The Road to Enterprise Artificial Intelligence: A Case Studies Driven Exploration

by

Saleha Saulat Siddique

MBA (2009)

University of California, Los Angeles

Submitted to the System Design and Management Program
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering and Management

at the

Massachusetts Institute of Technology

June 2018

Signature of Author____
System Design and Management Program
May 25, 2018

Certified by____
Jeanne Ross
Thesis Supervisor
Center for Information Systems Research

Accepted by____
Joan Rubin
Executive Director, System Design & Management Program

THIS PAGE INTENTIONALLY LEFT BLANK

The Road to Enterprise Artificial Intelligence: A Case Studies Driven Exploration

by

Saleha Saulat Siddique

Submitted to the MIT System Design and Management Program in Partial Fulfillment of the Requirements for the Degree of Master of Science in Engineering and Management

ABSTRACT

Increases in the volume of data and the availability of compute power have driven a number of advancements in the field of Artificial Intelligence (AI), and AI technologies and applications are getting a flood of publicity in the media. While four in five executives agree that AI is a strategic opportunity for their organization, only about one in five has incorporated AI in some offerings or processes, and only one in 20 has extensively incorporated AI in their offerings or processes. There is a gap between expectation and action, and we are still in the early days of enterprise AI adoption. This thesis explores the path enterprises need to take to close this gap and to build an enterprise AI capability, thereby realizing the full value of this disruptive technology. Through a literature review it proposes a seven component holistic framework that can guide enterprises through this journey. The framework is more 'wide than deep', and it is supplemented with five case studies that take deep dives into the real life journeys of enterprises from different industries. These stories provide a vivid illustration of best practices and challenges. The case studies cover Danske Bank fighting financial fraud with deep learning, Deutsche Telekom improving customer service with an intelligent digital assistant, General Electric deploying machine learning applications for monitoring workflows in the Industrial Internet of Things, General Mills automating insights for marketers, and Kaiser Permanente using state of the art Natural Language Processing techniques on unstructured triage notes to improve patient flow forecasting. Learnings from the case studies are synthesized into recommendations to aid practitioners on the road to enterprise Artificial Intelligence.

Thesis Supervisor: Jeanne Ross
Title: Principal Research Scientist, Center for Information Systems Research

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgements

This body of work, and my graduate program journey at large, has benefited from the support and encouragement of a number of people and institutions. I am deeply indebted to them all. I would like to particularly mention:

- **Jeanne Ross,** my advisor. Thank you, Jeanne, for the prompt reviews, the constructive criticism, the valiant efforts to save me from 'boiling the ocean', and the patience and encouragement while I battled my 'perfect is the enemy of good' demons. Thank you also for arranging funding for me to attend an Artificial Intelligence conference in New York, which was a fantastic learning experience. It was truly a joy working with you.

- **My parents,** who would be hard pushed to explain what I do but support me anyway. I am deeply grateful for their unflinching support and unconditional love.

- **O'Reilly Media,** who routinely arrange a number of great conferences in the area of Artificial Intelligence, and furthermore, make a treasure trove of excellently curated and practitioner oriented information available at student-budget-friendly prices.

- **The SDM program and community,** who enriched my journey at MIT in countless ways.

- **MIT,** it really is drinking from the fire hose here. Never in my life have I experienced so much intellectual stimulation, and being spoiled for choice on what conference to attend, course to take, talk to attend etc., and I certainly take away from here a commitment to lifelong learning and challenging myself.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

List of Tables

1 Introduction

“AI is the new electricity” – Andrew Ng, founder of Google DeepBrain project

“Machine learning and AI is a horizontal enabling layer. It will empower and improve every business, every government organization, every philanthropy — basically there’s no institution in the world that cannot be improved with machine learning.” – Jeff Bezos, Amazon

Self-driving cars, computer programs beating human champions at Go, digital assistants making phone calls to book haircut appointments and sounding indistinguishable from humans: Artificial Intelligence (AI) has been capturing the public imagination and dominating the headlines of late. The case for enterprise AI is no different - see a sample selection of headlines below.

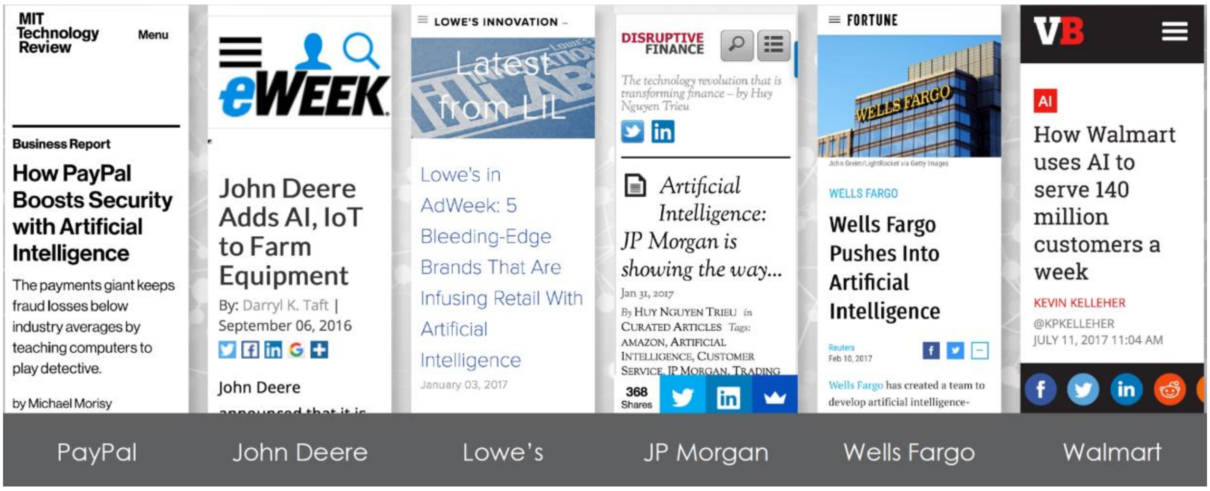


Figure 1.1 – AI is Happening in the Enterprise[1]

The potential for this disruptive technology is great, but despite the flood of publicity we are still in the early days of enterprise AI adoption. This thesis examines available literature and studies the road to building an enterprise AI capability through a case studies driven approach.

Section 1.1 articulates the motivations that led to this thesis. Section 1.2 further details its aim and objectives. Section 1.3 presents the research question this thesis explores. The section concludes by presenting a complete outline of the thesis in Section 1.4.

1.1 Motivations

Increases in the volume of data and availability of compute power have driven a number of advancements in the field of Artificial Intelligence. In narrow domains such as recognizing objects from images[2], the best AI systems can now exceed human performance. Artificial Intelligence technologies like deep learning, machine learning, and Natural Language Processing are at the peak of inflated expectations according to Gartner’s 2017 Hype Cycle for AI:

[1] (Bodkin, 2017)
[2] (Chui, Lund, Madgavkar, Mischke, & Ramaswamy, 2018)

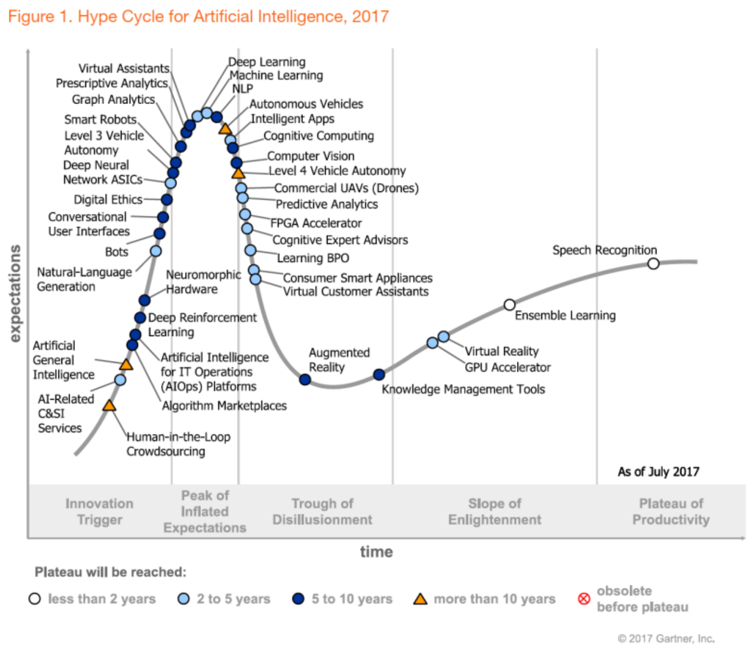


Figure 1.2 – Gartner Hype Cycle for Artificial Intelligence, 2017[3]

What does this mean for businesses? According to a BCG-Sloan Management Review survey of 3000 executives in 112 countries and 21 industries, "Expectations for AI run high across industries, company sizes, and geography…but although four in five executives agree that AI is a strategic opportunity for their organization, only about one in five has incorporated AI in *some* offerings or processes. Only one in 20 has *extensively* incorporated AI in their offerings or processes."[4] This disparity between expectation and action is echoed by other surveys. In Gartner's CIO survey from 2017, only "one in 25 CIOs described themselves as having artificial intelligence in action in their organizations.", though "Six in 25 are either piloting or have AI in their short-term plans. Five in 25 have it in middle-term plans."[5]

We are in the early days of enterprise AI adoption but it is expected to ramp up. As Gartner puts it "The risk of failing is great, but the risk of enterprise obsolescence or non-competitiveness in the digital business era is even greater. While the potential benefits are great, they will come with inherent failures, setbacks, and the 'disillusionment' typical of emerging technologies."[6]

The motivation behind this thesis is to help enterprises close the gap between expectation and action, and to illuminate the road to building an enterprise AI capability. It provides best practices and guidelines, and highlights challenges along the way, so enterprises can minimize the risks and setbacks in their AI journeys.

[3] (Brant & Austin, 2017)
[4] (Ransbotham, Kiron, Gerbert, & Reeves, 2017)
[5] (Harvey, 2018)
[6] (Brant & Austin, 2017)

1.2 Aims and Objectives

> "Over the next decade, AI won't replace managers, but managers who use AI will replace those who don't." – Eric Brynjolfsson and Andrew McAfee, Harvard Business Review
>
> "Management is demanding a 'Cognitive Computing' strategy. Engineering wants to get their hands on Machine Learning. Marketing wants to include 'AI' in product descriptions. Product is afraid of falling behind the competition. Everyone is getting calls from vendors" – Kristian Hammond, Narrative Science

This thesis aims to provide a practitioner oriented perspective on building an enterprise AI capability, with its audience being managers who are facing internal and external pressures to adopt AI. As such it does not dive deep into the mathematical underpinnings of the technologies, maintaining instead a functional perspective.

The objectives are to provide an understanding of the opportunities, challenges, best practices and common pitfalls in the road to enterprise AI, through a literature review and case studies analysis. The literature review is used to formulate a more 'wide than deep' framework which is complemented with case studies that are more 'deep than wide'. Together they illuminate the path to bringing AI into the enterprise and managing it as a strategic capability.

1.3 Research Question

To achieve the above aims and objectives, the key research question was designed using a To-By-Using framework that articulates the System Problem Statement[7] as:

System Problem Statement

To: Aid managers in building an enterprise AI capability
By: Highlighting opportunities, challenges, best practices and common pitfalls
Using: Literature review, case studies analysis

1.4 Outline of Thesis

This thesis consists of six chapters that describe the motivation, research methodology, literature review, framework for building an enterprise AI capability, case studies, and recommendations and future areas of work. This Section briefly describes the content of the remaining chapters.

Chapter 2

Outlines research methodology used in this thesis and explains the rationale for the approach.

Chapter 3

Reviews existing literature from journals, magazines and the internet on opportunities of AI and the challenges of AI adoption, which then motivates the subsequent parts of the thesis.

[7] (Crawley, Cameron, & Selva, 2015)

Chapter 4

Proposes a framework for building an enterprise AI capability so organizations have a holistic reference for the types of issues they need to be considering.

Chapter 5

Provides five in depth cross industry case studies – Danske Bank, Deutsche Telekom, General Electric, General Mills, and Kaiser Permanente.

Chapter 6

Synthesizes the learnings from the case studies into recommendations and discusses future areas of work.

2 Research Methodology

This chapter describes the research methodology used in this thesis. Section 2.1 discusses the research approach, Section 2.2 discusses the research design and methods, and Section 2.3 highlights the limitations of this methodology.

2.1 Research Approach

This thesis primarily uses an inductive, case studies based approach under an interpretative paradigm.

An inductive approach, where one reasons from specific instances to arrive at a general conclusion, is suited to the topic as no over-arching scientific or management theory of building an enterprise Artificial Intelligence capability exists. This is still an evolving, applied field rather than one with a theoretical basis. In the absence of sound theoretical underpinnings which would lead to stating hypotheses and then proving/disproving those (a deductive approach), this thesis opts instead to take the reader on a journey of discovering insights and then synthesizes that into overall recommendations (an inductive approach)[8].

A case studies based approach allows for in-depth, holistic analysis of a complex situation in a specific real-life context. This approach is used in a number of disciplines. It is particularly popular at leading business schools to impart management education. Given this thesis is aimed at managers who wish to build an enterprise AI capability, the case studies based approach is thus quite appropriate. The selected case studies develop and highlight a range of issues and how they were handled, illustrating common pitfalls and best practices. A common misunderstanding about case studies is that general theoretical (context-independent) knowledge is more valuable than concrete (context-dependent) case knowledge[9], but Flyvbjerg challenges that: "Concrete case knowledge is more valuable for social sciences than the vain search for predictive theories and universals."[10] Furthermore, even where one has context-independent predictive management theories, they can be quite superficial/trite, hindering their understanding, absorption and retention. Case studies that e.g. develop the same lesson through a richer, more substantive treatment will potentially fare better on learning outcomes.

Data analysis was conducted through an interpretative research paradigm (which posits that reality is a subjective construct, so "interprets" the reality through a "sense-making" process) rather than applying the positivist paradigm that presumes reality is relatively independent of context and can be studied using objective techniques like standardized measures.[11]

2.2 Research Design and Methods

Secondary data collection was used – i.e. case studies were mined from practitioners' stories recounted at various conferences. Primary data collection was not opted for, given a) this is a new area and only a limited number of enterprises are employing a best practice approach, so finding sources to interview

[8] (Fudge, 2015)
[9] (Flyvbjerg, 2011)
[10] (Starman, 2013)
[11] (Bhattacherjee, 2012)

in the timeframe of the thesis (one semester) was tough, b) challenges of proprietary data exist, and most of all c) even on finding practitioners who are doing it well, and who are ready to share their stories, the stories shared tend to be more superficial and less substantive than when a practitioner has actually prepared for a conference talk with an in-depth deck.

See the below table for the list of conferences consulted (top bolded one attended in person, for the rest video compilations were reviewed).

Conference Title	When	Where
O'Reilly AI Conference	**May, 2018**	**New York**
Strata Data Conference	March, 2018	San Jose
Strata Data Conference	December, 2017	Singapore
Re-work Deep Learning Summit	October, 2017	Montreal
Strata Data Conference	September, 2017	New York
O'Reilly AI Conference	September, 2017	San Francisco
O'Reilly AI Conference	June, 2017	New York
Strata Data Conference	May, 2017	London
O'Reilly AI Conference	September, 2016	New York

Table 2.1 – List of Conferences Consulted

The gathering of case studies was supplemented with a literature review of various articles, surveys, reports by market research firms, consultancies and others.

Qualitative analysis was performed which parallels the interpretive research paradigm and allows the drawing of general insights to address the research questions. Given the diversity of the cases, apples to apples quantitative methods were not applicable.

2.3 Limitations

Artificial Intelligence is a broad, diverse, dynamic field encompassing a range of technologies thus a comprehensive survey of literature and companies is not feasible. Instead this thesis aims to pull together a practitioner-focused practical, relevant and concise guide. A different approach could have been to focus by industry and technology – e.g. 'deep learning for recommender systems in retail' but this thesis aims to cater to a broader audience, and to take a high level CIO perspective rather than one deep in technical details.

The case studies based approach is also open to critiques, such the reliability and generality of the findings given one is extrapolating from a small number of cases, and whether 'intense exposure to study of the case biases the findings'[12]. As a means to address the first critique, an attempt has been made for a broad based selection of cases (see table overleaf) that are cross industry and that tackle a range of problem statements. As a means to address the second critique, the thesis also incorporates a framework that is more wide than deep, to counter bias/limitations in the selection of cases.

[12] ("The Case Study as a Research Method," 1997)

Company	Danske Bank	Deutsche Telekom	GE	General Mills	KAISER PERMANENTE
Industry	Banking	Telecommunications	Conglomerate[13]	Packaged Foods	Healthcare
Problem Statement	Fighting financial fraud	Improving customer service	Augmenting monitoring workflows for the industrial internet	Automating insights for marketers	Improving patient flow forecasting

Table 2.2 – Case Selection Spans a Number of Industries and Problem Statements

[13] Case study discusses examples in Aviation, Power and Oil & Gas

3 Literature Review

This chapter performs a literature review. This is a vast area, so this is just the tip of the iceberg. The next chapter extends the literature review into formulating a framework for building an enterprise AI capability.

In this chapter, Section 3.1 provides historical background, Section 3.2 discusses the opportunities in the space, Section 3.3 discusses benefits of applying AI technologies, Section 3.4 outlines some of the challenges, and Section 3.5 presents a helpful organizational maturity model.

3.1 Background

Artificial Intelligence technologies have existed since the 1950s, but smaller subsets - first machine learning, and then deep learning, a smaller subset of machine learning - have created the most recent disruptions.

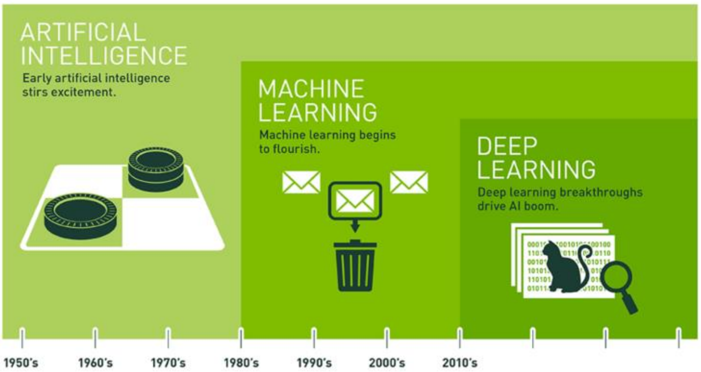


Figure 3.1 – AI Technologies Since the 1950s[14]

The field has experienced several hype cycles followed by periods of disappointment and cuts in funding ('AI winters') followed by renewed optimism. Given the key enabling factors now of explosion in data sources and the increase in compute power, AI technologies are now demonstrating substantive achievements and another AI winter is not likely. The below figure is often commonly cited in the literature, demonstrating the impressive performance gains of large neural networks over traditional machine learning as the size of datasets increases.

[14] (Copeland, 2016)

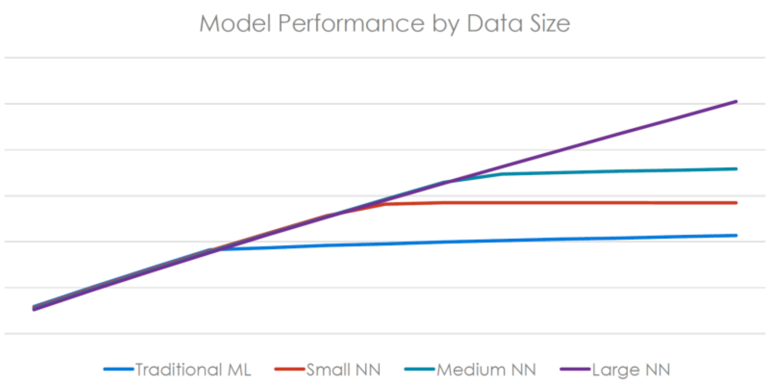


Figure 3.2 – For Large DataSets, Large Neural Networks Show Impressive Performance Gains[15]

3.2 Opportunities

Extracting value from data remains an untapped opportunity in many organizations

According to Harvard Business Review[16], less than half of an organization's structured data is actively used in making decisions—and less than 1% of its unstructured data is analyzed or used at all. Meanwhile more than 70% of employees have access to data they should not, and 80% of analysts' time is spent simply discovering and preparing data.

Sizing the AI opportunity

According to an McKinsey report, the total annual value potential of AI alone across 19 industries and nine business functions in the global economy came to between \$3.5 trillion and \$5.8 trillion. This constitutes about 40 percent of the overall \$9.5 trillion to \$15.4 trillion annual impact that could potentially be enabled by all analytical techniques. McKinsey analyzed more than 400 use cases across 19 industries, and found marketing and sales, and supply-chain management and manufacturing are among the functions where AI can create the most incremental value.

Examples of enterprise use cases

AI potential exists in every industry. Some common enterprise use cases are given in the table below to given an indicative idea.

Use Case	Description
Recommender Systems	Enable more effective suggestions, based on context for individuals, based on a particular objective such as purchase or lifetime value
Computer Vision	Enables dramatically more accurate visual recognition tasks that include image classification, detection and localization
Fraud Detection	Enables real-time detection of events in credit cards and e-banking. Enables fraud prevention, cybersecurity and system optimization
Text and Speech Understanding	Better service and automation for diverse applications such as call center, chat, field service, and medical records
Predictive Maintenance	Improves preventative measures & performance with greater accuracy at the asset & component level
Document Automation	Enables automation of processes that are human-intensive with higher speed and accuracy with paper or legacy apps

Table 3.1 – Examples of Enterprise Use Cases[17]

[15] (Bodkin, 2017)
[16] (DalleMule & Davenport, 2017)
[17] (Bodkin, 2017)

3.3 Benefits

Benefits from AI technologies lie in two distinct areas[18]:

- **Breakthroughs** that open new revenue streams, expand into new markets, create new products
- **Operational efficiencies** that compound through constant incremental improvement

A great illustration of both of these exists in Jeff Bezos's 2016 Annual Letter to Shareholders[19]:

> At Amazon, we've been engaged in the practical application of machine learning for many years now. Some of this work is highly visible: our autonomous Prime Air delivery drones; the Amazon Go convenience store that uses machine vision to eliminate checkout lines; and Alexa, our cloud-based AI assistant.
>
> …But much of what we do with machine learning happens beneath the surface. Machine learning drives our algorithms for demand forecasting, product search ranking, product and deals recommendations, merchandising placements, fraud detection, translations, and much more. Though less visible, much of the impact of machine learning will be of this type — quietly but meaningfully improving core operations.

3.4 Challenges

Applying AI technologies comes with several challenges. In a global survey of 3000 executives, researchers found that for 'Pioneers' talent was the main gap, while 'Passives' saw no business case.

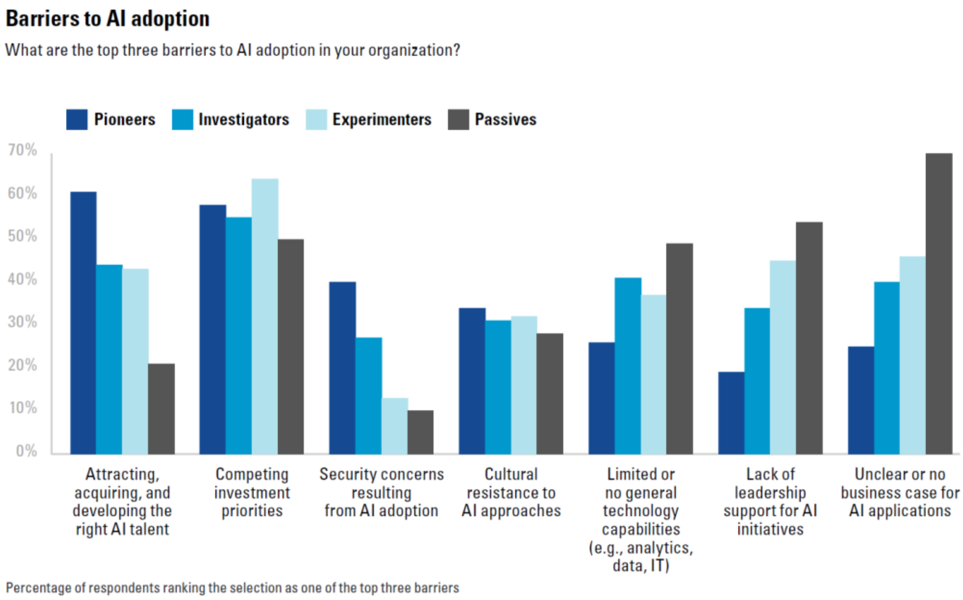


Figure 3.3 – BCG-Sloan Management Review Survey: Barriers to AI Adoption[20]

[18] (Elprin, 2018)
[19] (Heath, 2017)
[20] (Ransbotham et al., 2017)

Aside from the barriers in the figure above, the application of AI technologies comes with some distinct challenges:

Differences in Understanding AI's Abilities

There is very little consistency even in vendor and analytical reports as to what is and is not AI, and it is a very broad based term that encompasses a range of technologies. Thus there can be confusion in the enterprise as to what AI's abilities actually are, and that can hinder adoption and extracting value. To paraphrase from a practitioner – in an organization, one will encounter both those who are sceptics and those who think the technology will cure cancer, and expectations for both groups have to be managed to the reality. Educating executives on what is and is not possible should be key, in order to have business units submitting ideas for and driving AI projects.

Laborious Labeling of Training Data

Supervised learning (the more common variant used) needs labeled training data, and that can involve laborious annotations. For instance, self-driving cars are underpinned by an entire fleet of humans labeling things.

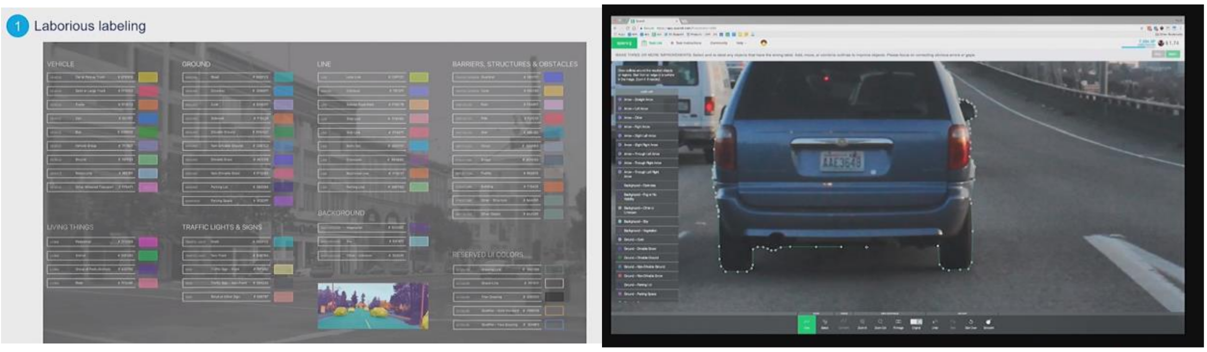


Figure 3.4 – Self Driving Cars Require Laborious Labeling[21]

Massive Data Needs[22]

As a rule of thumb, about 10 times as many examples are needed for training as there are degrees of freedom in a model. Today Google for instance performs image classification on sets of 300 million photos (this is significantly larger than ImageNet, which is the visual database used for deep learning challenges) and has released a dataset of 7 million YouTube videos. Organizations that need to use deep learning techniques should be thinking strategically about how they can acquire the data.

Need for Model Retraining

The moment you put a model in production, it starts degrading, known as 'concept drift', thus ongoing data acquisition for retraining AI systems is necessary. McKinsey's analysis of 400+ use cases found that one out of three use cases requires model refreshes at least monthly and almost one in four cases requires a daily refresh.[23]

The rate of degradation depends on the nature of the problem, not the algorithm, and the appropriate infrastructure needs to be in place to retrain as needed.

---

[21] (Chui, 2018)
[22] (Chui, 2018)
[23] (Chui et al., 2018)

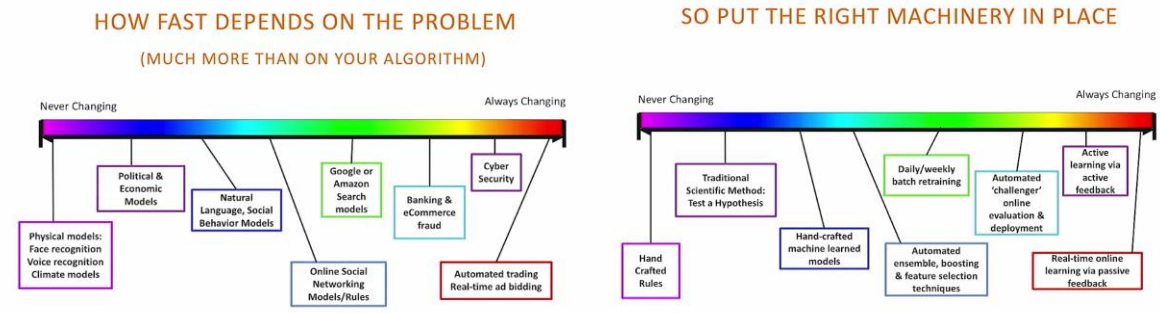


Figure 3.5 – Rate of Model Degradation Depends on Problem So Put the Right Machinery In Place[24]

Algorithmic Bias and Brittleness

Training data can have bias, and AI applications can absorb and amplify that. For instance, causal testing of a loan model by researchers revealed an 11% accidental gender bias, then when this was fixed, a 34% race bias[25] - a 'whack-a-mole' problem. Also of concern is the fact that opaque classifiers can be tricked by random noise imperceptible to humans to misclassify.

Wild West of Desktop Data Science

The top half of the figure below shows the aspirational connected loop of how model-driven organizations should be operating. Models are developed in a lab environment by data scientists, as they move from development to deployment there is a way for them to get validated/reviewed, and then once they are in production they receive ongoing feedback which they incorporate to improve. Most effective organizations will try to maximize speed around this cycle. However, the reality for most organizations can be different (as shown in the bottom half of the figure below). Data scientists can go off on tangents, they can be making science fair projects that never get deployed, their work can be stuck in the validation & review process, someone in IT could be re-implementing their Python or R code in Java, and there is ad-hoc one-off production that does not provide feedback to the original model.

[24] (Talby, 2017)
[25] (Black, 2017)

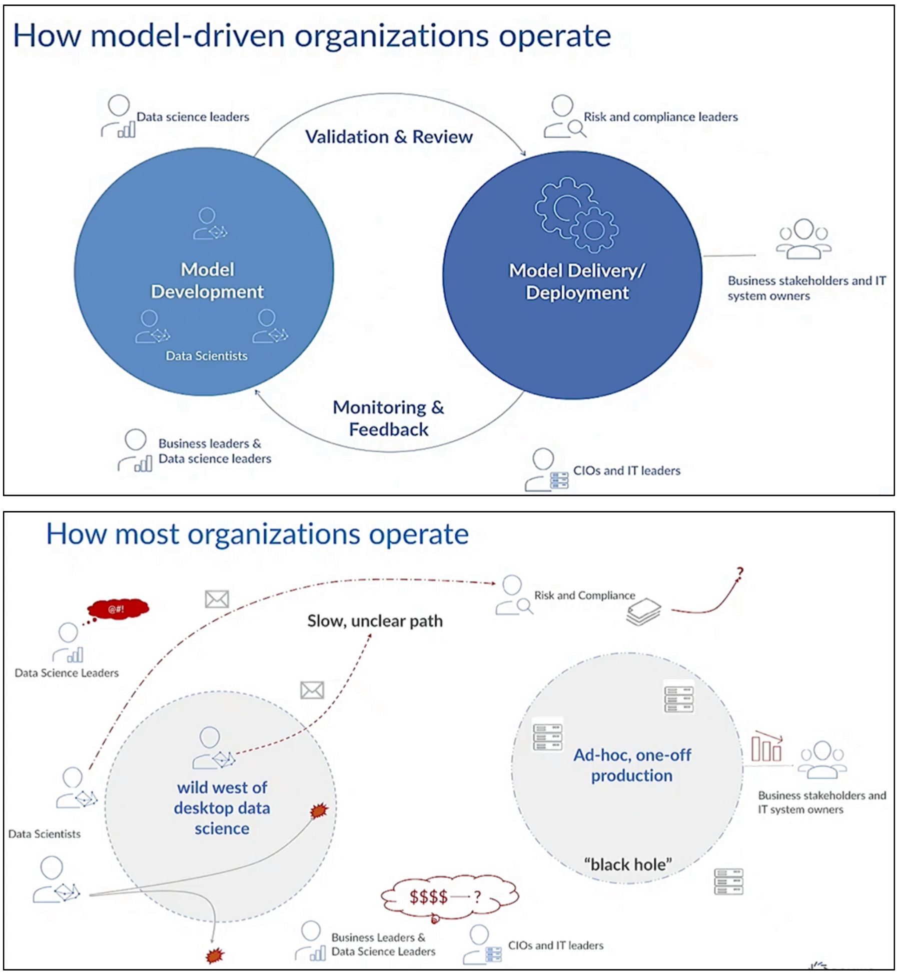


Figure 3.6 – Ideal vs. Common State of Organizational Data Science[26]

3.5 Enterprise AI Maturity Model

Enterprises vary where they are in their AI journeys, and based on that their needs can be different. Thus it is useful to think in terms of maturity models, and a few exist in the literature. A helpful one from Accenture is given below. As organizations move from ‘ad-hoc’ to ‘organize’ to ‘tactical’ to ‘mission critical’ to ‘industrial’, they progress on five dimensions to reach the below goals:

- **Strategy & Governance:** Run like a product
- **Architecture:** Data-centric and Secured
- **Development Process:** Agile and Dynamic
- **Regulation and Ethics:** Trusted & Transparent
- **User Support:** Self-service & Optimized

The figure below gives details what progress at each stage looks like.

[26] (Elprin, 2018)

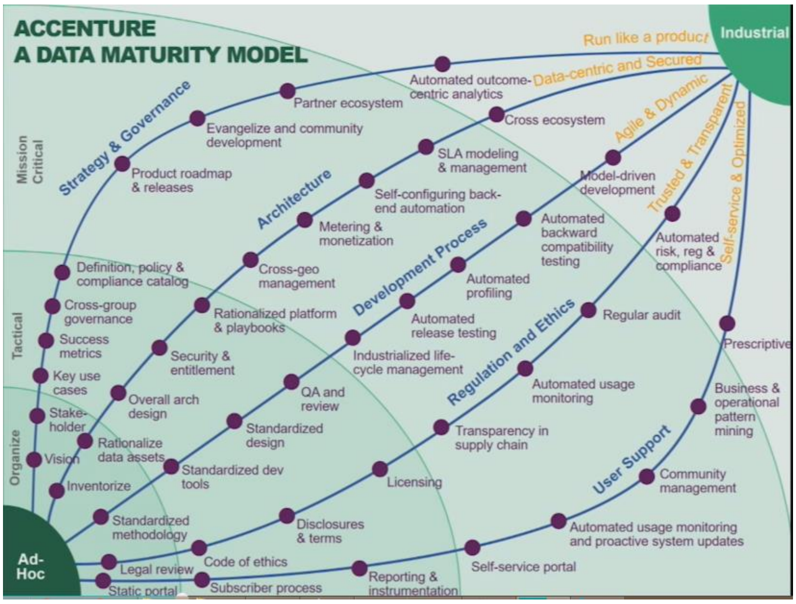


Figure 3.7 – Accenture Data Driven Enterprise Maturity Model[27]

[27] (Tung, 2017)

4 Framework for Building an Enterprise AI Capability

This chapter describes a framework for building an enterprise AI capability. The purpose of this chapter is to provide an overview of the various elements that go into building an enterprise AI capability. This section is more 'wide than deep', and it is meant to supplement the case studies in the following chapter, which are more 'deep than wide', and which highlight in detail how real organizations made decisions around certain elements in the framework. Each of the elements could in turn have a whole book written on them, but the objective of this chapter is not to give an exhaustive treatment but rather a holistic, high level overview.

This chapter is organized using a historic strategic management framework introduced by McKinsey, called the 7-S model. This framework "maps a constellation of interrelated factors that influence an organization's ability to change. The lack of hierarchy among these factors suggests that significant progress in one part of the organization will be difficult without working on the others."[28] The framework has been adopted and tailored towards building an enterprise AI capability, as that too is a major change initiative that requires the interplay of all of the elements below:

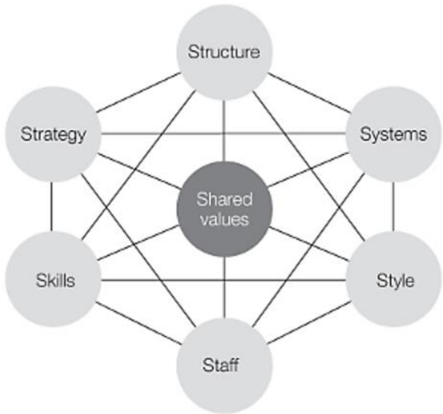


Figure 4.1 – 7-S Framework[29]

Sections 4.1 through 4.7 discuss briefly each of the elements of the above framework in the context of building an enterprise AI capability: shared values, strategy, structure, systems, style, staff and skills.

4.1 Shared Values

Shared values are at the heart of the model. These represent an organization's central beliefs and attitudes. It is important to cultivate the following values:

Product Mindset towards Data and Models

Successful AI driven organizations have a product mindset towards data and models. This means data and models drive business outcomes, are 'first class citizens', and have a similar lifecycle like traditional products (see figure below):

[28] (Bryan, 2008)
[29] (Bates & McGrath, 2013)

Figure 4.2 – Illustrative Product Lifecycle[30]

Think about a traditional product like say a car. You start off by conceptualizing – how will my car differentiate, e.g. is it economical, sporty or luxury? Similarly for an AI product consisting of data and models, you first need to start with defining the outcomes that will be unlocked through applying AI to data. Then you design your car – for an AI product, this can e.g. mean identifying the data sources and enriching them (e.g. adding user generated content, partner generated content, IoT etc.). You move on to manufacturing (building the models). Rolling out becomes your 'DevOps' – i.e. do you serve and manage data and models at scale in production, and also, how do you get your user community (business analysts, domain experts etc.) to actually use your AI product. Lastly, you have to think about service, i.e. have you built in feedback mechanisms from users, so your AI product can be continually improving and learning.

Data is a Shared Strategic Asset

In many organizations, data exists in silos. This can happen for many reasons – there can be non-integrated systems, different owners with different objectives, and/or lack of data sharing because of security concerns, or turf-ism, or sometimes even just embarrassment at the state of the data. It is important to overcome this, e.g. by integrating data and adapting the approach Brent Gleeson posits for breaking down organizational silos:

Figure 4.3 – Top 5 Ways to Break Down Organizational Silos (Brent Gleeson)[31]

Cloudera expands this value well, breaking it down into six principles of an enterprise data architecture:

[30] The figure and the car example are drawn from: (Tung, 2017)
[31] (Eggers, 2017)

Figure 4.4 – Cloudera's Six Principles of an Enterprise Data Architecture[32]

Large enterprises can also benefit from thinking strategically about data accumulation loops, as is done by tech companies and AI start-ups. Andrew Ng discusses a 'virtuous circle'[33] of AI, where companies that want to build defensible businesses in AI gather just enough data to launch a product: thus they get users; with users they get more data; and applying AI to that data yields an even greater product, which becomes a reinforcing loop that becomes difficult for competitors to replicate.

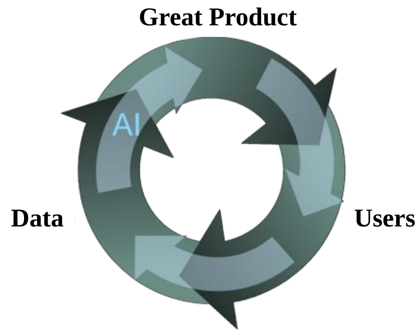


Figure 4.5 – Virtuous Circle of AI[34]

Being a Learning Organization / Experiment Driven Organization

The concept of a learning organization is not a new one – Peter Senge expounded on it in his 1990 book 'The Fifth Discipline'. The concept can be extended in an enterprise AI context to being an experiment driven organization. Wilder-James gives 3 core principles of experimenting[35]: experiments must be cheap in order to de-risk failure, they must be fast in order to learn quickly from feedback loops, and they must not break the important production processes of a business. He proposes six foundational elements of an experimental enterprise which the figure below illustrates along with a mapping of the capabilities they support.

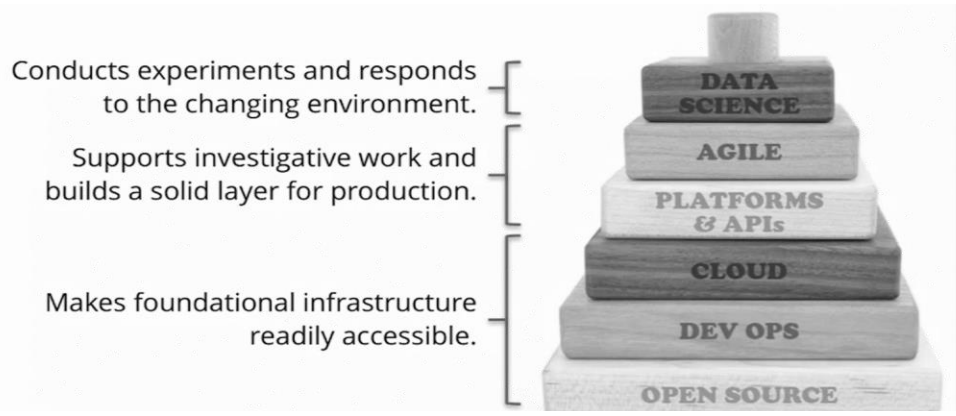


Figure 4.6 – Foundational Elements of an Experimental Enterprise[36]

Trust and Transparency

For the models to be useful their results must be trusted, and in order to be trusted, there needs to be transparency in how the AI models arrived at their conclusions. Building in explainability is important for the benefit of all stakeholders: data scientists, business users, customers, and regulators. This is an

[32] (Barron, 2017)
[33] (Ng, 2017)
[34] (Ng, n.d.)
[35] (Wilder-James, 2014)
[36] (Akred, 2017)

important area of research particularly with reference to deep learning algorithms that can be black boxes. Under the new European data privacy law (General Data Protection Regulation), businesses that cannot explain to consumers why decisions were made about them will be subject to heavy penalties.

Also related to trust is the concept of bias – human prejudices can exist in training data and there is the possibility that AI learns these and amplifies them. Organizations should be monitoring for and limiting bias but it is still early days in this area[37]. The AI Now Institute, an interdisciplinary group of researchers together with the American Civil Liberties Union, was founded in 2017 to identify and highlight algorithmic bias.

4.2 Strategy

Strategy defines what to do and what not to do – it is the detailed plan that allocates an organization's scarce resources over a period of time in order to achieve identified goals.

Whether firms engage in a full blown roadmapping exercise (e.g. a consultancy's methodology for this is shown in the figure below), or have a simpler mechanism say of asking business stakeholders to submit projects and reprioritizing these quarterly, the important thing to note is that business objectives/problems need to be driving the decision of what to do.

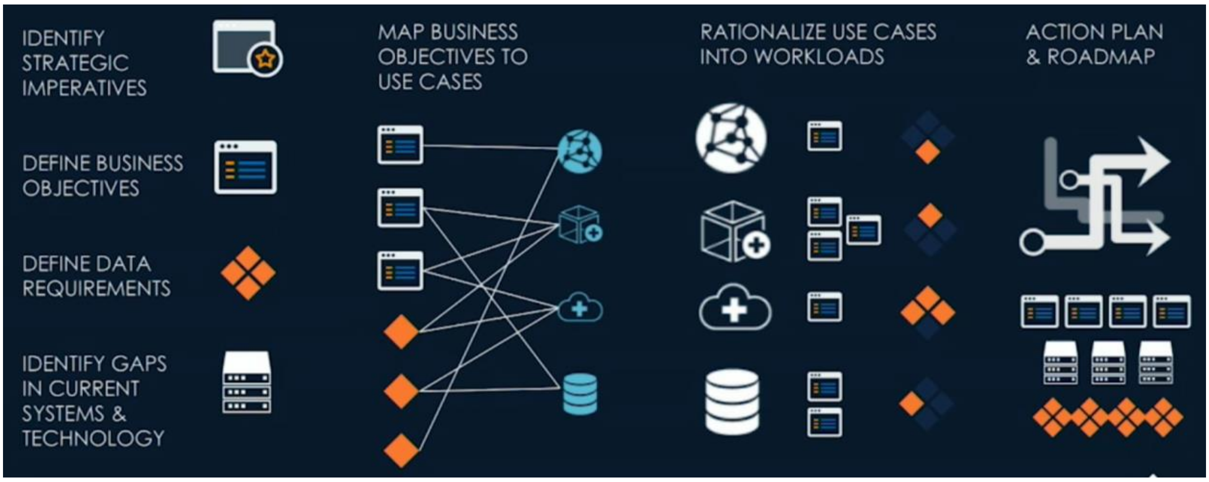


Figure 4.7 – Method for Data and Analytics Strategy (Silicon Valley Data Science)[38]

That may seem obvious – but it is surprisingly common in organizations for data scientists to let themselves loose on datasets, and to start poking around these, and from that analysis trying to define the end product and key performance indicators (KPIs). Projects can turn into science fair experiments, stakeholders lose enthusiasm, and measures of success are vague. The better method is to start with the business problem, define the relevant KPIs, define the product requirements, identify the necessary analysis, and source the data needed for the analysis. The following figure contrasts the two approaches:

[37] (Knight, 2017)
[38] (Akred, 2017)

Figure 4.8 – Typical Approach for AI/Data Science Projects vs. Better Method[39]

Project prioritization needs to take into account three factors: the value at stake, the estimated effort (which should include costs of integration, maintenance and retraining), and the forecasted risks (their likelihood x impact). Forecasted risks include potential barriers to adoption as well as consequences of the model not performing as expected (e.g. brand/reputation risk). These are depicted in the figure below, with the size of the bubble denoting risk.

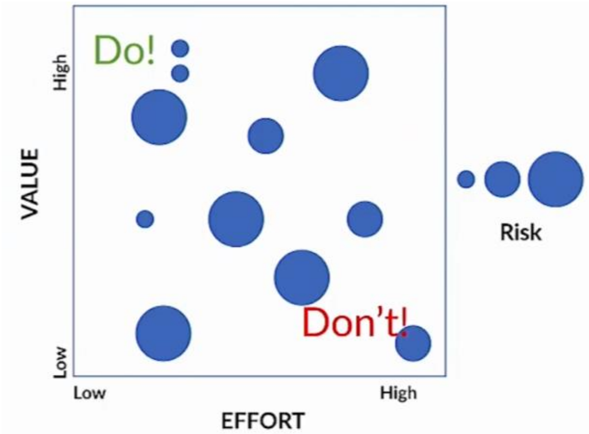


Figure 4.9 – Project Prioritization[40]

4.3 Structure

Structure refers to how an organization's units relate to each other. Enterprise AI capability typically starts out in a decentralized fashion – in this case the AI/data science function is duplicated per business unit. It has the advantage of being closer to the business, its issues and customers. It has the disadvantage of creating silo-ed functions that do not share best practices, with issues of redundancy, inconsistency, and lack of standardization. As these teams scale, and there are more commonalities to pull out, it is common to move to a semi-centralized model, where there is a central 'hub' which has developed shared expertise and codified best practises, while business units also have some embedded 'spoke' capability. A fully centralized model is less liked because it runs the risk of being too far removed from business priorities, with data scientists being confined to an ivory tower, and not accountable to business units. The following figure from Microsoft expands on the pros and cons of these different organizational structures.

[39] (Elprin, 2018)
[40] (Elprin, 2018)

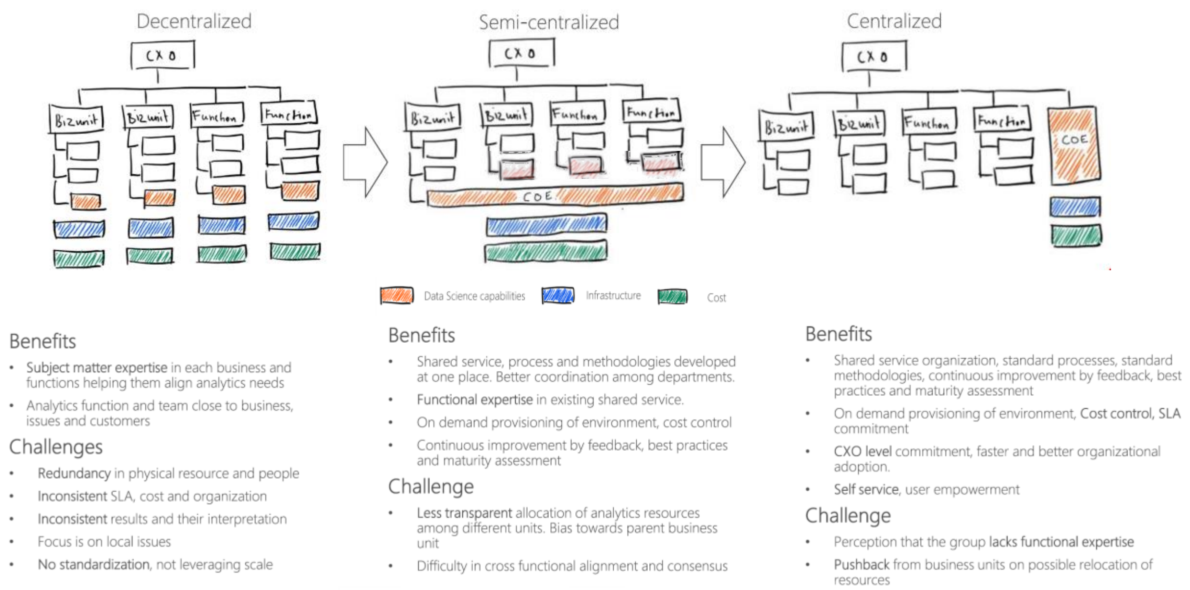


Figure 4.10 – Comparison of Different Organizational Structures for AI/Data Science Capabilities[41]

4.4 Systems

Systems refer to hardware and software systems as well as the processes that are used to undertake work. Let us touch on these in an enterprise AI context:

Hardware Systems

Enterprise AI systems need additional degrees of infrastructure flexibility and scalability, which cloud providers like Amazon Web Services cater to. Traditional slow procurement of servers by IT does not work to meet the demands of AI teams – they require on demand, scalable compute. Deep learning models are computationally intensive, and large neural networks can require specialized hardware like GPUs.

Software Systems

The case studies provide detailed examples of the production grade software infrastructure built by organizations, so refer to that section for more exposition. Here let us note that in enterprise AI systems, the actual AI models are only a tiny part of the software infrastructure. Many organizations struggle to put models in production and to have feedback loops for ongoing improvement. A 'DevOps' approach for AI is still in its early days, with the tooling yet to catch up to the conceptual approach.

Processes

Established processes are needed for reliability, repeatability and scalability. If a firm is early in its AI journey though, it should not get paralyzed by a complex process – it can do a lightweight version and focus on getting early wins in to build confidence and trust. For instance, the below table shows a 'Project Kickoff' checklist – a light version of 'Stakeholder mapping' could simply be 'Have you talked to business stakeholders?', similarly a light version of 'Model delivery plan' could be 'Have you made simple mockups of how your model will plug into existing business systems?'. For exploring the data, building the models and experimenting, many reference processes exist and can be adapted, such as CRISP-DM, SEMMA, KDD, TDSP etc.

[41] (Wright-Jones & Lidberg, 2017)

Project Kickoff Checklist	Best Practices	Common Pitfalls
Business case definition	• Define the value at stake, effort, risks (of false positives/negatives), retraining requirements and change management requirements	• Not taking into account integration and retraining costs
Stakeholder mapping	• Define responsible parties from each group: data science, business, DevOps, application dev, compliance etc. • Treat as cross-functional teams	• Lack empathy with goal of actual end user • Throw results "over the fence" to IT with no context
Technology needs	• Consider opportunities to accelerate research • Identify dependencies early	• "One size fits all" tooling • Underpowered infrastructure
Data availability	• Leverage existing sources first to build baseline • Create synthetic data with realistic characteristics • Track engagement with datasets to automatically discover experts	• Wait for "perfect" data • Buy external data without clear onboarding plan
Prior art review	• Review state of the art – internally and externally	• "Not invented here" culture • Nose-to-the-ground mindsets • No single source of truth
Model delivery plan	• Design multiple mock-ups of different form factors • Design approvers in advance (IT, analytics, business) • Create process flow to precisely show where model will impact • Proceed incrementally to get feedback from real usage	• Fail to educate end-users who revert to old habits • Over-engineer relative to the requirements
Success measures	• Pre-emptively answer "how will we know if this worked" • Frame in terms of business KPIs not statistical measures • Define needs for holdout groups, A/B testing etc.	• Not knowing when it is "good enough" • Fail to establish testing infrastructure and culture
Compliance and regulatory checks	• Consider consequences of errors (e.g. false positives/negatives) • State likely biases in training data • Track ongoing usage to prevent inappropriate consumers	• Assume no regulation today will last • Conflate model interpretability with model provenance

Table 4.1 –Project Kickoff Checklist[42]

4.5 Style

Style refers to the cultural style of an organization and how key managers behave in achieving the organization's goals. As such it derives directly from the shared values.

Agile and Experimenting

Iterate quickly in short sprints to unlock real value fast. The concept of agility extends to tooling – technologies in this space are evolving rapidly, no one single vendor is best of breed, and in order to keep pace teams need to not be locked in to any vendor/tool. Furthermore, the application of AI

[42] Adapted from (Elprin, 2018)

methods, deep learning in particular, is an experimental science not a theoretical one – you cannot predict what will work, so you need to keep experimenting. This means managers' expectations need to be set that not all attempts will work, but it is not 'failure', it is 'learning'.

Collaborative

Projects need to be done in cross functional teams with close collaboration between data scientists, business stakeholders, application developers, IT DevOps people etc. The collaborative style should also be backed up by the tooling – there should be a shared context / discussion area / knowledge management tooling etc.

Focus on Reproducibility and Reusability

Reproducibility and reusability underlie being a true learning organization. This means code, data, results, and environments can be recreated, and that management recognizes that reusable knowledge trumps producing an answer.

4.6 Staff

Staff refers to how to attract, assess, on-board/train, manage and retain top-notch talent. This area deserves focus as many organizations cite a talent gap as a key obstacle to overcome, and there is a lot of competition, from tech companies in particular, for a limited pool of trained resources. The following table gives some best practices and common pitfalls in this area. While the table has an external-hire orientation, companies can also find success by upskilling their existing resources.

How To	Best Practices	Common Pitfalls
Attract	• Have a differentiated offering and strategy • Advertise projects, not just the company • Offer modern tools and commitment to open source	• Write unrealistic job descriptions • Seek PhDs when need hackers (or vice versa)
Assess	• Be systematic: identify required attributes, design assessments for each • Be analytical: track interviewer and interview type efficacy • Include EQ and non-technical assessments • Sell while assessing: simulate real work • Set expectations on time allocation (time on job will be spent not only on just building models but also on data prep and on listening and talking to stakeholders)	• Over-rely on tech screens, not getting good sense of EQ • Subsequent churn from flawed expectations on time allocation

On-board/Train	• Reinforce mindsets, not just skills
	    ○ Develop culture of reuse, compounding
	    ○ Reward community-enhancing behavior
	• Provide "soft" skills training
		• "Not built here" mentality

Manage	• Share accountability with the business's KPIs • Focus on iteration velocity • Systematically capture stakeholder feedback and engagement	• Measure everyone but yourself • Over-index on any one project vs. factory performance

How To	Best Practices	Common Pitfalls
Retain	• Build in mentorship and community • Recognize and reward • Challenge and build in personal development	• Isolation - locked in ivory tower innovation silos • Repetitive work, lack of personal development

Table 4.2 – Staff: Best Practices and Common Pitfalls[43]

4.7 Skills

Enterprises looking to quickly build an enterprise AI capability may be seduced by the shortcut of hiring a number of ML and/or deep learning PhDs to staff a center of excellence. Taking that approach, however, will just yield overwhelmed and ineffective experts. The figure below from Microsoft illustrates some quotes when that approach is taken.

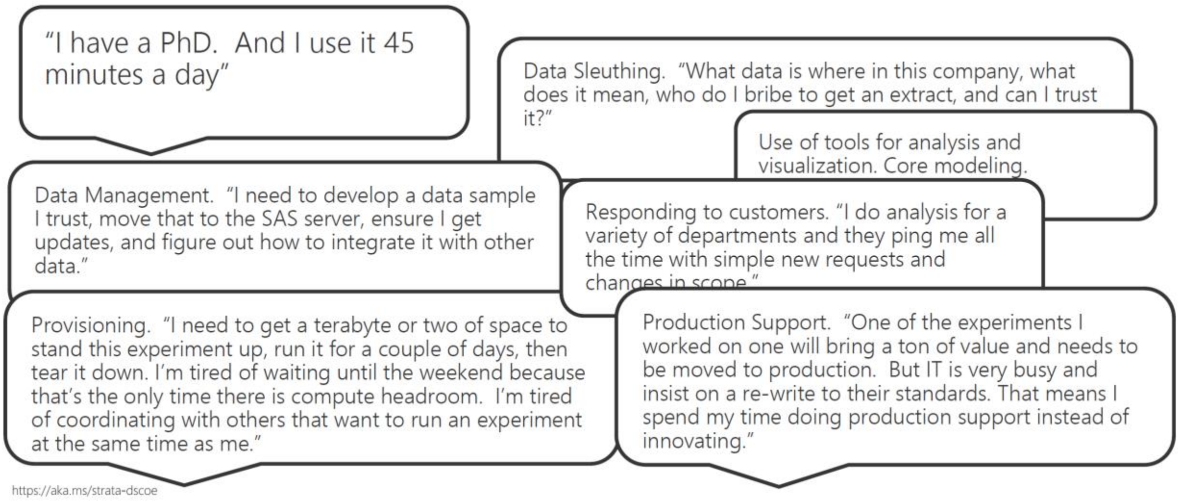


Figure 4.11 – Illustrative Quotes – Overwhelmed Data Scientists[44]

It takes a village of supporting roles other than data scientists to ensure production grade AI applications, as shown in the table below. If looking to hire the bare minimum, CIO Magazine quotes an IT executive:

"At first, we attempted to recruit for a single role – a data scientist – who had the all of the capabilities we needed. That approach did not work out," says Chris Brazdziunas, vice president of products at LogRhythm, a security intelligence company. "In our experience, we found that an AI group needs at least three distinct roles: a data engineer to organize the data, a data scientist to investigate the data and a software engineer to implement applications."[45]

Role	Priorities
Data Scientist	Generating and communicating insights, understanding the strengths and risks
Data Infrastructure Engineer	Building scalable pipelines and infrastructure
Developers	Implementing applications, productizing data science work
Solution Architects	Architecting the IT solutions such that experiments can be put into production
Business Architects	Organizational change management, working with business to find the right datasets

[43] Adapted from (Elprin, 2018)
[44] (Wright-Jones & Lidberg, 2017)
[45] (Harpham, 2017)

Role	Priorities
Business Stakeholders	Vetting the prioritization and ROI, providing ongoing feedback
Data Product Manager	Articulate the business problem, translate to day-to-day work, ensure ongoing engagement
Data Storyteller	Creating engaging visual and narrative journeys for analytical solutions

Table 4.3 – Roles in an Enterprise AI function[46]

[46] Adapted from (Elprin, 2018)

5 Case Studies

This section uses a case study format to explore how five large enterprises applied Artificial Intelligence technologies in varied problem domains. The case studies discuss the challenges they faced, the decisions they made, the capabilities that they built and the lessons that they learned. The five sections cover:

Section	**5.1**	**5.2**	**5.3**	**5.4**	**5.5**
Company	Danske Bank	Deutsche Telekom		General Mills	KAISER PERMANENTE
Case Title	Fighting financial fraud with Artificial Intelligence	AI driven customer care	Building machine learning applications for the industrial internet	Automating business insights for marketers through AI	Using AI to improve patient flow forecasting

Table 5.1 – Chapter Five Organization

5.1 Danske Bank: Fighting Financial Fraud with Artificial Intelligence[47]

About the Bank

Danske Bank, headquartered in Copenhagen, is a Nordic bank that is over 145 years old. In 2017, it had 19,000 employees, 2.7 million personal customers, 236,000 small and medium-sized business customers, and 1,800 corporate and institutional customers. Its vision is to be the most trusted financial partner in the Nordics, and gaining insight into its customers by being data driven is a key enabler of that vision.

The Problem Statement

For Danske Bank, fighting fraud is an ongoing challenge. There can be two types of fraud – either customer initiated (where customers themselves initiate the transaction, e.g. sending money to an investment scam), or fraudster initiated (e.g. identity theft). The figure below shows examples in each of the two categories.

Figure 5.1 – Two Types of Fraud

Fraud loss can cost Danske Bank tens of millions of euros each month. But detecting fraud comes with many challenges:

- **Low detection rate**: only around 40% of fraud cases were being detected
- **Many false positives**: the current approach to flagging fraud resulted in a staggeringly high rate of false positives - 99.5% of cases being flagged were not fraud related
- **Fast evolving fraud sophistication**: fraudsters keep getting increasingly sophisticated, in fact they too are evolving to use artificial intelligence

Launching the Fraud Project

The bank partnered with Think Big Analytics (acquired by Teradata), a big data consultancy with 500+ employees, that offered full spectrum consulting, data engineering, data science and support. Think Big marketed itself as being vendor-neutral with an open source focus, and had fixed fee offerings for data science and engineering. From the Danske Bank side, the fraud project had the full high-level executive support of the Head of Global Analytics, Nadeem Gulzar. Though the project's steering committee had some sceptics, Nadeem was committed to realizing the following ambitious goals for the project:

1. Reducing false positives and increasing the fraud detection rate

[47] All content in this case study, including all figures and tables, is sourced from this conference video: (Bodkin & Gulzar, 2017)

2. Moving from the traditional expert rules based fraud engine to a data driven approach with real time scoring of transactions
3. Building an advanced analytics blueprint. Nadeem wanted this project to serve as a blueprint/foundation for future projects and use cases, helping to realize Danske Bank's ambition of becoming one of the banks leading in advanced analytics capabilities.

Approach for Advanced Platform for Fraud

The high level approach to creating and running the advanced platform for fraud was:

1. Understanding the domain, to have an idea of the features that might be important
2. Gathering and preparing the data – this is where 80% of the time was in fact spent
3. Training models on historical data so they could automatically be able to generate rules/recognize fraudulent patterns.
4. Automatically maintaining the engine by retraining the model

The old approach to fraud detection entailed using expert-defined rules. The pros and cons of this new approach to fraud detection are summarized below:

Pros	Cons
Automatic/data-driven/objective inference of the rules	Might be unintuitive and hard to interpret
Ability to detect patterns in high dimensional data	Data preparation and feature aggregation is time consuming
Fast detection of new/changing fraudulent patterns	

Table 5.2 – Pros and Cons of Moving from Expert Rules Based to Data Driven AI Based Fraud Detection

Modeling Challenges

Though the bulk of the time was spent in gathering and preparing the data, building the models was not without its own challenges:

- Class imbalance: this means that the ratio of non-fraud to fraud training data is severely skewed in favor of the former, to the tune of 100,000:1
- Assigning fraud labels from historic data was also challenging. The further back in time you went, the worse the data quality. Fraud itself can be ambiguous
- Not all features predictive of fraud were available in real time, but from a customer point of view fraud detection needed to happen in real time - in milliseconds
- Most machine learning sees transactions atomically and is not seeing the bigger picture

Cross Functional Collaboration to Deliver Value in Each Iteration

An agile approach was employed, with a cross functional team. The first track to kick off (see timeline figure below) was the data science track, which focused on getting the data, preparing it so that it was of the right quality, and ensuring that all the features were in place. The engineering track kicked off a month later, and focused on getting the model in shadow production. The team, despite being large with 30+ participants, was highly energized and collaborative, and managed to get from PowerPoint to shadow production in just 8 sprints. It was important for the model to cut its teeth in a production-like environment. The stakeholders could thus get familiar with the setup and the model could get a taste of the real live transactions coming in. It was very important to monitor the performance of the model and decide if it needed retraining or not. After three months in shadow production, the model was working at the desired level, and moved to live production. This marked the end of Phase 1 of the project. The timeline for this phase is shown below:

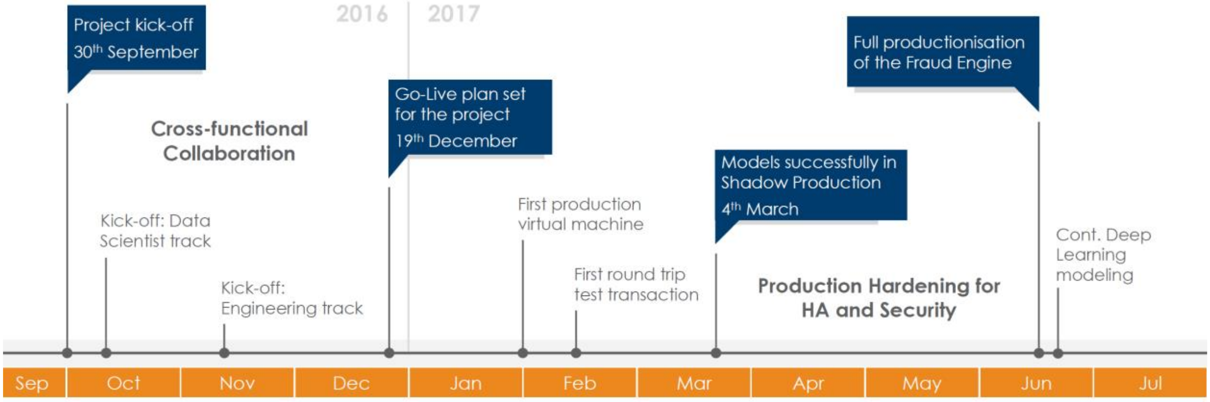


Figure 5.2 – Phase 1 Timeline from Kick-Off to Production

Three Capabilities of the Banking Anti-Fraud Solution

The complete solution solved for three distinct capabilities:

Data Modeling, Pipeline and Ingestion	Model Management Framework	Machine Learning and Artificial Intelligence
• Organization of silos of data • Real-time data integration • Security and procedures: following existing bank procedures	• Multiple models running in production at the same time • Mix of traditional (Phase 1) and advance deep learning (Phase 2) methods • AnalyticsOps: deploying machine learning models in production	• Hard to operationalize insights • Availability of analytic capabilities/skills and data • Interpreting the results of machine learning models

Table 5.3 – Three Capabilities of the Banking Anti-Fraud Solution

Augmenting the Existing Architecture

The advanced analytics platform for fraud (the top right box in the figure below) augmented the existing architecture. In the existing architecture, payment and banking systems go through large scale servers where a rules based fraud engine also sits. The team had to make an IBM mainframe call out to execute state of the art machine learning models.

Note that humans still stayed in the loop, just like in the existing architecture. The project had raised some concerns in the organization whether AI models would be taking over jobs.

The advanced analytics platform invoked a series of microservices that augmented transaction data with real time information and the context around the accounts involved in the transaction. Over time more and more contextual information could be added to enhance the models.

Ensemble techniques were used – i.e. the platform scored a number of different models in parallel and combined the results to come back with a decision. Besides increasing performance, this technique also created safety – if any model took too long to execute, the system could still come up with an answer. It also allowed A/B testing, and having incremental tests of new versions of models with a small percentage of traffic in a live environment.

As the project moved into the second phase (deep learning), the microservices architecture allowed for executing micro-batches of transactions on GPUs for efficient inference even though much of the rest of the work like logistic regression and boosted gradient decision trees was being done on CPUs.

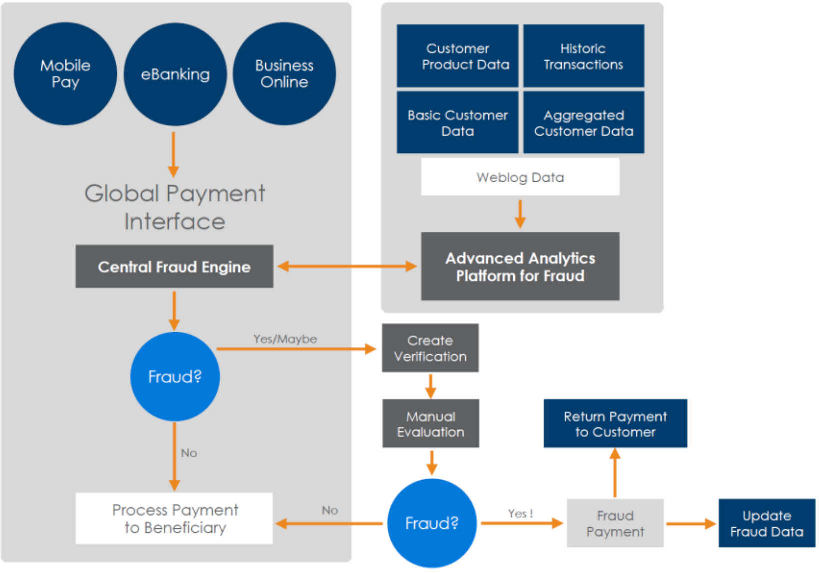


Figure 5.3 – Augmenting the Existing Architecture

Key Requirement: Model Interpretation

One of the key requirements for the system was model interpretability. Danske Bank needed to be confident in the models, and understand and trust their results for a number of reasons:

1. **Helping the investigators:** When a potential fraud is flagged, an investigator looks into it. They need to know what to look at. They always have too much work on their plate and not enough time, so anything that can help them improve their investigation is much appreciated. In fact, this ended up one of the biggest benefits of the project – the investigators found the information incredibly useful to improve their work.
2. **Consumer trust**: If a transaction is blocked consumers call up to find out why, and they need a good explanation else the bank would lose their trust.
3. **External regulatory requirements:** As a European bank, Danske Bank has to comply with the upcoming GDPR (General Data Protection Regulation). Per GDPR, companies that are unable to explain to consumers why decisions were made about them face severe penalties.
4. **Facilitating data science work:** E.g. if the data scientists needed to compare model performance (say pitting a champion against a challenger), they too needed to know what features were triggered, and why one was performing better than the other.

To solve this problem, the team deployed an open source solution that came out of the University of Washington called LIME (Locally Interpretable Model-Agnostic Explanations). The next phase of the project involved deep learning which is even more opaque, so LIME was essential to explain the key characteristics at the point of the decision that allow the model to classify as fraud/not-fraud.

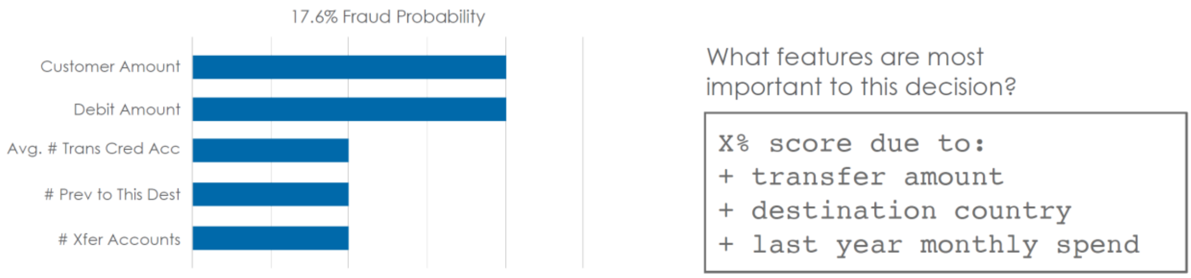


Figure 5.4 – Using LIME to explain the features most important to the decision

Machine Learning Results

So how did the machine learning results fare? Even better than expected – the team was shooting for a 30% false positive reduction rate, and a greater than 35% increase in the detection rate. It achieved a 60% reduction in the false positive rate. The increase in detection rate was anecdotally believed but the statistical evidence (as of the time of the talk from which this case study is sourced) would take longer, as it takes some time for consumers to check their statements and report fraud.

In the figure below, the red dot shows the performance of the traditional rule engine on the validation set. The green line shows the performance of the machine learning model, which is able to achieve the same and higher true positive rates as the rule engine at lower false positive rates.

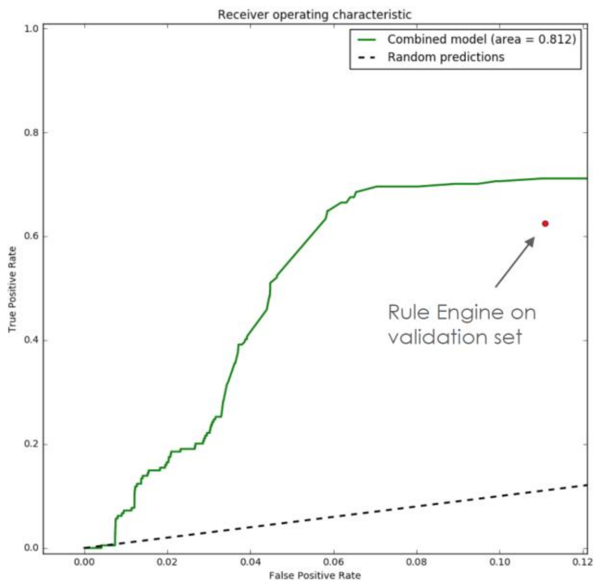


Figure 5.5 – Machine Learning Results Outperforming Traditional Rule Engine

Deep Learning Opportunity

The deployed machine learning models were only catching around 70% of all fraud cases. How could this be improved? The models had some limitations. Traditional machine learning models view transactions atomically and often missed fraud transactions that were part of a series, and could not capture correlation across many features. These limitations can be overcome by using deep learning – a technique usually applied in fields like computer vision and natural language processing. How can bank transactions fit that mold? The team transformed correlated features into a 2 dimensional representation – a ‘transaction image’ as it were where fraud and non-fraud cases looked different. These images were fed into neural networks. These networks had on the order of half a million nodes

– massive compared to traditional machine learning models for fraud. Visualizing and understanding the models is still ongoing work.

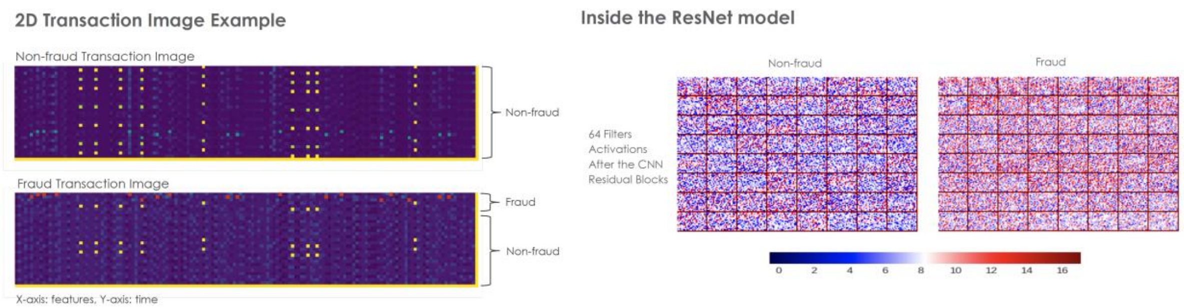


Figure 5.6 – 2D Transaction Images Fed to Residual Convolutional Neural Networks

Note that deep learning is more an experimental science than a theoretical one. The team had the choice of three possible deep learning architectures: convolutional neural networks (CNNs), LSTMs (Long Short Term Memory networks), and auto-encoders. Rather than go in trying to predict which architecture would give the best result, the approach was to try them all. It turned out that a version of convolutional networks (i.e. residual CNNs) worked best.

Deep Learning Results

The deep learning models (green, mustard and red lines in figure below) gave a really impressive improvement compared to the traditional machine learning ensemble method (blue line below, with red dot showing rules based engine performance). Note the graph is really zoomed in, as the business required a minimal false positive rate. Also, with all of these models, unlike the traditional rules based engine, the bank had the ability to make tradeoffs – i.e. to increase the detection rate at the expense of the false positive rate (slide along the axis).

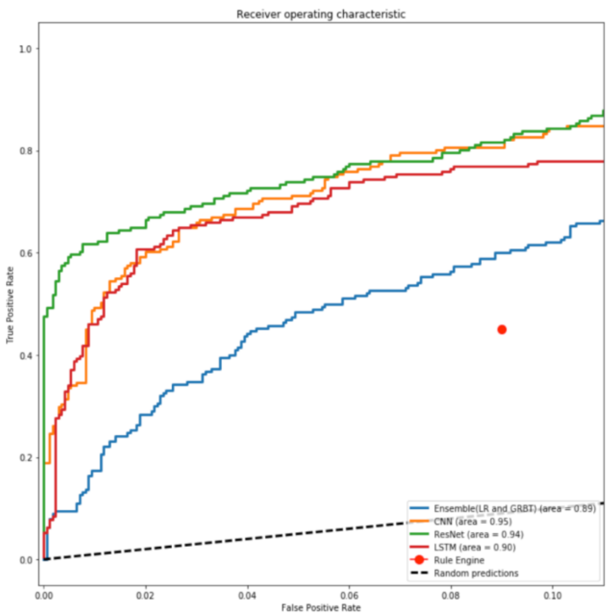


Figure 5.7 – Comparison of Three Deep Learning Models and the Traditional Machine Learning Ensemble Model

Conclusion

On the technical side, Danske Bank built the following capabilities:

- **Deep learning adoption** from pictures to financial transactions
- Enhancement of data quality & cluster capabilities with **data ingestion**
- Building **AnalyticsOps** capabilities to support business units
- **Leveraging experience** from fraud advanced analytics to deliver extra use cases

Equally important were the lessons learned on the soft skills side – going from PowerPoint to shadow production in 8 sprints was very much due to factors like executive sponsorship, a great collaborative team effort, and an agile approach. The team successfully spearheaded innovation in all involved systems and set up an inspirational blueprint for combatting new challenges in advanced analytics.

5.2 Deutsche Telekom: AI Driven Customer Care[48]

About the Company

Deutsche Telekom (DT), headquartered in Bonn, Germany, is a global telecommunications company with presence in over 50 countries. It is the largest telecommunications provider in Europe by revenue[49]. It has a number of subsidiaries worldwide, including T-Mobile in the US. In 2017, it had 74.9 billion euros in revenue, over 217,000 employees, 168 million mobile customers, 28 million fixed-network lines, 19 million broadband lines, and 7.4 million TV customers[50].

The Problem Statement

With over 160 million mobile phone customers, DT had ongoing massive volumes of customer calls and suboptimal customer care experience with customers having to wait, or repeat themselves as they were passed on to a next level of care without resolution. Customer care agents meanwhile were also tired of helping people with the same routine simple enquiries and would have preferred to focus on solving complex customer problems. DT figured it could use Artificial Intelligence to handle a large chunk of queries that are simple and repetitive so customers would not have to wait in service lines, and agents could be freed up to focus on more high value tasks.

Launching the AI Innovation Project: eLIZA

DT launched this AI initiative in mid-2015 as an innovation project dubbed eLIZA (based on the ELIZA[51] program created at the MIT Artificial Intelligence Laboratory in 1966 that allowed people to engage in discourse with the program, and that appeared to 'pass' the Turing test).

The eLIZA project was structured as a large distributed team – spread across 5 sites, in 3 different countries, with over 60 people. It was a collaborative effort between many different divisions of DT, including: Product Innovation in the board department Technology & Innovation, T-Mobile Austria, Telekom Deutschland, Telekom Servicegesellschaft and many more[52].

The team focused on agile development with quick and ongoing user feedback. The team decided to pilot in Austria and then do a broader rollout.

The team's vision was to create an intelligent digital assistant – not a chatbot that frustrates a user, but a virtual friend/assistant that helps customers out via human-like dialog. It would have potential far beyond customer service, as once one starts to engage customers in service problems, there are also opportunities to up-sell, cross-sell etc.

The assistant needed to be just around the corner for everybody – i.e. it had to be everywhere, every device, via any channel (text and voice capable), and it should listen to and learn from customers and improve the way it answers.

[48] Except where otherwise explicitly cited, all content in this case study, including all figures, is sourced from these two conference videos: (Hoffman, 2017), (Lynam-Smith, 2017)
[49] ("Top 10 European operators revenue is €51 billion in Q2 2017," 2017)
[50] ("Deutsche Telekom: At a glance," 2018)
[51] (Weizenbaum, 1966)
[52] ("eLIZA – the innovation project of Deutschen Telekom | welove.ai," 2018)

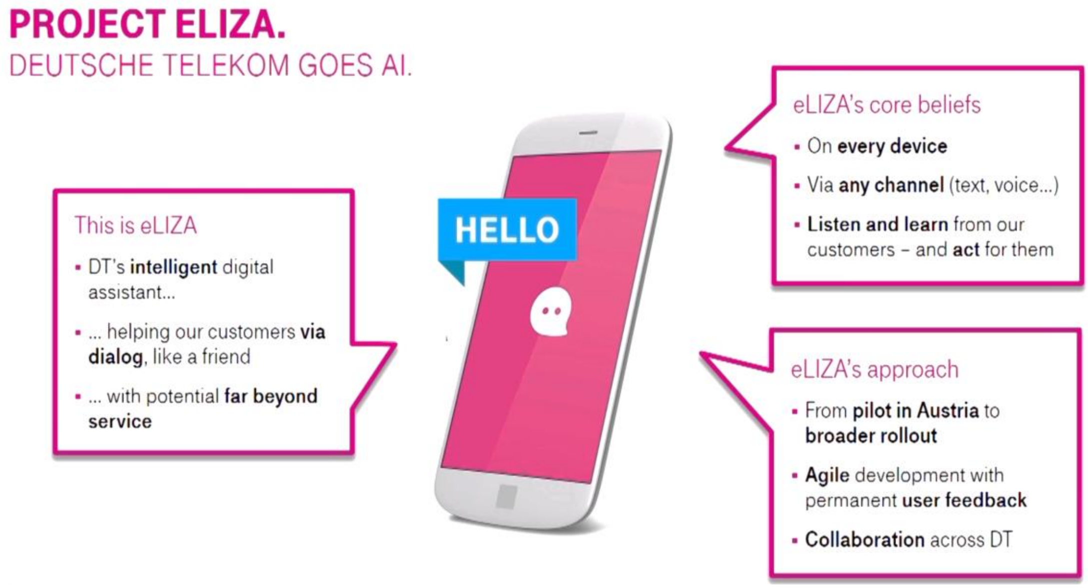


Figure 5.8 – Vision and Approach for Project eLIZA

Think Big, Start Focused

eLIZA had ambitious goals. The team picked a focused starting point. DT already had a rule based chatbot (this worked with simple keyword detection – with no AI), and the team decided to see how far it could get by improving the rule based chatbot by tweaking the conversational interface. The team also worked on making sure this chatbot was available at various touchpoints, including internal ones like the DT website and app, and external ones like Facebook Messenger. The avatar dubbed 'Tinka' was piloted in Austria.

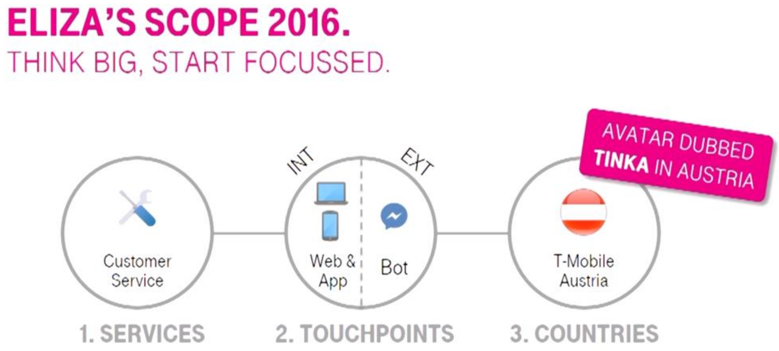


Figure 5.9 – eLIZA's 2016 Scope

Tinka's conversational interface featured multi-stage dialog, i.e. unlike a search engine that spits out an answer, Tinka would guide through dialog. With a responsive design, she ran on mobiles and large browsers. She could give step by step tutorials, with enriched media - video, gifs and buttons were all integrated into the chat flow. The team connected different pre-existing self-service touchpoints like forums, FAQs etc. so customers would have a one stop shop for help. Customer history was also integrated so that Tinka would remember a customer and his/her problems.

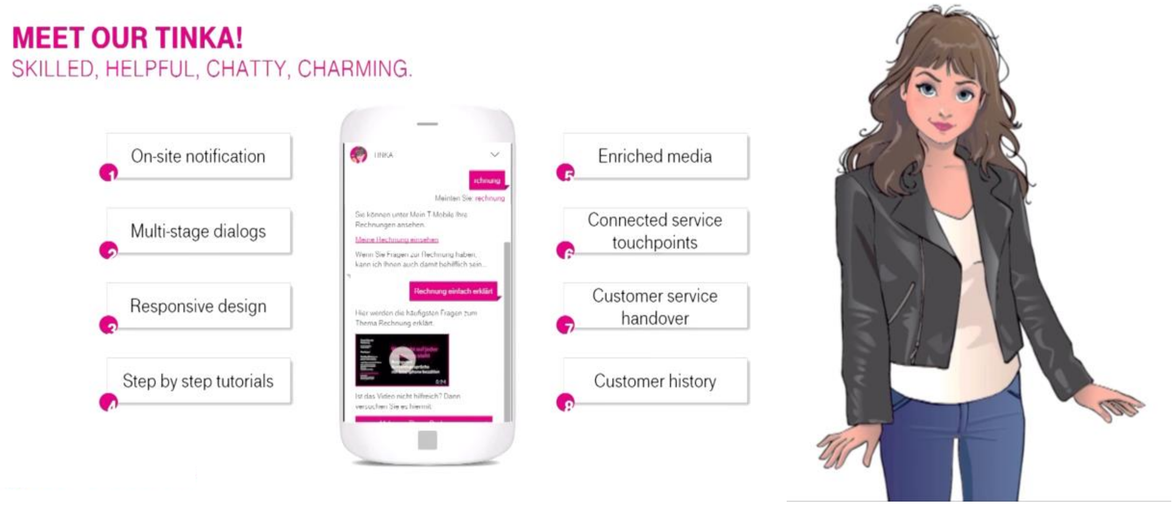


Figure 5.10 – Tinka’s Features

Going Where the Customer Is

Many companies usually want control over what their customers do with them and that can translate into having their digital touchpoints on their premises, but the team figured it had to go where its customers were. Facebook was a good start for that, and Tinka was integrated with Facebook Messenger. The team also began to think about other places where customers go, e.g. future integrations with Amazon Alexa and Cortana.

Learnings from Rule-Based Tinka

Tinka vastly improved the customer experience. She worked great as a solid one-stop shop for customer help. The team learned there was a lot of value in optimizing the user experience, even before you start to introduce Artificial Intelligence. Conversational interfaces are still new compared to websites and apps that have been optimized over many years, so there was a lot to learn there for the team.

Tinka also led to creating the tooling to enrich content on an ongoing basis. In a business environment where things are constantly changing (e.g. there could be new products/tariffs or feedback that a current version of a dialog is not working and needs to be changed), there needs to be an efficient way to feed new content into the virtual assistant. This cannot feel like coding, i.e. there needs to be a graphical user interface, as content is driven by non-technical customer service specialists.

But there were still shortcomings in the system. Handling complex queries was uncertain – they could maybe be handled if the customer put in the right keywords, but e.g. if a customer used keywords with negation like “I’m not interested in iPhone, I’m looking for a Sony Xperia phone”, that would be tricky for the rule-based system to understand.

Scaling content was also not simple. Yes, the tooling had been put in place, but it was still a manual process to feed new content.

As the team wanted to next launch Tinka in Germany, which, at 40 million customers, represented a big jump in system breadth compared to Austria, it began to think about taking Tinka to the next level by using Artificial Intelligence.

Sourcing an AI Vendor

AI represented a steep internal learning curve. DT in general uses a lot of partners, it is in fact one of its core focuses in its company strategy to win-win with partners. So the team decided to pick a partner and began to look at the landscape of AI solution providers and startups. It was overwhelming – the choice worldwide was huge. The team started with more than 75 companies, and then had a funnel to narrow these down. It make Proofs of Concept (PoCs) with real data with 4 of these companies, and finally there was one winner. That winner is now the new Natural Language Understanding (NLU) and Dialog Management Engine that is plugged into the stack, along with Speech to Text and Text to Speech components, effectively giving Tinka a new brain.

DT of course collects big data, some of it sourced from the customer journey, and the team is also working on making this available and useful for the digital assistant.

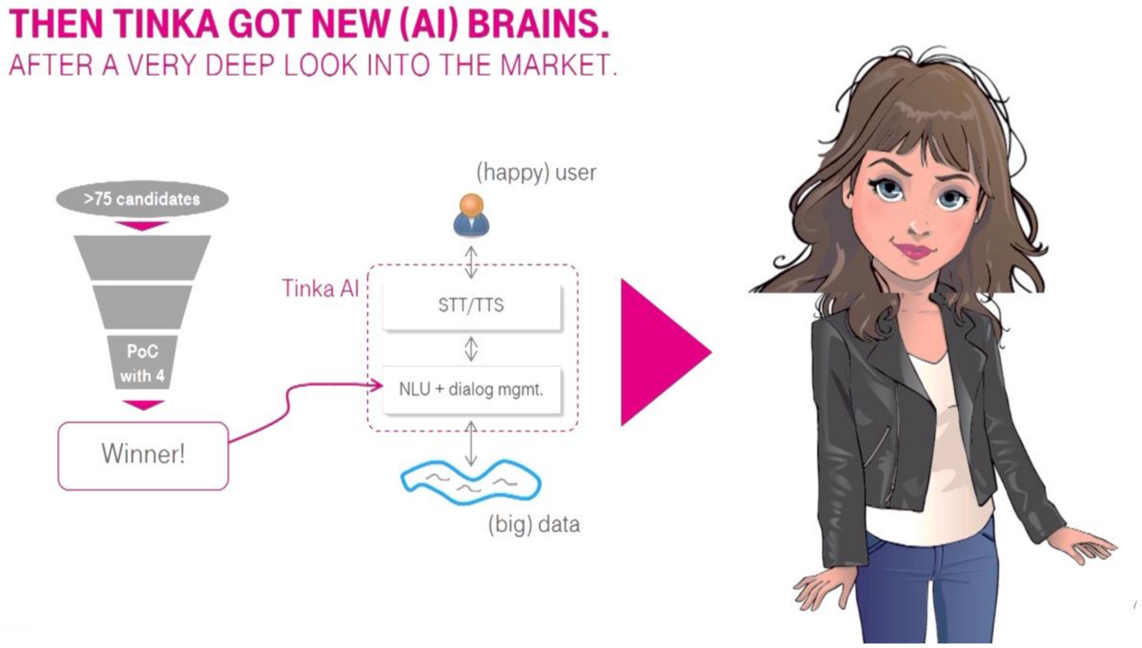


Figure 5.11 – Making Tinka Smarter

Results of AI-Based Tinka

AI-based Tinka had more human-like understanding and smarter dialog steering. For instance, in a human dialog people refer to things they said several steps before, or use pronouns – AI-based Tinka could deal with these and with many complex queries. She got great feedback. Tinka was chatting with more than 270 customers a day, handling 120,000 questions per month[53]. Tinka was able to handle around 80% of questions put to her[54], and where she could not answer she forwarded it on to a human agent.

Learnings from Introducing AI

The team learned that:

- **Comprehensive dialog understanding is key, but off the shelf it is hard to find:** While the speech to text and text to speech components were at this stage commodities, comprehensive dialog understanding , while available in research, was hard to find off the shelf as a product.
- **Everybody expects full on "self-learning" but the market is nowhere near:** Business expects AI to be "self-learning", that is you just plug in historic dialogs and the AI will learn from them, the reality is much less seamless than that.

[53] (Morgenthal, 2017)
[54] (Fulde, 2018)

- **Training data is plentiful but prep is tough – and which data to take?:** You would think training data would be plentiful given DT has tens of thousands of customer dialogs, but these need to be cleansed, anonymized and annotated, and have to be the right dialog for the right systems, so it takes a lot of effort to train the AI well.
- **Tooling should be the easy part but is in its infancy:** The assumption would be that tooling – like having an interface to plug in new questions and intents – would be the easy part, given that is standard programming. But even that is in its infancy in many systems.
- **Do detailed evaluations of vendors:** The team short listed vendors based on the complexity of their solutions, and whether they had the availability of data (e.g. if working in a specific domain, with a European language like German, the data might not be available with solution providers). Another thing to look at is transaction costs and the transparency thereof – e.g. the team found that some vendors themselves did not understand the transaction costs involved in building the solution and thus the team was getting bills weeks after the PoCs had finished.

Next Nuts to Crack for Higher Business Value

Jan Hoffman, leader of project eLIZA, reflected on the next nuts the larger AI ecosystem needed to crack for higher business value. He drew a 'back of the napkin' wish-list:

Figure 5.12 – Napkin for Next Nuts to Crack for Higher Business Value

- **Training Efficiency:** In a business context, you need to make sure the system you setup works over many years. Training a neural network efficiently is a major thing, and Jan wanted to see this enabled via e.g. automated annotation of raw data, and automated updates of intents in the intent classifier. There needs to be a closed learning loop, where you have your NLU + Dialog engine, you have intake of new data, automated cleansing, and that feeds into the engine to improve it. Currently this is not really existing.
- **Language Transfer:** DT has a number of subsidiaries in different parts of the world, and it would love to see a mechanism for more efficient language transfer – i.e. the ontology of the telco domain transfers into the next language easily.
- **Beyond Intent Classification:** Deep neural networks are mostly doing intent classification but true NLU is broader than this. The field needs more off the shelf dialog comprehension, and more attractive long tail answering ('long tail' queries are those unpopular individually but that in aggregate make a large proportion of traffic).
- **Open Systems:** This is key to business world adoption: no one party can do everything best, so there need to be APIs and open systems.

Critical Success Factors for DT's AI Journey

Miles Lynam-Smith, AI Chief Programme Owner, responsible for democratizing AI at DT's headquarters, reflected on the critical success factors in the AI journey taken by project eLIZA, which were essential in setting up DT for its next moves. Per Miles: "The next steps will be towards bringing AI from an innovation on the periphery of the company to an innovation in the centre – automation of processes will totally change the way we work".

Miles looked back at his early corridor conversations in the building about AI "..most people when they thought about AI, thought about Terminator 2, and losing their jobs, which as you can admit, isn't a great start".

So how did DT change this? First of all, it started by focusing on the customer problem (i.e. the poor customer care experience), and not the technology. Secondly, it defined a role for AI. AI is a big buzzword and can mean absolutely anything. At DT they defined it as something that can listen, learn, optimize and act, and they linked it to the larger corporate strategy imperative of 'Making our customers' lives easier' (a phrase engraved on walls at DT). In Miles's opinion, what the company is doing in AI needs to be linked to the larger company strategy, structure and vision that employees and shareholders have bought into, so that people can trust it, otherwise there will be a lot of questions/fear/sabotage.

Choosing the right partner is hard and thus Proofs of Concept are vital. It is important to be clear on the complexity of the solution, the data being offered by the solution providers, and the costs they are charging. At DT, the team tried to develop an AI ecosystem that is vendor agnostic and flexible, so that it can change and grow as new products come to the market place; thereby allowing DT to always offer the best solutions to its customers.

Lastly, DT strove to create an attitude in its large, distributed team to be the best team that it could be, by being the team that learns the fastest. The team has an external website at welove.ai, which is a good way of showcasing its work both to the rest of DT as well as prospective hires. Per Miles, "We try and have fun…we empower people, we let them go for it. We explain that they're part of one team, one ethic. And we allow people to go and innovate. That sometimes means go and fail, of course, but we try to term it as going and learning."

5.3 General Electric: Building Machine Learning Applications for the Industrial Internet[55]

About the Company

General Electric (GE), headquartered in Boston, Massachusetts, is a conglomerate company with products and services ranging from aircraft engines, power generation, and oil and gas production equipment to medical imaging, financing and industrial products. GE describes itself as a 'global digital industrial company, transforming industry with software-defined machines and solutions that are connected, responsive and predictive.'[56] In 2017, it served customers in over 180 countries, had over 313,000 employees worldwide, and revenues of $122 billion.

Consumer vs. Industrial Machine Learning

When we think of machine learning applications, often the first examples that pop into mind are from the consumer internet – Netflix's recommendation engine, Amazon's Alexa voice assistant, Google's Gmail Smart Reply etc. For these companies, data was already being aggregated and is relatively accessible.

Industrial machine learning is a field that is now emerging. Until recently the data was not available but we are now reaching a point where every single asset in the field is laced with sensors, collecting data on short timescales, and transmitting it to the cloud, enabling more advanced analytic approaches. This unlocks a number of use cases across a number of industries. By 2020 the industrial internet will have more than 50 billion connected machines.

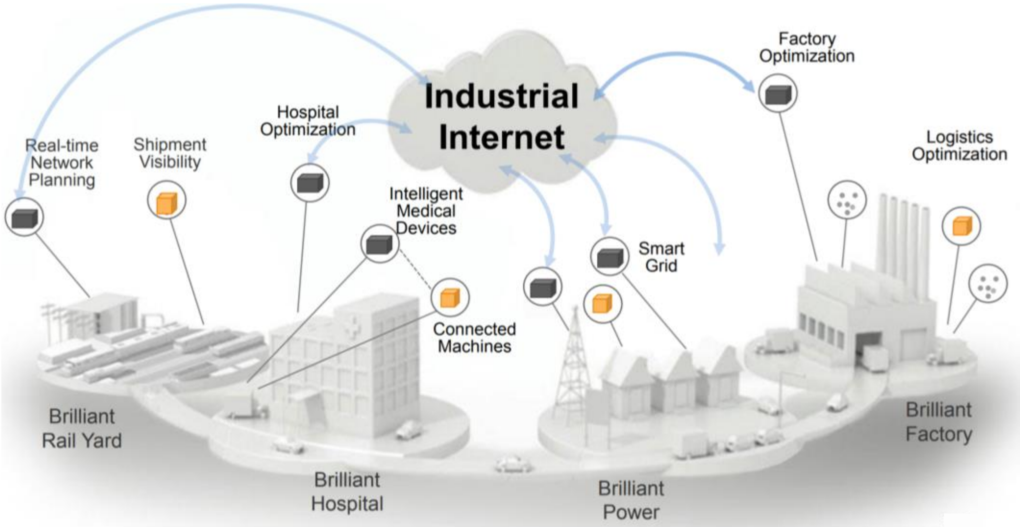


Figure 5.13 – Industrial Internet of Things – Connected Assets Will Enable Use Cases Across Many Industries

This is an area of tremendous opportunity because even a 1% gain in efficiency can translate into tens of billions of dollars saved within GE's business units, as shown in the figure below.

[55] Except where otherwise explicitly cited, all content in this case study, including all figures and tables, is sourced from this conference video: (Richards, 2018)
[56] ("GE 2017 Annual Report," 2018)

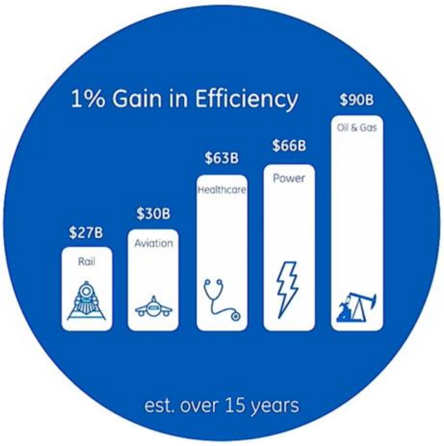


Figure 5.14 – What's at Stake for the Industrial Internet of Things

At the same time, this is also an area with a number of challenges as compared to the consumer internet, as summarized in the table below. Firstly, the Industrial Internet has much more data to manage – e.g. a day's worth of Twitter data is 500 GB but a single flight's data is double that at 1 TB. Connectivity is also problematic as these assets are in the field – e.g. a sensor could be in an aircraft engine – so getting connectivity at timescales that enable support is a challenge. Assets like engines and turbines are meant to be in operation for decades, so the sensors that are on it also need to work reliably for years/decades. Security is also far more necessary and many of these assets are 24/7 mission critical, e.g. in power, aviation and health care, and thus must be hack-proof. Finally, unlike the consumer internet, privacy in these industries is highly regulated.

	Consumer Internet	**Industrial Internet**
Data Management	Day's worth of Twitter: 500 GB	Single flight: 1 TB
Connectivity	Biggest cell phone complaint: dropped calls	Mission critical, rough & remote
Device Support	Avg. wearables lifetime: 6 months	Lifetime of a Turbine: 20+ years
Security	Time to hack most devices: minutes	24/7 Mission Critical
Privacy	Privacy is no longer a 'social norm' – Mark Zuckerberg	HIPAA, ITAR, …

Table 5.4 – Challenges for Industrial Internet vs. Consumer Internet

Considering all these challenges, it is not too surprising that most well-known machine learning applications happen to be on the consumer side, given the data there is easier to collect, manage and share. Just getting the infrastructure set up in the industrial internet to collect data at scale, aggregate it, send it to the cloud, and have people securely access it has been a complicated and time consuming process. However, GE made substantial progress on that front and was finally at the moment when the data infrastructure to enable machine learning applications was in place.

Acquiring Wise.io To Build Machine Learning Applications Across GE's Business Units

In the fourth quarter of 2016, GE acquired Wise.io to build and deploy machine learning applications for GE and its customers across all of GE's business units. Wise.io was a startup that had built and deployed more than 100 machine learning applications for its customers and that had a platform for deploying machine learning applications in production at scale. The Wise team sought to brought that mentality to GE – i.e. how to tackle common problems at a much larger scale, with an impact across millions of assets, in a way that is repeatable, scalable and unlocks real value.

First Example: GE Aviation

About GE Aviation

GE makes 60% of the world's airplane engines. It has more than 33,000 engines in service worldwide, and each of those engines has 50-100 sensors on it, recording several times per flight, which will soon be upgraded to recording once per second (i.e. at 1 Hz). GE manufactures these engines and sells them to airlines, and also services these engines.

The Problem Statement

GE's Aviation Fleet Monitor process, which was built and refined over decades, was also very labor intensive, manual, and suboptimal (not in terms of engine parts failing, but in that GE tended to overmaintain/service when not needed).

The Fleet Monitor process supports GE's global engine fleet by capturing sensor data from engines multiple times per flight. This data is aggregated into the cloud, physics based models run on the data to identify anomalies and create alerts that get surfaced to the global Fleet Monitor team. This team has deep domain experts who then make a decision about each alert, and decide when to issue 'Customer Notification reports' to airlines when necessary and when to dispatch technicians on-site to perform inspections and repairs.

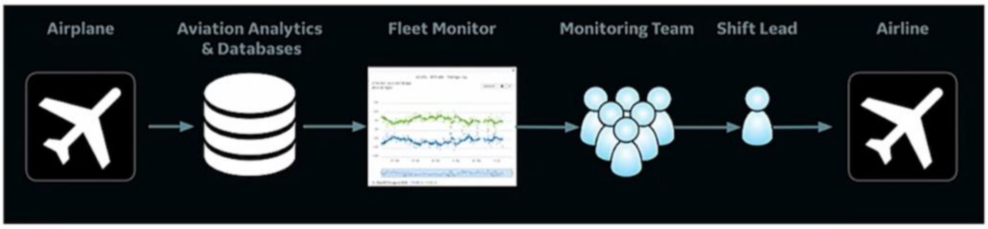


Figure 5.15 – Fleet Monitor Process at GE Aviation

The bottleneck in this process was for the experts to review the alert and all the data associated with it, and to potentially look up in the knowledge base data about similar alerts. This was very time consuming, required many people hours and created delays.

Introducing Machine Learning

The Wise.io team aimed to inject machine learning into this workflow in a way that was not disruptive. The team took all the historical alerts and outcomes of those alerts, across the entire fleet, and used machine learning to learn the patterns that are associated with the actual outcome of each alert.

The machine learning application then served recommendations to the Fleet Monitor team, using machine learning to point them in a direction (serving both suggestions and confidence levels in those suggestions), and empowering the team with a tool to make their lives easier.

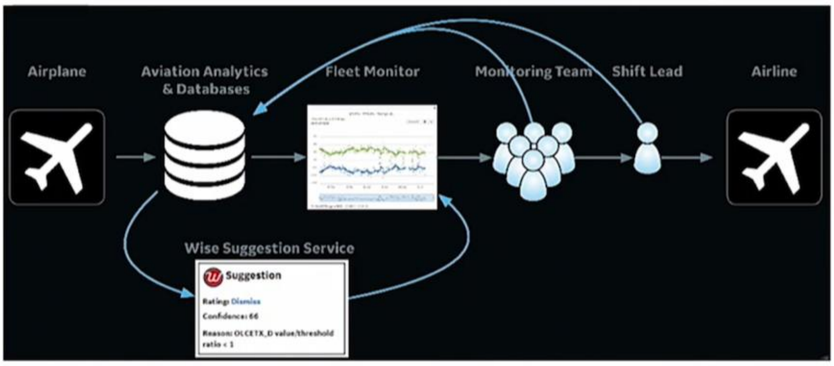


Figure 5.16 – Injecting Machine Learning Suggestions into the Fleet Monitor Process

The Fleet Monitor engineers enjoyed working with the application. It decreased the time to case valid alerts, and it freed up time for the engineers to spend on more high value work, like communicating with the airline company, doing more proactive maintenance etc. It also drove higher consistency in the monitoring process.

Extrapolating to Other Business Units

What the team had improved was in fact a typical problem across many of GE's business units. There were many time consuming, repetitive processes where experts (analysts, engineers) looked at data and reference knowledge bases to make decisions. The team built similar machine learning applications for GE Power, and BHGE (Baker Hughes GE) Oil and Gas, as discussed in the examples below.

Second Example: GE Power

About GE Power

GE technology delivers 1/3rd of the world's electricity, and manufactures e.g. gas and steam turbines, windfarms, solar solutions etc. In power the amount of data gathered is several orders of magnitude larger than in aviation. For each power plant, GE captures time series from 3,000 to 10,000 sensors, each collecting data at 1 Hz. GE builds, monitors and services the equipment.

The Problem Statement

Just like in aviation, GE Power has similar workflows for its Monitoring & Diagnostics (M&D) team where alerts come in from the field, the team looks at data to support the alert and makes a decision and communicates that to the end customer.

Introducing Machine Learning

Machine learning was inserted in the GE Power M&D team workflow, similar to how it was inserted in the GE Aviation Fleet Monitor workflow, with the slight change that it automated some of the case creation, i.e. inching the way forward to humans out of the loop. This empowered the engineers to spend less of their time on alerts that were not very risk/consequential.

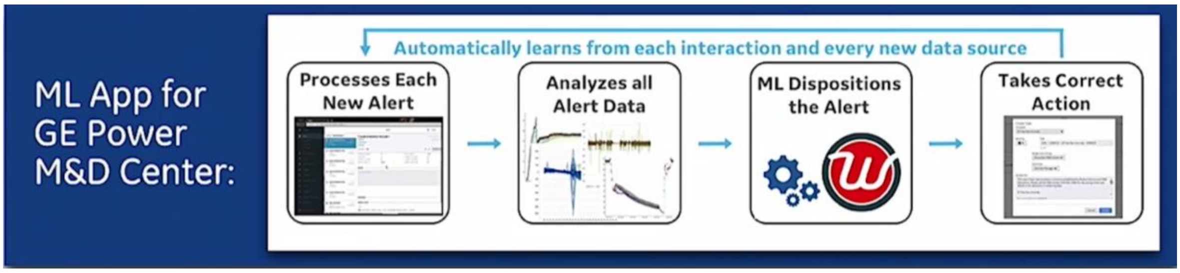


Figure 5.17 – Machine Learning Application for GE Power Monitoring & Diagnostics Center

Third Example: BHGE Oil & Gas

About BHGE Oil & Gas

Baker Hughes GE Oil & Gas delivers products and services across the full-stream oil and gas value chain. GE Pipeline Solutions within it is in charge of inspecting oil and gas pipes worldwide. These are imaged by ultrasonic or magnetic sensors. GE Pipeline Solutions has inspected more than 1 million miles of pipeline, each pipeline is scanned at 3 mm granularity. Over the past decade over 120 million at risk areas have been identified in 18,000 customer reports, with each inspection comprising around 1 TB of data.

The Problem Statement

The workflows mimic those seen in aviation and power, i.e. data is collected, uploaded to the cloud, and served to an army of people sitting and scanning through hundreds of kilometres of pipe to identify defects. Each inspection generates data at 3 mm granularity captured by 500+ sensors for 100+ kilometres pipeline. 350 trained analysts (it takes around 2 years to train them) search for anomalies. It takes months to deliver an inspection report to a customer.

There are not enough of these highly trained analysts, and it takes far too long to take the data, process it and make the customer report. There is also too much noise in the pipeline data hindering them from doing their job quickly and effectively.

Introducing Machine Learning

The machine learning application was able to reduce manual work by 20%. It automatically dismissed regions in the analysis pipeline that were not likely to contain a reportable defect, enabling the analysts to spend more time on likely regions. It preserved all reportable anomalies, automatically learned from analyst feedback and ensured high quality inspections.

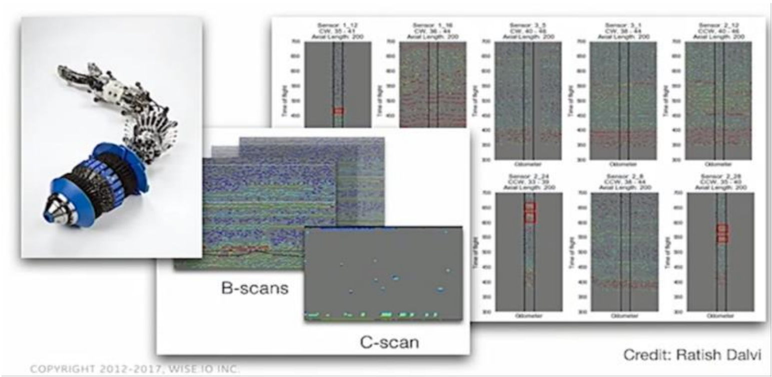


Figure 5.18 – Wise Pipeline Inspection Application

To get to production grade on a use case like this required back testing on all the historic data (i.e. hundreds of historical inspections), to demonstrate that the algorithm never missed something that could be an injurious defect.

Impact of Machine Learning Decisions in Consumer vs Industrial Internet

That leads us to drawing the important distinction between the impact of machine learning decisions in the consumer vs. the industrial internet. The confusion matrix in the figure below illustrates this.

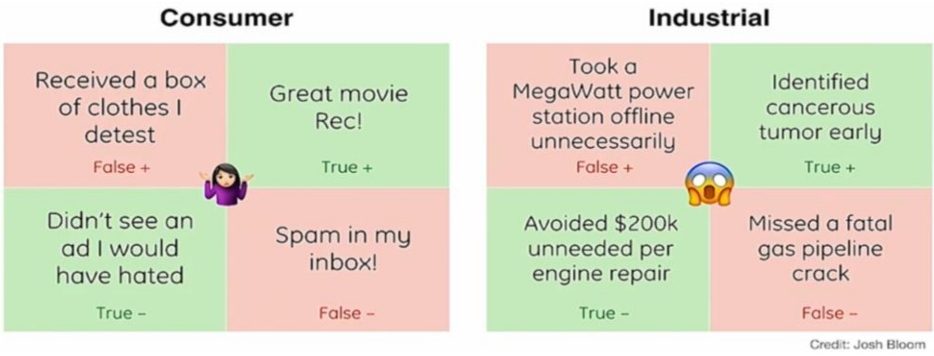


Figure 5.19 – Impact of Machine Learning Decisions in Consumer vs. Industrial Internet

On the consumer side, a true positive would be receiving a great movie recommendation, and a false positive would be receiving a box of clothes one detests. On the industrial side there are far higher stakes: a true positive would be identifying a cancerous tumour early, a false positive might have one taking a power station offline causing thousands of dollars of disruption. On the consumer side, a true negative would be not being shown an ad one would have hated; a false negative may mean spam in one's inbox. On the industrial side, a true negative may mean one has avoided unnecessarily taking an engine off a plane and grounding it. A false negative meanwhile is the most dangerous in the industrial world – it could mean e.g. one has missed a gas pipeline crack that could lead to a fatal explosion.

Rigorous, Repeatable Processes for High Quality Machine Learning Systems

Given the high stakes impact of machine learning decisions in the industrial internet world, it was important to ensure production grade, high quality machine learning systems whose quality was on par with GE's standards of excellence. The Wise team focused on creating that mindset around production grade machine learning. They created a platform and approach to ensure machine learning applications created worked well in production (and not just as offline prototypes) with minimal need for customization/hand holding to ensure they work well. The Wise team put in place a rigorous, repeatable process for the design, development, testing, evaluation, deployment and maintenance of machine learning applications, enabled by a full-stack ML application platform.

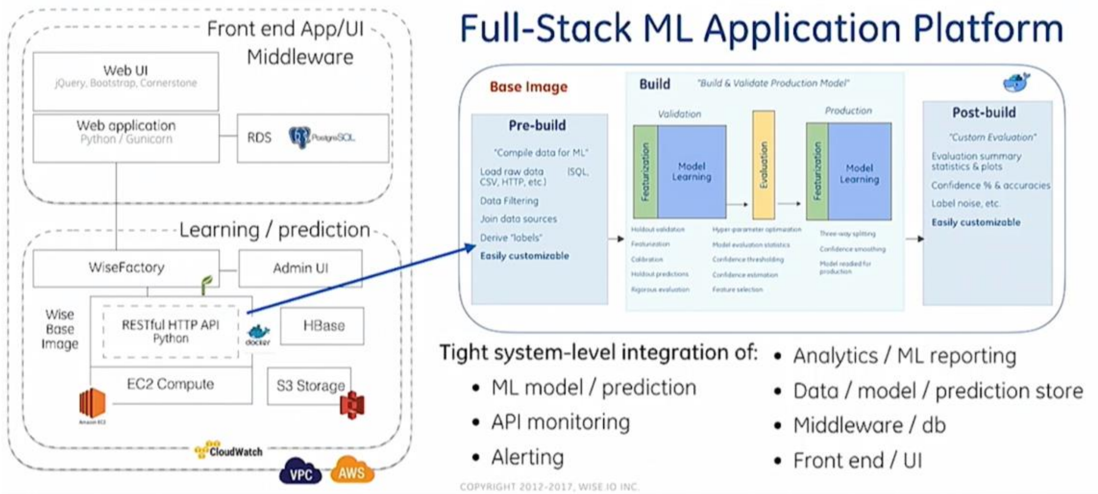


Figure 5.20 – Full-Stack ML Application Platform

The platform featured tight system-level integration of all the components one would need in an ML system in production – so it catered for how you serve models; do retraining; monitor API endpoints; do the alerting; reporting; store the data, model and predictions; customize middleware to interact with external systems; and the user interface/custom user interfaces to be built. Having all these components encapsulated in a system made it easy to build use case templates that could be deployed horizontally/customized across similar use cases across industries.

Approach for Building Valuable ML Applications

To sum up, the Wise's team approach for building valuable ML applications at GE included:

- **Enhancing and improving workflows** with ML working behind the scenes. This involved having a deep understanding of the existing battle-tested workflows, as well as how ML works, and putting those together in a non-disruptive way.
- **Building ML applications for domain users**, i.e. the "people looking at data"
- Having robust methodology for **deciding when to augment or automate** manual processes, assessing risk, confidence levels and various trade-offs. Less risky decisions e.g. can be automated.
- **Working alongside of, not replacing, the domain specific systems and applications** the domain experts were already using. GE already had detailed physics based models of how equipment should be working in an idealized environment – the team needed to figure out how to leverage the output of these physics based models and have the machine learning based models reside side by side with them.
- Capturing and using **ongoing feedback** from users' normal interactions. In these systems the number of events is small, e.g. there are only dozens of alerts a day and it is not like Amazon with millions of interactions a day, so it is very critical that there is continual learning and that the system captures and uses feedback to improve.
- Creating **machine learning application templates** that scale. The team is not creating one off applications but embedding in the system in a way that scales cross industry and to adjacent use cases.
- **Empowering the other data scientists** to build and deploy full stack ML applications. It is relatively easy to make prototypes, it is much harder to get those prototypes working at scale in production.

5.4 General Mills: Automating Business Insights for Marketers through Artificial Intelligence[57]

About the Company

General Mills (GM) is a US based global consumer foods company with $15.6 billion in net sales in 2017[58]. 75% of its sales came from five categories – cereal, snacks, yogurt, convenient meals and super-premium ice-cream. The company manufactures and markets over one hundred well-known brands, such as Cheerios, Häagen-Dazs, Yoplait, Pillsbury, Betty Crocker and many more[59], sold through retail stores.

The Problem Statement

For General Mills, the problem was **how to help over 1,000 people in marketing with data analytics**. All these people should be using data in their day to day workflows, but they were hampered by the large number of different data sources, none of which talked to each other. The vast majority of data in marketing is data that GM does not generate, but buys/obtains from different sources, as it is a food manufacturer that does not sell direct to consumers. Instead it sells to retailers who then sell to consumers. There are also many data silo owners. These owners have built their careers on the data and are the experts on it. They may not be looking for it to be available to everybody, and typically do not have data democracy in their vocabulary.

All these factors were preventing analytics to be built across the data.

Building a Dedicated Decision Sciences Team in Marketing

GM went ahead with creating a dedicated decision sciences team in marketing, staffing it with 3 types of roles:

Data Stewards
Ensured quality of data and maintained connections

Data Visualizers
Developed easy way to use the data

Analytics & Data Science
Created custom built analytics
Engineered new data sources and access

Figure 5.21 - Decision Sciences Team at GM

The **Data Stewards** were responsible for stitching the disparate sources of data and establishing a common nomenclature – for instance, one data mart needed the stitching together of 47 different data sources! They also had to make sure these remained connected. **Data Visualizers** developed visualizations that made the data easier to use and work with, while the **Analytics and Data Science** function primarily created custom built analytics.

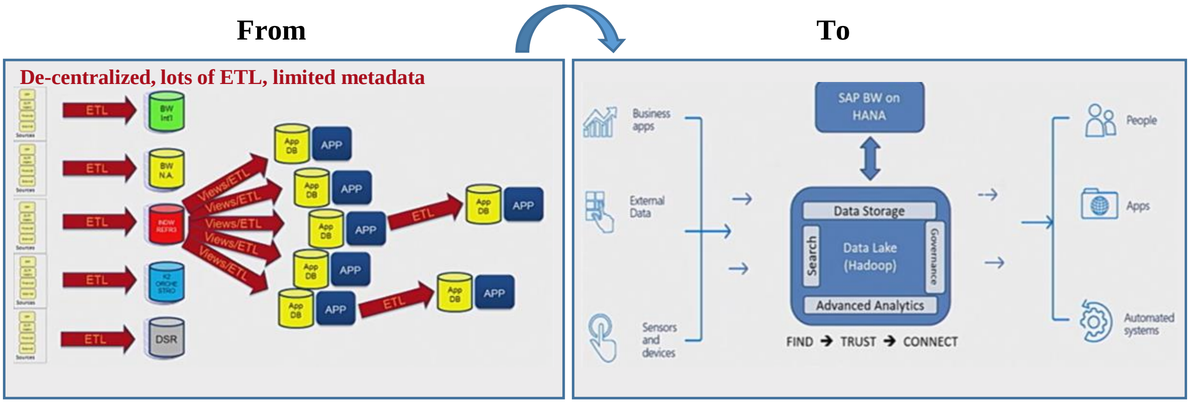
Partnering with IT to Create a Data Lake

[57] Except where otherwise explicitly cited, all content in this case study, including all figures and tables, is sourced from this conference video: (Fleener, 2018)
[58] ("General Mills, Inc. 2017 Annual Report," 2018)
[59] ("General Mills: Brands overview," n.d.)

The decision sciences team partnered with IT to create a data lake, going from a landscape of many decentralized independent data warehouses, to a data lake where all the data was housed together, enabling:

- not moving data around when it was needed
- finding what was needed; supported through data governance
- having analytic capabilities reside in the same place as the data
- storing unstructured data

From **To**



Multiple Platforms: Oracle, Exadata, SQL Server, SAP HANA, SAP BW Etc.

Figure 5.22 – Moving to a Data Lake

Creating Impact

All told this was a five year journey – the first three years of which were mainly just about getting the data organized. During this time, building visualizations that brought access to data was an important way of showing some early wins and gaining buy-in. Over the last two years there was more data science work being done. The journey created meaningful impact along the way in the form of:

Dimension	Impact
Data Availability The team enabled access to data people could not access before, and provided documentation to educate.	**New possibilities on what can be asked**
Data Connectivity When the data is stitched together it speaks to each other. The data is aligned the way marketers think.	**Reduces time to prepare analysis**
New Insights Purpose built visuals answer common business questions, with analytics to answer advanced questions	**Reduces the time to answer or enables the ability to answer questions**

Table 5.5 – Creating Impact at General Mills

Deep Dive into A Specific Analytics Solution and its Evolution

Let's take a deeper look at one of the specific analytics solutions provided by the decision sciences team and how it evolved. This was the **New Product Dashboard**. In the US, GM launches several hundred new products each year, such as a new cereal line, granola bar etc. In consumer packaged goods, it is a well-established rule of thumb that around 80% of new products are deemed a failure within two years. That is a lot of launches then that are not successful, and a number of measures along the way are used to judge whether they are successful or not.

Build a Dashboard and They Will Come (…Not!)

At General Mills the decision sciences team estimated around 40 people in marketing were spending 25 hours a month trying to gauge new product success by looking at measures on their own. So the team decided to automate that through the new Product Dashboard.

- *50+ charts, 120 metrics all important*
- *15-20 minutes for a user to see what is going on with each new item*
- *Data warrants watching it weekly*
- *A marketer can have 5-8 new items at any given time*

Figure 5.23 – New Product Dashboard

The New Product Dashboard consolidated a wealth of information in 50+ charts – but it took 15-20 minutes to go through the visualizations for each new item and see that everything was on track. The data warranted watching it weekly, with marketers having 5-8 items at any time. The trouble was that marketers were extremely busy (see figure below for a typical schedule of a marketing manager). So while the dashboard had some users, most marketers were not using it, simply not having the time to do so.

Jane,
Marketing Manager
on Nature Valley

Time	Priority
8am	Finish what they did not yesterday
9am	Prepping for a retailer conversation
10am	Got an urgent request from leadership
11am	Still scrambling to get an answer
12	Talk to media buying group about an ad in flight
12:55	Try to grab something to eat
1	Meet with Finance
2	Work with product development on a future flavour
3	Discuss what to do about plant restrictions
4	Finalize an answer to the urgent request from leadership
5	Run and grab the kids
9pm….	Answer all the emails from the day

Repeat

Figure 5.24 – Typical Schedule of a Marketing Manager

Time for a Rethink

The team had to think differently. How could it cut through the clutter and identify what products had something unique going on? How could it quickly guide the marketer to what was unique about the product and increase the speed to action. Out of the 50+ visualizations, what were the 3 most important things the marketer should be paying attention to?

Adopting a Machine Learning Approach to Anomaly Detection

The team decided to look for various types of anomalies, using a machine learning approach. As they did not have a training set identifying anomalies, they went with an unsupervised approach, using a number of algorithms. Then they created a master algorithm that combined all these to quickly identify anomalous products and anomalous metrics.

Types of Anomalies
Point: Individual data point can be considered anomalous with respect to the rest of the data
Collective / Contextual: instance is anomalous in a specific context (but not otherwise), then it is termed as a contextual (conditional) outlier. If group of points, then termed collective outlier.
Global: Anomalous data points are defined by measuring the global deviation of a given data point with respect to its neighbors, globally.
Local: Anomalous data points are defined by measuring the local deviation of a given data point with respect to its neighbors.
Business Transactions: Anomalous data points related to some business transaction often measured via a KPI ($s or volume) over time (think time series)
Reference Data: Anomalous data points related to the metadata or reference data about some entity

Table 5.6 – Types of Anomalies

Approach	Example Algorithms	Considerations
Rule-Based	zscore, frequency, hard-coded rules	• Must be able to define abnormal • **Must define rules** • Does not adapt to change
Supervised	Classification, Logistic Regression	• Need data labeled as abnormal • **No need to define rules** • Adapts to change
Unsupervised	Clustering (KNN, DBSCAN), Isolation Forest, Deep Learning Autoencoder	• **No need to have labeled data** • **No need to define rules** • Adapts to change

	Anomaly Type					
Algorithm	Point	Collective	Global	Local	Time Series	Reference Data
HDBSCAN		x				x
K-NN			x			x
Isolation Forest	x	x				x
LOF				x		x
Rule-based (Z-scores)		x				
GESD			x		x	
PR		x			x	
LOESS		x			x	
RQR	x				x	
RSD	x				x	

Figure 5.25 – Applying Unsupervised Machine Learning to Anomaly Detection

The algorithms were deployed in a visualization, to make it easier to digest. A business user could log into the visual and see what launches were anomalous, they could see what the anomaly event dates were, and then they could drill down into specific business metrics.

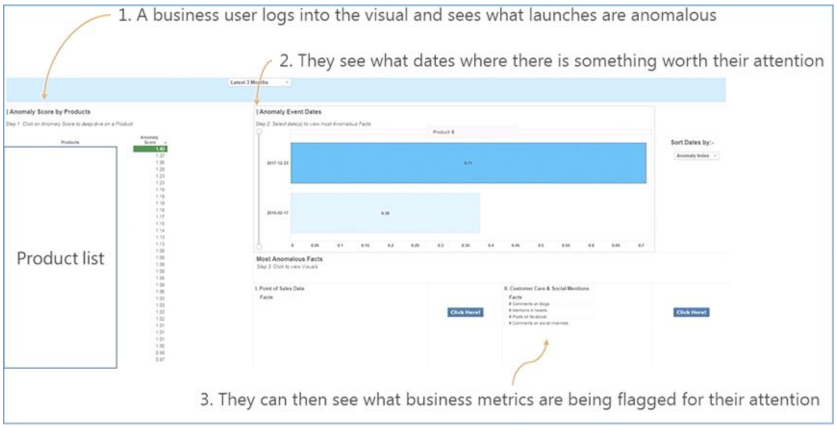


Figure 5.26 – Deploying Machine Learning Algorithms in a Visualization

A Tale of Two Products

The algorithms worked great. They were able to pick up anomalous products 6 weeks into launch whereas normally it takes a year. Consider the case of two limited run products A and B, both similar but different flavors, and both detected as anomalous based on people calling the call centre about it at different rates, and based on the volume being sold. For Product B, a number of people were calling in and requesting that it become a permanent product; for Product A, a number of people were calling in and asking if they could exchange it. The difference in sentiment was corroborated by anomalous volume events.

So if Product B was a clear winner, GM's marketers should have said let's keep B very early on. They should have started having discussions with retailers about keeping it, they should have alerted manufacturing to keep sourcing it, and decided on a strategy to keep B as a permanent product.

Could Have, Should Have, Would Have

Unfortunately none of the above transpired. Even though the decision sciences team had made it much easier to quickly go through the visual, it still required a business user to proactively go and look at it. This was competing with all their priorities and still not simple enough.

Meet the Business Users Where They Are

The decision sciences team realized it had to bring the automation of insights to where the business users already were. Despite management support, in a large organization, changing entrenched patterns of behaviour was a big endeavour. So finally, what did work at GM was sending prescriptive email alerts to the business users. Up to now the decision sciences team had not built the right user experience and thus was not getting adoption. The team realized it had to treat the business user experience like a consumer experience. The decision sciences team trying to help business users become smarter and take action off data is an ongoing journey, where the user experience is just as important as the insight.

5.5 Kaiser Permanente: Using Artificial Intelligence to Improve Patient Flow Forecasting[60]

About the Company

Kaiser Permanente (KP), founded in 1945 and based in Oakland, California, is one of America's leading health care providers and not-for-profit health plans. Its mission is to "provide high-quality, affordable health care services and to improve the health of our members and the communities we serve."[61] In 2017, it had 11.8 million members, 211 thousand employees, $72.7 billion in operating revenue, 39 hospitals and 682 clinics and other facilities.

The Problem Statement

Hospitals today face numerous challenges that are straining their existing bed and service capacity and driving the need for improved patient flow management. These challenges include increased demand for services, clinical staff shortages, lack of tools and technology to adequately measure and manage patient flow, the risk of patient deterioration due to prolonged hospital stays, and fewer available beds.

With the continued aging of the US population and accelerated clinical technology advances, demand for inpatient bed capacity is projected to rise by nearly 4-5% every year. The supply side is not keeping up, which means hospitals need to become more intelligent in how they manage this demand.

In order to solve this problem, traditionally hospitals have used rudimentary forecasting methods that are very top down – they identify how many patients are flowing through different channels and look at time series and historical trends.

Objective

KP partnered with the vendor Pacific AI to optimize patient flow models and provide insights for real-time decision making and for strategic planning by predicting:

- **Bed demand** – could KP predict bed demand on an hourly basis, or real time?
- **Safe staffing levels** – every medical facility has to maintain safe staffing levels to be able to serve patients
- **Hospital gridlock** – there are times when demand cannot be met by current capacity – when hospitals enter a 'gridlock' situation it can take a lot of time, effort and money to get out of it. Could the model provide insight ahead of time that a gridlock situation might occur?

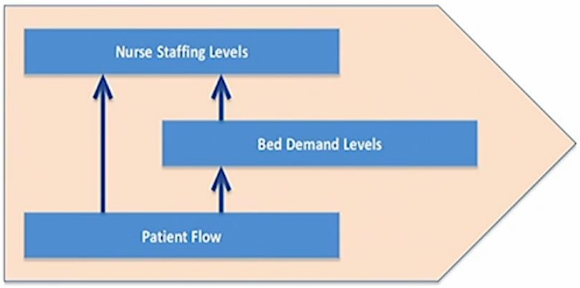


Figure 5.27 – Optimizing Patient Flow Models for Real-time Predictions

---

[60] Except where otherwise explicitly cited, all content in this case study, including all figures and tables, is sourced from this conference video: (Talby & Kulkarni, 2018)
[61] ("Kaiser Permanente 2017 Annual Report," 2018)

Looking Into Key Factors that Influence a Patient's Flow

The team started looking into the key factors that influence patients' flow, determined through answering questions like how likely are they to be admitted? For how long? For what?

A key factor was volume of arrivals from different ports of entry – within which Emergency Department (ED) arrivals was the most tricky as anybody could come through the doors. Other factors included the time of arrival (what hour/day/whether holiday), the admission speciality (e.g. oncology, cardiology) etc. and various other factors shown in the figure below.

When the team tried to get this data, it discovered that some of the most relevant factors (such as acuity level of the patient, pain, ongoing treatment – circled in red in the figure below) were only available within free-text clinical notes. These notes are highly dependent on the ability and willingness of the person who is documenting the notes; furthermore, standards and nomenclature for these notes vary from facility to facility. They feature various abbreviations and non-standard vocabulary.

So the challenge the team faced was how to unlock value from these notes, which is a Natural Language Processing (NLP) problem.

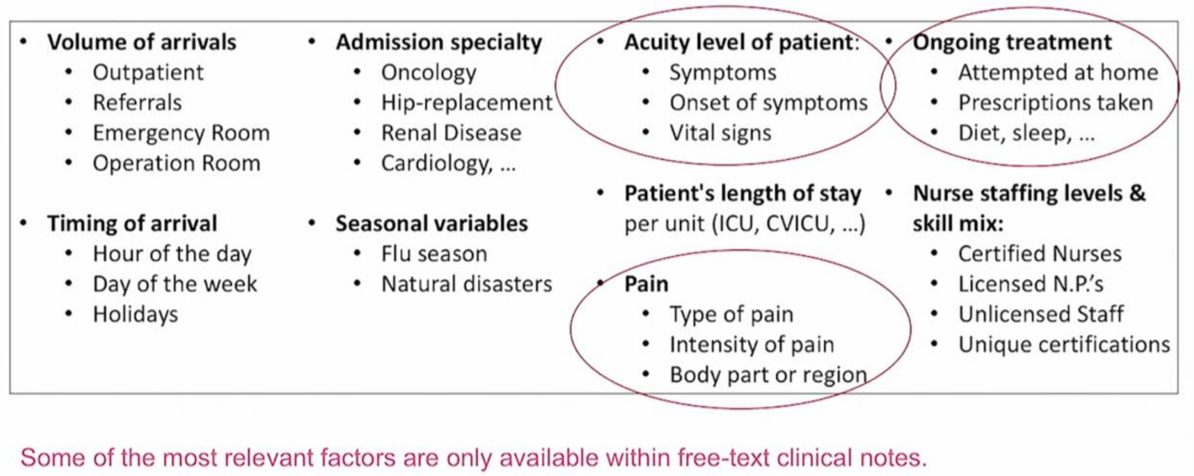


Figure 5.28 – Optimizing Patient Flow Models for Real-time Predictions

NLP is Just a Small Part of Building an NLP AI Solution

This initiative was not about building a research project that shows feasibility, but rather about having a system that goes into production in a real hospital setting, that can scale to multiple hospitals, and that can deal with the real data quality issues that are out there.

As the following figure from a paper published by Google in 2015 shows, in a real-world machine learning (ML) system, only a tiny fraction of the code is being leveraged to do prediction and is your actual ML code. The vast majority is essentially surrounding plumbing, the complexity of which represents a huge technical debt that needs to be paid down over time. This requires organizational commitment and a shift in culture. Deploying an ML system can be easy and cheap – it is maintaining it that is very expensive.

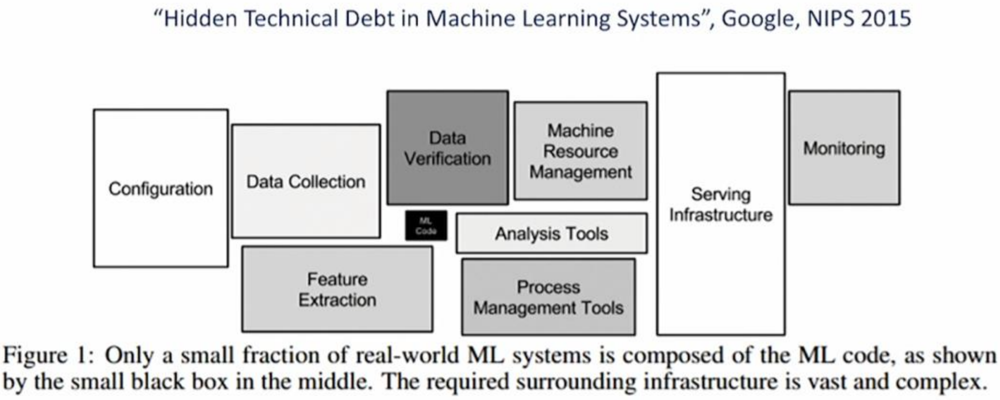


Figure 5.29 – ML Code is only Small Fraction of Real-World ML System[62]

KP adopted the concept of a data factory, i.e. a platform that takes a pipeline perspective, connecting producers of data to analysts and consumers. Data owners, engineers, analysts, scientists, and data operations people all have their roles in the pipeline. The platform addresses all questions from how to get the data to how to productionize the models and addresses many of the challenges related to technical debt.

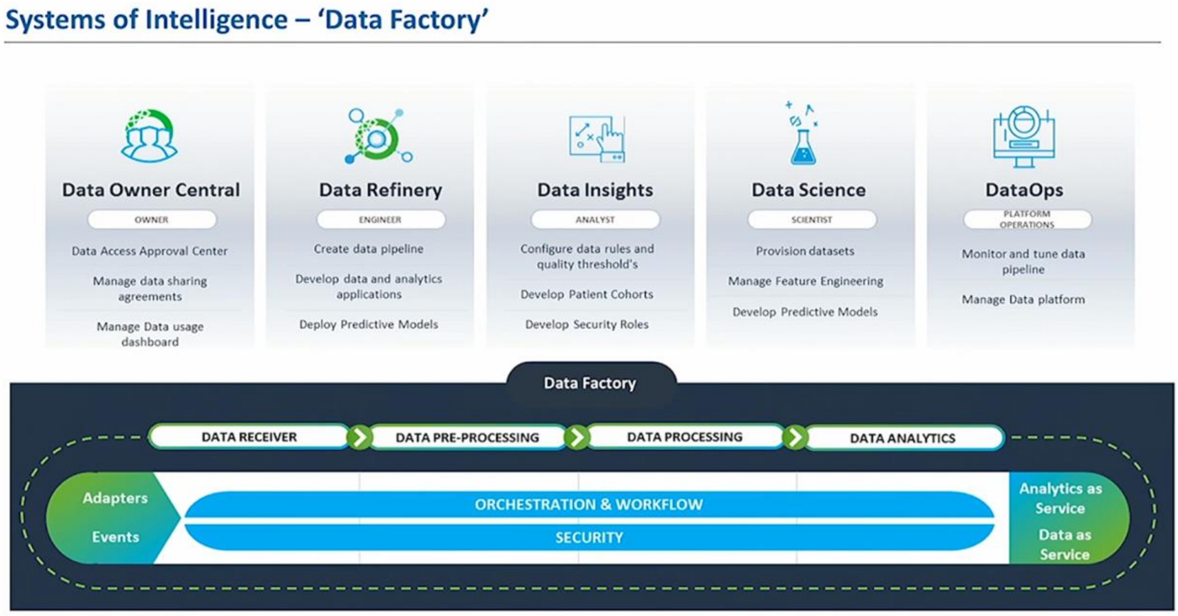


Figure 5.30 – KP's Data Factory

If we zoom into the 'Data Science' box in the figure above, it can be expanded as in the next figure. The data science platform has a set of **content packs** that come in, and get curated and updated by clinical experts so they can be used as a reference. It includes the capability to do **interactive analysis and visualization** without coding. It also has the capability to build and train **models**, do feature engineering and run **experiments**. It also manages how to **send models to production**, and builds in **scaling, security and monitoring**. At KP, the team could not e.g. just install Jupyter Notebook and start working with it – the main challenge was building everything around it, i.e. getting to a point where a community of data scientists at KP could use the whole platform appropriately and be productive.

[62] (Sculley et al., 2015)

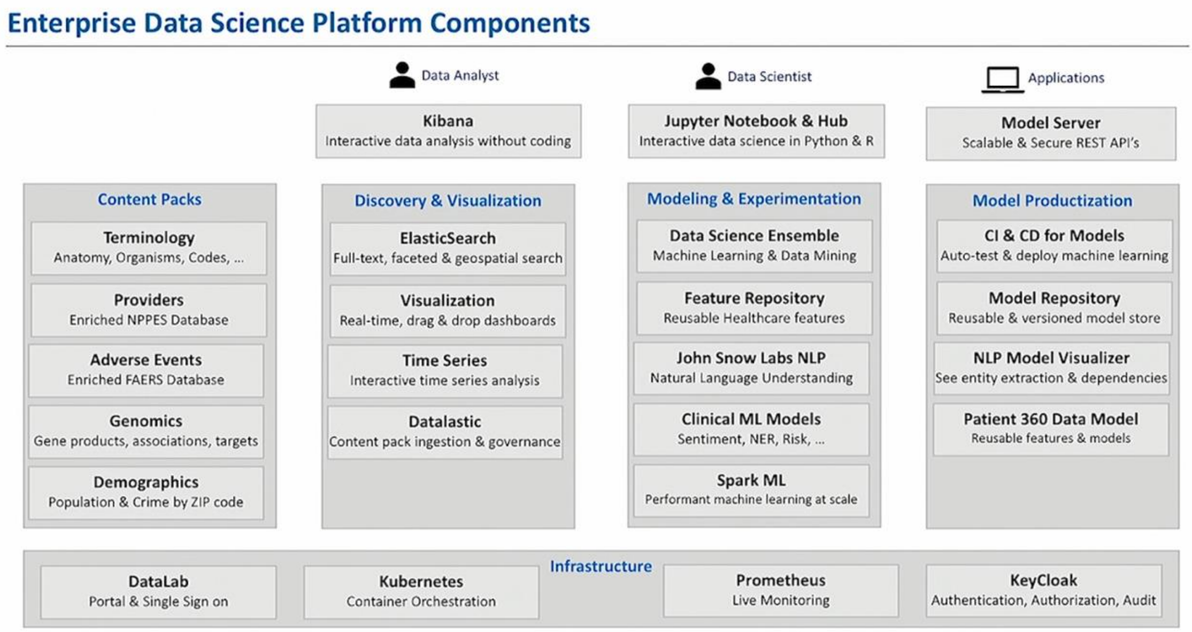


Figure 5.31 – KP's Data Science Platform

It was a lot of work to setup the entire platform, but it was necessary to build the below key capabilities (bolded ones specific to this healthcare challenge).

Enterprise Scale and Enterprise Grade Capabilities
Machine learning, data mining & deep learning on unstructured natural language
Out-of-the-box, reusable, healthcare-specific models & datasets
Continually updated, clean, linked & enriched content packs
High productivity toolset for data scientists working in programming languages like Python or R
Cutting-edge algorithms for a broad variety of data science problems
Self-service data discovery, visualization & analysis without coding
Productive machine learning models quickly, at enterprise-grade scale & reliability
Tools supporting best practices for validating, versioning, sharing & reusing models
Seamless integration with big data platforms, using Spark like execution engines

Table 5.7– Enterprise Scale and Enterprise Grade Capabilities of KP's Data Science Platform

The NLP Problem

Let's get back though to the actual NLP problem. Natural language understanding is hard as language can be nuanced, fuzzy, contextual, medium specific and domain specific. Take a look at the below 3 examples of Emergency Department triage notes – none of them even mention the word 'patient', or 'pain', but this is the type of language the system had to deal with and extract information like what are the symptoms, when did they start, what is the type and level of pain, where is it happening, and what did the patient try at home.

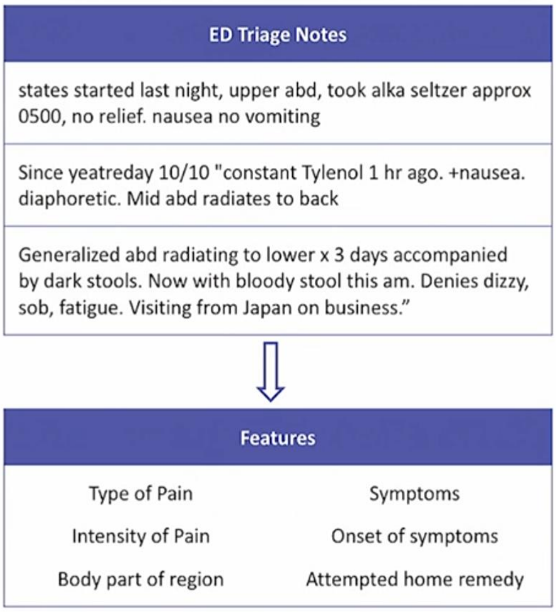


Figure 5.32 – Extracting Features from Emergency Department Triage Notes

Solving this problem required adding two healthcare specific components sourced from John Snow Labs:

- Data: 300+ expert curated, clean, linked, enriched & up to date datasets (covering things like terminology, clinical guidelines, measures etc.)
- Custom Algorithms: Health specific NLP annotators (e.g. doing entity recognition, word embeddings, sentiment analysis etc.)

Building in State of the Art Performance

KP wanted to claim state of the art performance for this problem. If you want to do state of the art you have to read up on academic literature to find out what that actually is. The initiative had a whole group of people who spent their time doing that. Then the team decided what they wanted to productize, and how to make it production grade and how to make it scale – some things reproduced well, and some did not from the academic research. The below figure shows three components that the team built that worked and their academic research inspirations:

NLP Library Feature	State of the Art Research
Named Entity Recognition	"Entity Recognition from Clinical Texts via Recurrent Neural Network". Liu et al., *BMC Medical Informatics & Decision Making*, July 2017.
Word Embeddings	"How to Train Good Word Embeddings for Biomedical NLP". Chiu et al., In *Proceedings of BioNLP'16*, August 2016.
Assertion Status Detection	"Improving Classification of Medical Assertions in Clinical Notes". Kim et al., In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011.

Figure 5.33 – Drawing on Academic Research to Build State of the Art Algorithms

The Results

At the end of the day the team wanted to demonstrate an improvement in the demand forecasting of admission from the Emergency Department (ED) over the baseline. This was the forecasting point important for hospitals, as ED is the wildcard, while it is easier to forecast demand when you have e.g. scheduled surgeries and transfers from other hospitals.

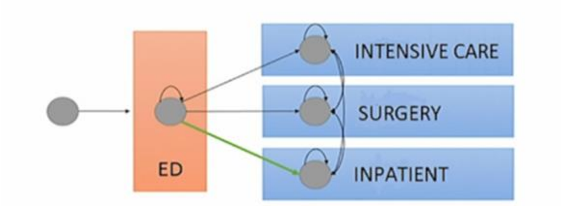


Figure 5.34 – Demand Forecasting of Admission from the Emergency Department

The baseline is human manual prediction – what a vast majority of hospitals do, i.e. they ask people, and have multiple meetings per day. The team found that it was fairly easy to beat this by using structured data:

Figure 5.35 – Adding Features from Structured Data

But while adding structured data improved performance versus human manual prediction, the more significant uplift came from adding features from the unstructured clinical notes:

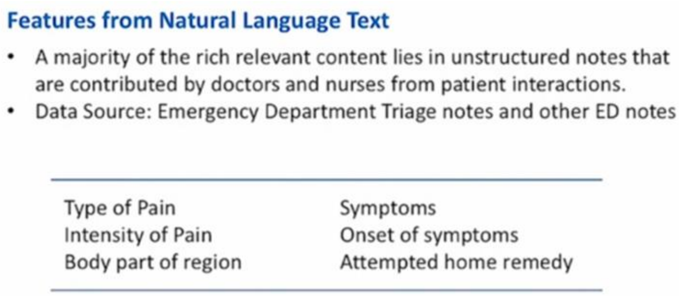


Figure 5.36 – Adding Features from Unstructured Data

The chart below depicts the relative improvements in prediction performance:

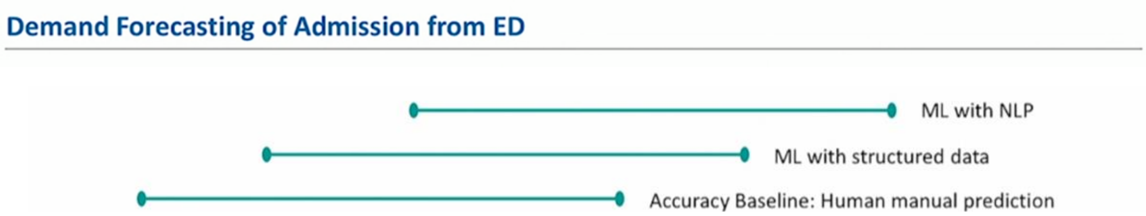


Figure 5.37 – ML with NLP is significant uplift over ML with structured data and over human manual prediction

Epilogue

In the health care domain, valuable information is in the unstructured notes, and leveraging this can facilitate a number of use cases. 2 case studies in recent healthcare literature also illustrate this:

1) **Detecting Sepsis Using Machine Learning**

   Early detection of sepsis is vastly improved when using unstructured notes:

   "Compared to previous work that only used structured data such as vital signs and demographic information, utilizing free text drastically improves the discriminatory ability (increase in AUC from 0.67 to 0.86) of identifying infection."[63]

2) **Cohort Selection in Oncology**

   Selecting a cohort is a common class of problem in the healthcare space – it means trying to find similar patients e.g. to identify who is a good fit for a clinical trial. Here too, using both structured and unstructured data was vastly more effective:

   "Among the 8324 people in the cohort generated using structured and unstructured data, only 2472 were also in the cohort generated using structured data only. Furthermore, 1090 people were included in the Structured data only cohort that are unlikely to meet the true parameters of the study population and would be erroneously included in an analysis that only uses structured data to select the study population."[64]

[63] (Horng et al., 2017)
[64] (Berger, Curtis, Smith, Harnett, & Abernethy, 2016)

6 Recommendations & Future Areas of Work

This chapter synthesizes the learnings from the case studies into a focused set of recommendations in section 6.1. Section 6.2 discusses future areas of work that can advance the subject of this thesis.

6.1 Recommendations

Ten key themes that emerge from the case studies are summarized below:

1. Always start with the business problem

All five case studies started with specific business problems: fraud loss for Danske bank, poor customer service for Deutsche Telekom (DT), time-consuming and labor-intensive asset monitoring workflows for General Electric (GE), being slow to gauge success of new product launches at General Mills (GM) and not being able to effectively forecast patient flow to predict bed demand and staffing levels at Kaiser Permanente (KP).

What these organizations did not do was attempt to bring in technology for technology's sake / go with the hammer of 'Artificial Intelligence' and start looking for nails. It is a common pitfall at many enterprises for data scientists to start by playing with the data. That is a recipe for failure – business needs should drive enterprise AI projects.

2. Invest in bringing together your data first

These enterprises had to invest significantly in building their data infrastructure before AI could add value, e.g. GM took a five year journey, the first three years of which were mainly just about getting the data organized, and the case details how disparate data sources were stitched together in a data lake. For GE, just getting the infrastructure set up in the industrial internet to collect data at scale, aggregate it, send it to the cloud, and have people securely access it was a complicated and time consuming process that laid down the foundation for subsequently building machine learning applications.

3. Training data is the 'new new oil'

It is not sufficient to just have data, a popular saying now is that if 'data is the new oil', training data is the 'new new oil'. For supervised learning problems (which comprise the majority of use cases today), to have a high quality annotated training dataset suitable for the decision problem at hand is also a substantive effort. Amongst the set of case studies, GM outlines an unsupervised approach in anomaly detection whereby they were able to circumvent the need for training labels, but this is less common. More typical is the wish expressed by Jan Hoffman of DT for the broader AI ecosystem to achieve training efficiency e.g. by automated cleansing and annotation of new data that feeds into the system in a closed learning loop.

4. Domain knowledge is key

Data scientists need to work closely with domain experts (such as fraud experts at Danske, or highly trained engineers at GE) to understand the domain so they can engineer meaningful features. AI systems also need to have the tooling in place such that new content can be fed into the system by domain experts in a way that does not feel like coding, e.g. at DT, non-technical customer service specialists provide information on e.g. new products/tariffs, feedback that a current dialog is not working etc. At KP, 300+ datasets covering terminology, clinical guidelines, measures etc. get curated and updated by clinical experts so they can be used as a reference. Without this domain knowledge, standard NLP techniques would be of no use on clinical notes.

5. Recognize that it is a journey and incrementally improve analytic techniques

Danske Bank started with machine learning and then added deep learning in the second phase. DT made Tinka first as a rules-based engine, and then added AI to make the assistant smarter. KP started with structured data and then added unstructured data, with an ongoing effort of looking at academic research to make sure their NLP algorithms were state of the art. Start with where data is available and the analytic techniques are more accessible, it will build in quick wins to build political capital, and build up the team expertise and confidence.

6. AI models are only a small part of the overall AI solution

Recognize that AI models are only a small part of the overall AI solution and invest in the surrounding infrastructure (e.g. the data ingestion pipelines, the model management framework etc.) Have the organizational commitment in place to build and maintain that platform. The platform will enable scalable development, deployment and maintenance of production grade AI applications.

At KP, the main challenge was building this platform and getting to a point where a community of data scientists at KP could use the platform appropriately and be productive. The GE case study also has a similar theme of empowering other data scientists to build and deploy full stack ML applications using the integrated platform built by the Wise team.

Besides empowerment of data scientists, platforms enable rigorous, repeatable processes for creating high quality systems. At GE, there was a focus on creating machine learning application templates that scale cross industry and to adjacent use cases. Danske's anti-fraud solution also built in data ingestion pipelines and a model management framework, and the bank viewed it as a blueprint for future projects and use cases.

7. Choose a partner wisely

Developing AI applications has fundamental differences from traditional software engineering. Not only the technologies but also the basic workflows are different. E.g. in AI, it is experiment driven development as you do not know upfront what models will work and whether the results will be useful. Even once you get a model working, a lot of ongoing work happens post production in model retraining. Given the new technologies and ways of doing work, it makes sense for enterprises to accelerate their AI journeys by partnering with an experienced vendor/consultancy initially, and we see that in all the case studies (except GE which took the route of acquiring Wise.io).

The landscape of potential partners is large and can be overwhelming. DT for instance started with more than 75, ultimately narrowing it down to 4 with which it did proofs of concept before finally picking one. Do detailed evaluations and aim for developing an AI ecosystem in the organization that is vendor agnostic and open source to have flexibility, given the technology landscape is very dynamic and evolving, and no single solution is best of breed, and you will want to plug in latest technologies as they emerge.

8. Work in cross functional, agile, collaborative teams

The Danske bank and DT case studies in particular highlight the importance of this. At Danske bank, cross functional, energized, collaborative teams delivered value in each iteration and were able to get from PowerPoint to shadow production in just 8 sprints of 2 weeks each. DT had a large distributed team, but strove hard to create the feeling they were all part of one team, and stressed the importance of being the best team that it could be by being the team that learns the fastest. Given the experimental nature of AI work, iterating quickly is key to success.

9. Integrate AI solutions into existing workflows and pay attention to the user experience

Give upfront thought to how the AI solutions can be integrated into existing workflows – will you augment or automate? What will the user experience look like? Will change management initiatives / user training be needed? At GM, the decision sciences team did not give this due consideration upfront and then found that even though their solution had great results (it detected which product launches were failures in only 6 weeks whereas previously it took a year), it was not being used as it was not integrated into marketers' workflows. Conversely, DT extracted value simply from optimizing the user experience even without AI, and plugged AI in as a subsequent step. At GE, there was a special focus on integrating machine learning applications into battle tested workflows with minimal disruption.

10. Build in explainability

AI applications cannot be black boxes and need to be able to communicate not just their results but also the data and reasoning that supports these. The Danske Bank case study for instance details how the team used open source technology called LIME (Local Interpretable Model-Agnostic Explanations) to build in explainability for the benefit of fraud investigators, customers, regulators and data scientists, and how work on visualizing and understanding deep learning networks is ongoing. Similarly at GE, e.g. the machine learning application for the aviation fleet monitor team clearly gave a suggestion, a confidence level, and a reason for the suggestion.

6.2 Future Areas of Work

The Chapter 4 framework components and sub-components can be delved into deeply – e.g. areas like explainability and algorithmic bias are active areas of research.

The case studies can be expanded. One direction to take might be to focus by industry and technology, e.g. 'NLP systems in health care', 'recommender systems in retail'. That would yield a deeper, domain-specific set of guidelines which practitioners could refer to. Expanding the sources of case studies might also be fruitful – in this thesis they originate from conference talks where companies are interested in marketing, public relations and recruiting, so there is a 'success bias' and failures, if any, are a blip in the trajectory to success. Doing primary research and talking to practitioners directly could yield some interesting insights from modes of failure as well as success.

As AI technologies grow and mature and are increasingly adopted by enterprises, more and more practitioner stories should be available and more best practices will coalesce. It is still early days in a fascinating, diverse and dynamic field.

Bibliography

Akred, J. (2017). Developing a modern enterprise data strategy - (Silicon Valley Data Science) - Part 1 - Strata Data Conference 2017 - Singapore. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/strata-data-conference/9781491985373/video317275.html

Barron, R. (2017). Executive Briefing: How to structure, recruit, operationalize, and maintain your insights organization - (InfoStrategy) - Strata Data Conference 2017 - Singapore. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/strata-data-conference/9781491985373/video317322.html

Bates, B., & McGrath, J. (2013). Theory 61: The McKinsey 7-S framework model - The Little Book of Big Management Theories. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/the-little-book/9780273785262/html/chapter-074.html

Berger, M. L., Curtis, M. D., Smith, G., Harnett, J., & Abernethy, A. P. (2016). Opportunities and challenges in leveraging electronic health record data in oncology. Future Oncology, 12(10), 1261–1274. https://doi.org/10.2217/fon-2015-0043

Bhattacherjee, A. (2012). Chapter 12 Interpretive Research | Research Methods for the Social Sciences. Retrieved May 25, 2018, from https://courses.lumenlearning.com/suny-hccc-research-methods/chapter/chapter-12-interpretive-research/

Black, D. (2017). Unlearning Racism and Sexism in Learning Machines. Retrieved May 25, 2018, from https://www.enterprisetech.com/2017/08/31/unlearning-racism-sexism-learning-machines/

Bodkin, R. (2017). Deep learning in the enterprise: Opportunities and challenges: Artificial Intelligence Conference. Retrieved May 25, 2018, from https://conferences.oreilly.com/artificial-intelligence/ai-ca-2017/public/schedule/detail/63424

Bodkin, R., & Gulzar, N. (2017). Fighting financial fraud at Danske Bank with artificial intelligence - Ron Bodkin (Teradata), Nadeem Gulzar (Danske Bank Group) - O'Reilly Artificial Intelligence Conference 2017 - New York, New York. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/oreilly-artificial-intelligence/9781491976289/video311819.html

Brant, K., & Austin, T. (2017). Gartner Hype Cycle for Artificial Intelligence, 2017. Retrieved May 25, 2018, from https://www.gartner.com/doc/3770467/hype-cycle-artificial-intelligence-

Bryan, L. (2008). Enduring Ideas: The 7-S Framework | McKinsey & Company. Retrieved May 25, 2018, from https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/enduring-ideas-the-7-s-framework

Chui, M. (2018). Executive Briefing: Artificial intelligence—The next digital frontier? - (McKinsey Global Institute) - Strata Data Conference - San Jose 2018. Retrieved from https://www.safaribooksonline.com/library/view/strata-data-conference/9781492025955/video319044.html

Chui, M., Lund, S., Madgavkar, A., Mischke, J., & Ramaswamy, S. (2018). McKinsey Global Institute: Notes from the AI Frontier. Insights from Hundreds of Use Cases. Retrieved from https://www.mckinsey.com/~/media/mckinsey/global themes/artificial intelligence/notes from the ai frontier applications and value of deep learning/mgi_notes-from-ai-frontier_discussion-paper.ashx

Copeland, M. (2016). The Difference Between AI, Machine Learning, and Deep Learning? | NVIDIA Blog. Retrieved May 25, 2018, from https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

Crawley, E., Cameron, B., & Selva, D. (2015). System Architecture. Retrieved May 25, 2018, from https://dl.acm.org/citation.cfm?id=2821260

DalleMule, L., & Davenport, T. H. (2017). What's Your Data Strategy. Retrieved May 25, 2018, from https://hbr.org/2017/05/whats-your-data-strategy

Deutsche Telekom: At a glance. (2018). Retrieved May 25, 2018, from https://www.telekom.com/en/company/at-a-glance

Eggers, J. (2017). It's the organization, stupid. - (Nara Logics) - O'Reilly Artificial Intelligence Conference 2017 - San Francisco, CA. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/oreilly-artificial-intelligence/9781491985250/video314957.html

eLIZA – the innovation project of Deutschen Telekom | welove.ai. (2018). Retrieved May 25, 2018, from https://www.welove.ai/en/about/the-project.html

Elprin, N. (2018). Managing data science in the enterprise - (Domino Data Lab) - Strata Data Conference - San Jose 2018. Retrieved from https://www.safaribooksonline.com/library/view/strata-data-conference/9781492025955/video319071.html

Fleener, W. (2018). Automating business insights through artificial intelligence - Wayde Fleener (General Mills) - Strata Data Conference - San Jose 2018. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/strata-data-conference/9781492025955/video319137.html

Flyvbjerg, B. (2011). Case Study. In The Sage Handbook of Qualitative Research (pp. 301–316). Retrieved from https://www.researchgate.net/publication/235953309

Fudge, T. (2015). Inductive vs. Deductive Writing | Purdue Global Writing Center and Academic Support Center. Retrieved May 25, 2018, from https://purdueglobalwritingcenter.wordpress.com/2015/02/25/inductive-vs-deductive-writing/

Fulde, V. (2018). Deutsche Telekom: How Telekom uses artificial intelligence (AI). Retrieved May 25, 2018, from https://www.telekom.com/en/company/digital-responsibility/details/how-telekom-uses-artificial-intelligence-524374

GE 2017 Annual Report. (2018). Retrieved from https://www.ge.com/investor-relations/sites/default/files/GE_AR17.pdf

General Mills, Inc. 2017 Annual Report. (2018). Retrieved May 25, 2018, from http://investors.generalmills.com/2017-interactive-annual-report/HTML1/general_mills-ar2017_0003.htm

General Mills: Brands overview. (n.d.). Retrieved May 25, 2018, from https://www.generalmills.com/en/Brands/Overview

Harpham, B. (2017). How to build a highly effective AI team | CIO. Retrieved May 25, 2018, from https://www.cio.com/article/3200190/careers-staffing/how-to-build-a-highly-effective-ai-team.html

Harvey, C. (2018). InformationWeek - AI Begins to Infiltrate the Enterprise. Retrieved May 25, 2018, from https://www.informationweek.com/big-data/ai-machine-learning/ai-begins-to-infiltrate-the-enterprise/d/d-id/1330809?print=yes

Heath, A. (2017). Read Amazon CEO Jeff Bezos 2016 letter to shareholders - Business Insider. Retrieved May 25, 2018, from http://www.businessinsider.com/read-amazon-ceo-jeff-bezos-2016-letter-to-shareholders-2017-4

Hoffman, J. (2017). AI-driven customer interactions - Jan Hofmann - Deutsche Telekom - IntelliSys 2017. Retrieved May 25, 2018, from https://www.youtube.com/watch?v=jmIvb2y28ls

Horng, S., Sontag, D. A., Halpern, Y., Jernite, Y., Shapiro, N. I., & Nathanson, L. A. (2017). Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLOS ONE, 12(4), e0174708. https://doi.org/10.1371/journal.pone.0174708

Kaiser Permanente 2017 Annual Report. (2018). Retrieved May 25, 2018, from https://share.kaiserpermanente.org/static/kp_annualreport_2017/#/numbers

Knight, W. (2017). Biased Algorithms Are Everywhere, and No One Seems to Care - MIT Technology Review. Retrieved May 25, 2018, from https://www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care/

Lynam-Smith, M. (2017). Launching AI in Enterprise: Success Factors- Miles Lynam-Smith. Rework AI Conference. Retrieved May 25, 2018, from https://www.youtube.com/watch?v=407zOZkGn74

Morgenthal, J. F. (2017). Tinka – the clever chatbot of T-Mobile Austria | DEUTSCHE TELEKOM B2B EUROPE. Retrieved May 25, 2018, from https://www.b2b-europe.telekom.com/blog/2017/10/25/tinka-the-clever-chatbot-of-t-mobile-austria

Ng, A. (n.d.). Deep Learning. Retrieved from http://cs229.stanford.edu/materials/CS229-DeepLearning.pdf
Ng, A. (2017). AI is the new electricity. - (Coursera) - O'Reilly Artificial Intelligence Conference 2017 - San Francisco, CA. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/oreilly-artificial-intelligence/9781491985250/video315023.html
Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping Business With Artificial Intelligence. Retrieved May 25, 2018, from https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/
Richards, J. (2018). Machine learning applications for the industrial internet - Joseph Richards (GE Digital) - Strata Data Conference - San Jose 2018. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/strata-data-conference/9781492025955/video318973.html
Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., … Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. NIPS. Retrieved from https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf
Starman, A. B. (2013). The case study as a type of qualitative research. Journal of Contemporary Educational Studies, 1, 28–43. Retrieved from http://www.sodobna-pedagogika.net/wp-content/uploads/2013/03/Starman1.pdf
Talby, D. (2017). When models go rogue: Hard-earned lessons about using machine learning in production - (Pacific AI) - Strata Data Conference 2017 - New York, New York. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/strata-data-conference/9781491976326/video316301.html
Talby, D., & Kulkarni, S. (2018). Spark NLP in action: Improving patient flow forecasting at Kaiser Permanente - David Talby (Pacific AI), Santosh Kulkarni (Kaiser Permanente) - Strata Data Conference - San Jose 2018. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/strata-data-conference/9781492025955/video318957.html
The Case Study as a Research Method. (1997). Retrieved May 25, 2018, from https://www.ischool.utexas.edu/~ssoy/usesusers/l391d1b.htm
Top 10 European operators revenue is €51 billion in Q2 2017. (2017). Retrieved May 25, 2018, from https://www.broadbandtvnews.com/2017/12/06/top-10-european-operators-revenue-is-e51-billion-in-q2-2017/
Tung, T. (2017). Executive Briefing: Becoming a data-driven enterprise—A maturity model - (Accenture Labs) - Strata Data Conference 2017 - Singapore. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/strata-data-conference/9781491985373/video317339.html
Weizenbaum, J. (1966). ELIZA---a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36–45. https://doi.org/10.1145/365153.365168
Wilder-James, E. (2014). The Experimental Enterprise. Retrieved May 25, 2018, from https://www.forbes.com/sites/edddumbill/2014/02/05/the-experimental-enterprise/#7a80080b76e4
Wright-Jones, B., & Lidberg, S. (2017). The value of a data science center of excellence (COE) - (Microsoft) - Strata Data Conference 2017 - Singapore. Retrieved May 25, 2018, from https://www.safaribooksonline.com/library/view/strata-data-conference/9781491985373/video317295.html