FOUR IT/IS PILLARS FOR ARTIFICIAL INTELLIGENCE MACHINE LEARNING/DEEP LEARNING APPLICATIONS

Robert E. Samuel, Robert.Samuel@ieee.org
Gerald Cormier, gcormier@charter.net
Shannon Fascendini, Shannon.Fascendini@gmail.com
Christina M. Stubanas, tinastubanas@comcast.net
Katherine A. Yacko, kyacko1@gmail.com

ABSTRACT

Artificial Intelligence Machine Learning/Deep Learning (AI ML/DL) technology adoption for business applications is impacting many Information Technology/Information Systems (IT/IS) roles. Bridging the divide between data, insight, and action requires revisiting the development operating models. The evolution of four key IT/IS pillars will be necessary to successfully implement today's AI ML/DL business solutions. Based on qualitative research within a Fortune 50 U.S.-based healthcare company, this paper assesses the IT/IS pillars for role definition, required technical skills, and behavioral competencies. The evolution of these pillars could influence how individuals learn and prepare for the roles as AI applications gain business adoption and acceptance.

Keywords: Artificial Intelligence, Machine Learning, Deep Learning, IT/IS Roles

INTRODUCTION

McKinsey Global Institute brought national spotlight to the term "big data" and the exponential growth of data in major industries (Manyika, et. al., 2011). Autonomous vehicles, advanced healthcare research and pervasive virtual assistants, social media analytics, and Internet of Things (IoT) are a few use cases where data science and machine learning are being rapidly adopted. The adoption of artificial intelligence machine learning/deep learning (AI ML/DL) applications in corporations is forcing information technology/information systems (IT/IS) practitioners to reassess the pillars necessary to successfully implement business solutions (Banavar and Cooper, 2016) (Samuel, et. al., 2017). The IT industry has experienced computing era shifts beginning with the first era of tabulating machines to the second era of programmable computers to today's era of AI ML/DL. Tabulating machines were heavily focused on the application of mathematics. Programmable computers were coded with "if, then, else" instructions to produce a deterministic output. Today's AI ML/DL applications are about the probabilistic outcomes based on context (Banavar and Cooper, 2016).

A literature review indicates that the industry remains fragmented on the exact classification of AI (Harper, 2017) (Chen, 2017) (Skansi, 2018). Skansi (2018) further elaborates that there are two major industry societies that provide a formal AI classification used to classify research papers: the American Mathematical Society (AMS) and the Association for Computing Machinery (ACM). The AMS maintains the Mathematics Subject Classification 2010 which divides AI into the following subfields:
- General,
- Learning and adaptive systems,
- Pattern recognition and speech recognition,
- Theorem proving,
- Problem solving,
- Logic in artificial intelligence,
- Knowledge representation,
- Languages and software systems,
- Reasoning under uncertainty,
- Robotics,

- Agent technology,
- Machine vision and scene understanding, and
- Natural language processing.

The ACM classification for AI provides their subclasses as well (Note that ACM identifies machine learning is a parallel category to AI, not subordinated to it). The subclasses are:
- Natural language processing,
- Knowledge representation and reasoning,
- Planning and scheduling,
- Search methodologies,
- Control methods,
- Philosophical/theoretical foundations of AI,
- Distributed artificial intelligence, and
- Computer vision.

Skansi (2018) concluded from these two classifications that there are a few broad fields of AI that can summarized as:
- Knowledge representation and reasoning,
- Natural language processing,
- Machine Learning,
- Planning,
- Multi-agent systems,
- Computer vision,
- Robotics,
- Philosophical aspects.

This research used the Skansi broad fields of AI as a basis for discussion and interviews. These fields provide the background to assess the IT/IS pillars for AI ML/DL business application development. The identified IT/IS pillars are:
- Data Architecture - the primary role to gather high level business needs and requirements (often in partnership with a data analyst) and design the solution. Typically, the architect is a generalist but often has depth of knowledge in particular technology domain (Forrester, 2017).
- Data Engineering - the primary role to deploy a runtime implementation of the application and monitor application performance, reliability, and stability. The engineer is often a specialist in a particular operations-based technology stack (Forrester, 2017)
- Data Science – the primary role is to follow "a set of fundamental principles that support and guide the principled extraction of information and knowledge from data" (Provost and Fawcett, 2013, p. 52).
- Data Analysis – the primary role is judge the value of generated insights and to ensure applications address important business problems (Provost and Fawcett, 2013). Data Analysts do basic descriptive statistics, data visualization, data source assessments, and communicate data points for conclusions.

The potential of AI ML/DL depends on the availability of talent and technology to harness its value (Boisvert, et. al., 2017). As business AI applications are increasing in frequency and complexity, IT/IS skills are evolving. Forrester Research (Goetz, 2017) discovered that enterprises want to get value out of their data faster and at scale, however technology alone can't solve this challenge and adding roles with the sole responsibility of activating data is necessary. Davenport and Patil's Harvard Business Review article (2012) identified data scientists as an important and emerging role that has little consensus on where the role fits in an organization, the value to the organization, or the process of discovery. As a result, it was necessary for data scientist to craft their own tools and research approaches. William Chen (2017), a Data Scientist at Quora, summarizes the five skills and competencies that he sees is important for a data scientist role. This includes 1) programming (augments, large datasets, create tools); 2) quantitative analysis (experimental design and analysis, modeling of complex economic or growth systems, machine learning); 3) product intuition (generating hypothesis, defining metrics, debugging analysis); 4) communication (communicating insights,

data visualization and presentation, general communication); and 5) teamwork (being selfless, constant iteration, sharing knowledge). The authors notice that three of the five competencies are not technical skills.

In recent years, the emphasis has been on establishing and cultivating the data scientist role within corporations. The focus on the data scientist role has resulted in a bottleneck when there are inadequate resources to efficiently build the AI applications. Forrester (2017) discovered the need to grow the role of data engineer equally to the data scientist role to successfully deploy AI ML/DL applications by stating:

> *"As organizations began developing data scientists, they expected them to carry the load of developing data lakes and data pipelines as well as creating sophisticated analytics. But individuals with computer science and statistical skills are rare. At one aerospace manufacturer, data engineers take on the tasks to source, wrangle, and democratize data in data lakes, allowing data scientists to focus on insight creation."*

They continue by stating that data development roles are highly outsourced and transient. This transient nature of these engineer and scientist roles means that knowledge from development projects are lost when resources are reassigned. Data architects are often necessary for design work for platforms and to provide data access to ad hoc requests. Therefore, as data engineering evolves, data architecture is becoming a peer and guides the investments required for data sources and services while providing the platform, frameworks, and reference architectures.

The role of a data analyst continues to evolve. Nicholas Chamandy (2018), the Scientific Director at Lyft, recent published an article stating that have "maintained a fairly strong semantic distinction between the two roles: analysts extract insights from data, track the health of our business and drive better decision-making; scientists build the mathematical models and algorithms that power the core components of our product." This further highlights the industry challenges at defining the pillars and roles of AI ML/DL.

RESEARCH METHODOLOGY

There is a limited volume of academic peer-reviewed literature that addresses how AI ML/DL is impacting IT/IS pillars. The authors found the clear majority of material published is in trade journals. Even in the trade journals, the topic is addressed with significant variance. While the field of machine learning, deep learning and artificial intelligence spans three decades, the academic research regarding IT/IS pillars and industry roles is rather sparse.

Using the four IT/IS pillars as a framework, the researcher centered upon the following research questions:

> *R_1: How has the IT/IS pillars definition evolved with respect to the development of ML/DL business applications?*

> *R_2: Which skills align to the IT/IS pillars for the development of ML/DL business applications?*

> *R_3: Which competencies align to the IT/IS pillars for the development of ML/DL business applications?*

To address these research questions, the authors used a qualitative opinion-based research methodology approach using informal individual interviews of subject matter expert practitioners and managers. A review of peer-reviewed scholarly publications was used to determine the boundaries (skills, competencies, technology maturity) of the problem space and IT/IS pillar framework. The interview questions derived from articles that are within the computer science, software and systems engineering, information science, and artificial intelligence domains. The interview questions were assembled based on key domain topics and issues observed during the review of existing peer-review scholarly publications. The questions were structured to inquire on the participant's understanding of how the pillars and roles have evolved over the past five years. The interviews involved a convenience sample size of nine industry practitioners at a U.S. based Fortune 50 company. The industry experts spanned the four IT/IS pillars of data architecture, data engineering, data science and data analysis. The interviews were conducted during the March and

April 2018 timeline via phone, in-person, and email communication. Interview data was normalized and grouped into categories for commonality and role association.

RESULTS

The first general observation is that there are a variety of roles in each pillar that essentially have the same, or similar, definition, technical skills and competencies. Depending on the industry, these roles could go by a different title (e.g. Information Architect, Machine Learning Engineer, Data Specialist, Database Analyst, Big Data Analyst). Regardless of the title, it's important to understand the key distinctions between each role and how they fit together across the pillars to make better data-driven business applications. The data volume, veracity, and variability are increasing and becoming more prominent within the systems in which users interact every day.

Analysis of the interviews resulted in the following observations across the four pillars:

- Data Architecture – Traditionally data architects are not included in the early phase of development AI ML/DL business applications. Industry changes are requiring data architects to have a broad understanding of AI ML/DL technologies and assist in providing the "holistic" view of the problem domain. This role requires increased competencies for teamwork and facilitation along with technical skills.

- Data Engineering – This role is becoming more important and more inclusive of skills previously held by the data scientist. The data engineer has the primary role of data preparation and manipulation, from ingestion to formatting/transformation to storage for the consumption by the data scientist. Of the four IT/IS pillars, the data engineer has had the most impact on amount of technical skills required to be successful.

- Data Science – With the evolution of the data engineer, the data scientist has been afforded more time to focus on discovering insights. There is less focus on the tools needed for the data preparation. The technical skills/competencies have not been significantly impacted, but their productivity has shifted given the less time needed for data preparation.

- Data Analysis – This role was historically aligned to corporate business intelligence reporting has expanded to include business acumen and data semantic understanding. This role has increased emphasis on the visualization of the data for business understanding. The data analyst role will need to expand into new ways of delivering insights that go beyond the current deliverables of reports, dashboards, and messaging alerts.

Overall, many technical skills and competencies documented in literature aligns to the research findings. This research discovered that due to the sacristy and higher salaries of experienced data scientists, the role of data engineer is increasing throughout the industry to help augment, and sometimes replace, the data preparation tasks for the data scientists. This finding aligns with the research performed by Forrester (2017).

Interview data collected associated with the research question R_1 identified that the pillar definitions have evolved as shown in Table 1. With respect to research question R_2, the technical skills for each pillar is shown in Table 2. Table 3 highlights the pillar competencies that addresses research question R_3.

Table 1. IT/IS Pillars Revised Role Definitions

Data Architecture	Data Engineering	Data Science	Data Analysis
Role:	Role:	Role:	Role:
A data architect defines how the data will be stored, consumed, integrated, and managed by different data entities and IT systems with an organization's data architecture.	A data engineer identifies and ingests data, builds reliable data pipelines, prepare features for a predictive model, productionalizes predictive models and transforms the data into formats that data scientists can consume.	A data scientist combines data inferences and algorithmm development to discover insights from massive amounts of structured and unstructured data. These discovered insights help enable companies to make smarter business decisions.	A data analyst retrieves, organizes and performs quantitative analysis on data to generate reports that will identify trends that the business can leverage.

Table 2. Technical Skills for IT/IS Pillars

Data Architecture	Data Engineering	Data Science	Data Analysis
Data Modeling tools	Data modeling and mining	Programming Tools (Python, R, SAS, C/C++)	Programming Tools (SAS and/or R)
Hadoop Technologies	Hadoop Technologies	Unstructured Data	Data Visualization and communication
RDBMS (DB2)	Big data tools (Spark, H2O)	Concepts of Machine Learning	Statistics
Big Data	Software Engineering	Predicitve Modeling	Data Modeling
Data Lakes	Data Warehousing solutions	Data Visualization and Reporting Techniques	SQL
	Database architectures	Hadoop Technologies	
	Scripting Tools (Python and Shell)	Cloud tools (Amazon S3)	
	SQL	SQL	
	Linux Fundamentals		
	Network Basics (IP address, ports, DNS)		

Table 3. Competency Skills for IT/IS Pillars

Data Architecture	Data Engineering	Data Science	Data Analysis
Holistic View	Data Intuition	Insight (Behavioral, Medical, Financial, Marketing, Security)	Quantitative Analysis
Communication Skills	Industry Knowledge	Quantitative Analysis	Filter to useful data
Problem-Solving Skills	Data industry standards and best practices	Business Acumen	Data Intuition
Attention to detail	Communication Skills	Math/Statistics skills (linear algebra, algorithms)	Problem-Solving Skills
	Problem-Solving Skills	Intellectual Curiosity	Intellectual Curiosity
	Persistence	Communication Skills	
	Innovative Approaches	Problem-Solving skills	
		Investigative skills	
		Domain Knowledge	

Cao (2016) identified the revolution of data science and analytics by the three key indicators of 1) a disciplinary paradigm shift; 2) a technological transformation; and 3) innovative data products. The authors agree with this observation and extends it to include the evolution of the IT/IS pillars. Cao and this research align with the findings that most data scientists simply conduct normal data engineering and descriptive analytics. An organization requires different roles, skills and competencies according to the maturity level of business applications through the effective use the IT/IS pillars.

Additional suggested research could be comparing these findings to the current career paths and available resources for individual become educated. Sagheb-Tehrani (2015) concluded that "the college curriculum in information systems (IS) is revisited and often changed in institutions for higher education to reflect the changes in the field. It is important to make necessary changes to the IS curriculum to make programs challenging and to better prepare graduates for today's job market." The authors believe that it is possible to capture the evolution of the IT/IS pillars for AI ML/DL in the redesign of college curriculums. The availability of undergraduate curriculums for the data architect role is particularly sparse (Aasheim, et. al., 2015). Other than undergraduate curriculums, the IT/IS industry has an opportunity to increase understanding through a more formalized association to the pillars and roles.

SUMMARY

This research indicates that AI ML/DL technology is impacting many IT/IS roles when developing business solutions. The initial heavy emphasis on the data scientist role is resulting in a resource constraint that can be augmented by the data engineer, data architect and data analyst. The evolution of four key IT/IS pillars (for role definition, required technical skills, and competencies) will be necessary to successfully implement today's AI ML/DL business solutions. The evolution of these pillars could influence curriculum changes for university IT/IS programs as AI applications gain business adoption and acceptance.

The four IT/IS pillars will need to expand and evolve their technical skills and competencies to include Internet of Technology (IoT) data and event driven technologies. IoT will impact organizations with new data sources, tools, and architectures for analysis and insights. Event driven concepts and technologies will help manage and throttle multiple

data streams to deliver additional insights. The authors recommend further research in respect to IoT and event driven technology as the relate to AI ML/DL technologies.

REFERENCES

Aaschiem, C. L., Williams, C., Rutner, P., & Gardiner, A. (2015). Data Analytics vs. Data Science: A Study of Similarities and Differences in Undergraduate Programs Based on Course Descriptions. *Journal of Information Systems Education, 26*(2), 103- 115.

Banavar G., & Cooper M. (2016). Turing Lecture 2017 Cognitive Computing. *ITNOW* 2016; 58 (4): 62-63. doi: 10.1093/itnow/bww117

Boisvert, D., Topi, H., Harris, M. D., & Yohannan, K. (2017). Exploring the Landscape of Data Science. SIGITE'17, October 4-7, 2017. Rochester, NY.

Cao, L. (2016). Data Science: Nature and Pitfalls. IEEE Intelligent Systems. September/October 2016. Pg. 66- 75.
Davenport, T. and Kirby, J. (2016). Only Humans Need Apply. HarperCollins Publishers. New York, NY.

Chamandy, N. (2018). What's in a name? The semantics of Science at Lyft. https://eng.lyft.com/whats-in-a-name-ce42f419d16c Retrieved: 18 April 2018.

Chen, W. (2017) What Are the Top Five Skills Data Scientists Need? https://www.forbes.com/sites/quora/2017/06/15/what-are-the-top-five-skills-data-scientists-need/#19f0efdd7c0c

Davenport, T. H. & Patil, D.J. (2012). Data Scientist: The Sexiest Job of the 21st Century. Harvard Business Review. October 2012.

Goetz, M. (2017). Data Engineers Have Become More Important Than Data Scientists. Forrester Research. November 9, 2017.

Harper, J. (2017). The Future is Now: Cognitive Computing Throughout the Enterprise Today. KM World. April 29, 2017. http://www.kmworld.com/Articles/Editorial/Features/The-future-is-now-cognitive-computing-throughout-the-enterprise-today-117755.aspx

Lane, D. (2015). The Skills Implications of Cognitive Computing. April 29, 2015. http://dalelane.co.uk/blog/?p=3299

Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., & Hung Byers A. (2011). Big data: The next frontier for innovation, competition, and productivity. https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Provost, F. & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data, 1*(1), 51-59.

Samuel, R., Stubanas, C., Yacko, K. (2017). Unique Skills and Competencies for Cognitive Computing Implementations. Proceedings of the International Association for Computer Information Systems 57th International Conference, Philadelphia, Pennsylvania, p. 33.

Sagheb-Tehrani, M. (2015). Towards a Consistency of Information Systems Curriculum. *Issues in Information Systems. 16(*11), 28-37.