the IEEE standards for the Waterfall development process and could not make the transition. Those companies and software developers became legacy companies and developers overnight.

I foresee the same happening in the transition from the current SDLC process over the next two to three years. In other words, those companies and software developers who do not transition will become legacy companies and software developers overnight.

When there is a new emergent technology at this scale, it is common place for there to be an initial period of inertia. In each industry segment, one company will break out utilizing these new emergent technologies and development/deployment processes. The other companies will either afterwards have to adapt or die.

It is better to be the leader in your industry, than facing the consequence of ‘Can we adapt? Will we perish?’

***The Change***

In the current SDLC methodology, companies develop a solution. The solution consists of algorithms that are predefined. The same input essentially produces the same output. When a new problem is encountered or a new feature is needed, a new solution has to be developed and deployed to the customer base.

Over the years the industry has improved on this through Business Intelligence (BI). Through BI, one collects the domain knowledge of an expert and codifies it into a set of rules. When new problems are encountered, one goes back to the domain expert to develop a new set of rules, and the new rule base is deployed to the customer base, without the need to update the software.

But the problem remains that after the update, the same inputs produce the same new output. Your company continues to be dependent on the long-term expertise of the domain experts you employ, the human labor in codifying that knowledge and the lengthy deployment process.

ML/IA changes that. Software Development Engineers (SDE) no longer will be developing a solution, but developing search and learning algorithms. We see this already happening with technologies’ big five, such as Google’s search, Apple’s Siri and Amazon’ Alexa. While a human expert can learn rules of thumb with much smaller amount of data over years, machine learning with vast amounts of data can learn far more accurate models in seconds.

What changed in AI? Why is there a breakout today? Up to a few years ago, AI tackled problems by injecting semantic knowledge into algorithms using domain experts. These rule-based and business intelligence algorithms made slow progress in advancements over the years. With the advent of commodity memory and computing and big data sets, the industry shifted to algorithms based on statistical distributions. Overnight, once the computing power and data was there, these algorithms outperformed the legacy methods out-of-the-gate and continue to advance. That’s what happened. The science shifted from coding human semantics to learning semantics through large scale statistical distributions.

Combined with Intelligent Agent (IA) and performance metric, these systems can continuously learn and self-improve once deployed to the customer base. This will cause a change from companies deploying updates to their solutions, to deploying updates to their search and learning algorithms. Once underway, technology advancements will move at a pace not seen in the industry.

***Skills***

The roles and titles of software developers will change to titles such as data scientist, data engineer, machine learning engineer, machine learning analyst, and data ops. The data scientists, typically seen as having PhDs in computer science or computational theory, are being employed to develop new machine learning algorithms and techniques. Currently, there is a very strong competition for these individuals from the deep pocket companies, deploying a “first to market strategy”. These individuals will have skills in graph theory, computational theory, linear algebra, neural networks, and computational algorithms like Bayes Theorem, Markov principles, Bellman equations, and Kalman filters. As advancements are published each month, they will be on top of understanding the computation and theory behind them. As of this writing, we’ve seen new advancements being published in stacking, adversarial attacks and capsules.

Next on the tier are the machine learning engineers (MLE). These developers will specialize in applying machine learning, utilizing the latest frameworks, such as Google’s TensorFlow, or machine learning services, such as from Amazon and Microsoft. While they may not need to have deep knowledge of computational theory and algorithms, they still need a good breadth understanding, particularly in linear algebra and conditional probabilities to effective do hyper-parameter tuning of models. This task is still considered an art form vs. a science. Effective decision making and testing of hyper-parameters will impact the performance, resources and accuracy of the models the MLE will train.

For the MLE working in the autonomous world or robotics, they will be developing programs that will learn and adapt in a dynamic real world environment. They will need additional skills in graph theory, reinforcement learning and dynamic programming, as well as sensor fusion. As software only applications become more and more real-world dynamics (e.g., financial trading), these MLE engineers will also need to develop the same skills.

**Data Scientist**

S

S

**Role  
Hierarchy**

MLE

**Pay Scale**

**Data  
Ops**

**Data Analyst**

**Data Engineer**  
E

Next on the tier are data engineers, data analysts, and data ops (operations). Data engineers will perform work understanding and preparing data for machine learning. They will work with both structured and unstructured data. They will need skills in data formats, like CSV and JSON, data ontologies, like RDF and OWL, data encodings, tools for handling large data sets, like Hadoop, Spark and Pig, and databases for unstructured data, like MongoDb. Data engineers will need to have skills in the data preparation, which include data cleaning (i.e., data wrangling), missing values, categorical conversion, and feature engineering. For these tasks, the will need skills with machine learning libraries that perform these tasks, such as numPy, pandas, Sci-Learn and Torch libraries. With the ever increasing amount of human text data, some will need additional skills in toolkits for natural language processing, such as NLTK, for preprocessing narratives in datasets into numeric values which are required for machine learning.

The data analyst will perform roles similar to those of the traditional business analyst, but will work with new data analysis and visualization tools that are being released, such as Tableau. While these tools may handle much of the math under the cover, it would be helpful for them to have basic understanding of linear algebra and linear and logistic regressions. Knowledge of clustering, such as K-means, and feature co-variance, such as PCA, would also be helpful. They will continue to need to have skills in plotting and visualizing both data and results from trained models.

The data ops engineer is comparable to the dev (development) ops engineer. These engineers will be responsible for deployments into the cloud of distributed systems which handle vast amounts of data. Their focus will be on how data is distributed, managed and flows through the system. They will need skills in the partitioning, storage and retrieval of data as efficiently and reliably as possible. They will also manage the distribution and updates of deployed trained models. They will need skills in big data systems, such as Hadoop and Spark, managing/configuring unstructured databases, such as MongoDb, and performing in-production monitoring of performance and accuracy of trained models.