EXPERIAN
INNOVATION
SUMMIT
2019

DATA ENGINEERING : NOUVEAU MOTEUR DE CROISSANCE

DATATECHNOLOGYANALYTICS

experian™

Scaling AI in the Enterprise

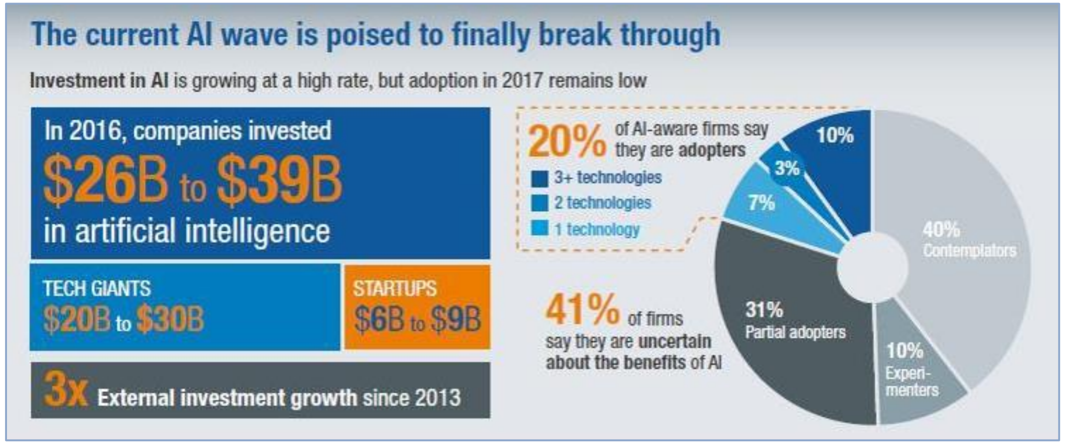
Javier Campos

Head of DataLabs EMEA and UK&I
Experian

AI: The new electricity or another buzzword?

Why the hype?

Tech giants already investing +$40bn and making substantial returns. AI technologies will be the most disruptive over the next 10 years

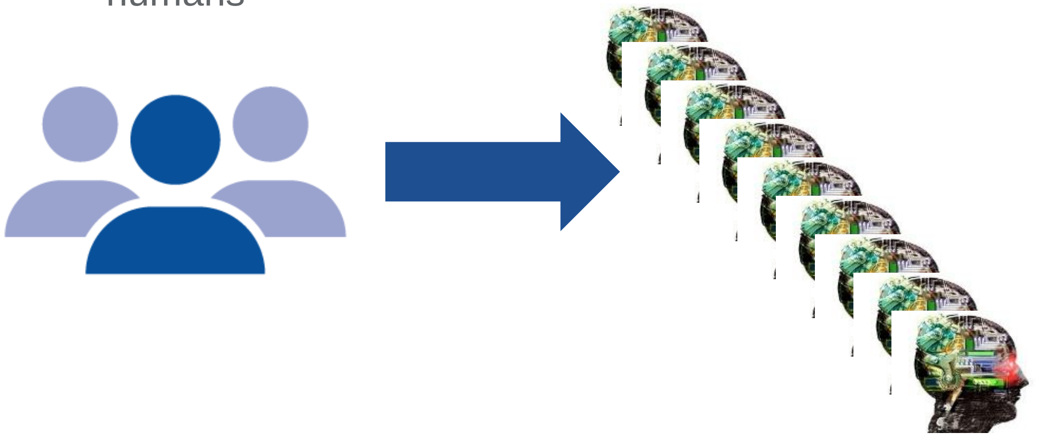


Why new Electricity?

AI technologies **automate Intelligent decisions** by adapting and learning from **local data at scale**

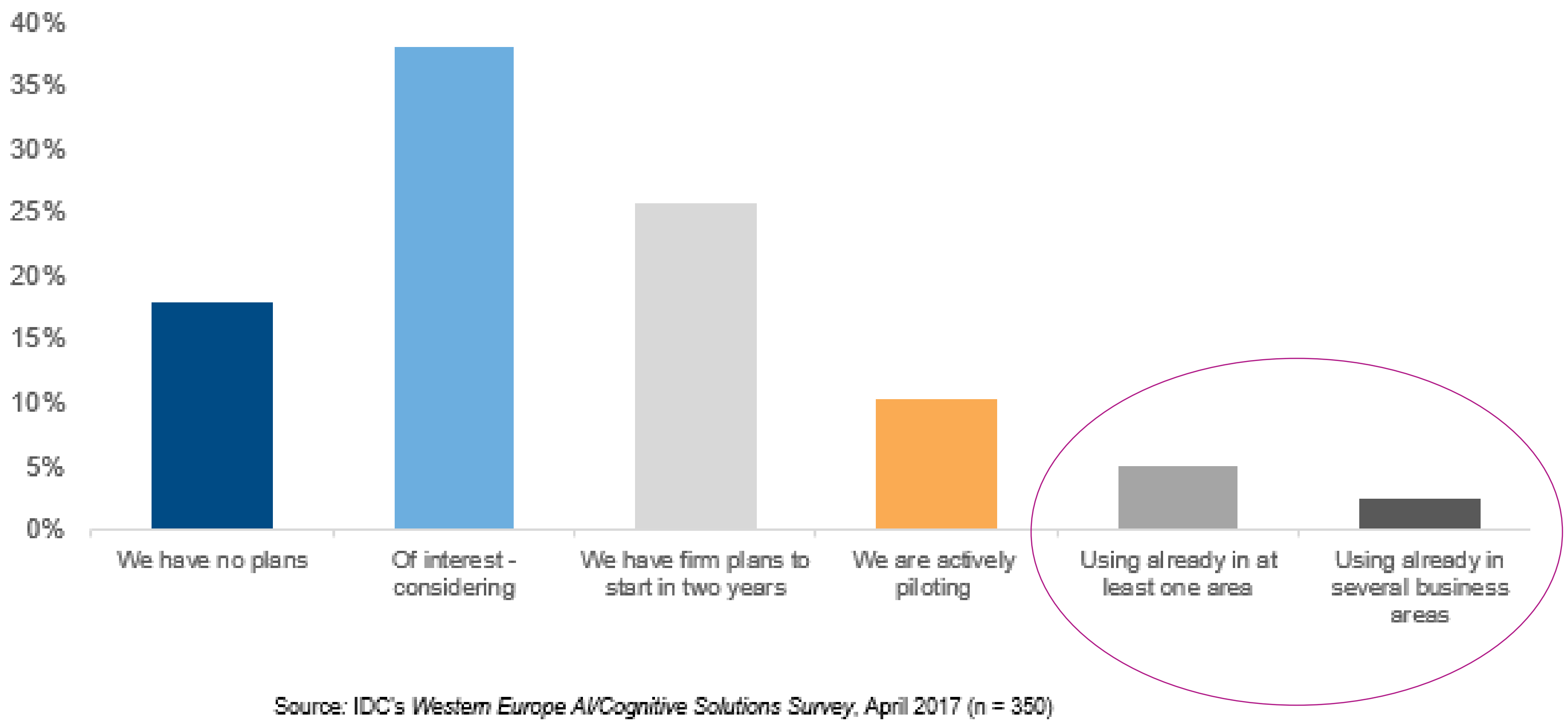
From central operations with humans

To distributed intelligent decisions



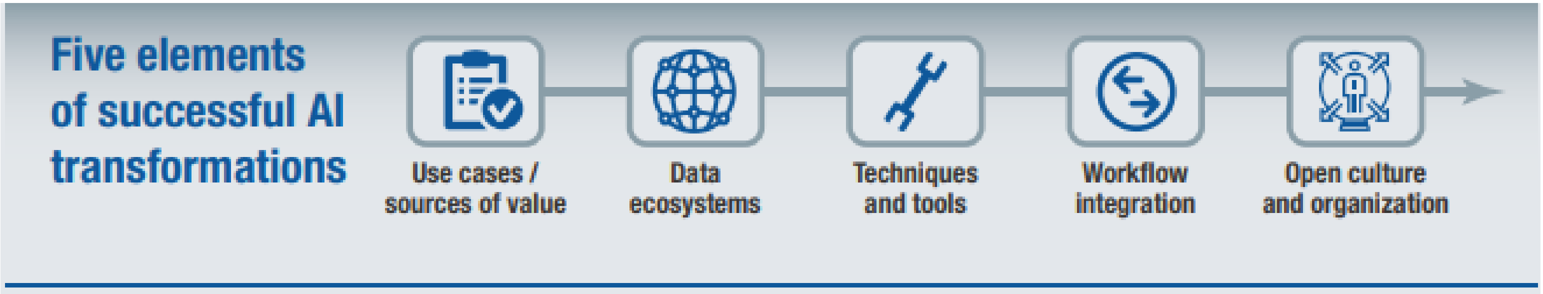
Despite the potential, still very few models in production

Current usage of AI Solutions in Europe:



Source: IDC's *Western Europe AI/Cognitive Solutions Survey*, April 2017 (n = 350)

Why is scaling so difficult - How hard can it be? …



Data & Technology	Data & Technology	Governance

• Having the right approach and Skillsets
• …
	• Manage new Data (IoT)
	• AI platforms:
	  • Uber: Michelangelo
	  • Google: TFX
	• …
		• Who's job it is? Chief AI?
		• How to align different teams
		• …

experian™

The HOW …

An approach: A Framework for AI

DATA ⟷ Algorithms ⟷ INFORMATION ⟷ Algorithms ⟷ DECISIONS

Experiments
Monitoring
Sensing

Theory
Models
Understanding

Physical systems
Financial systems
Human systems
Etc, etc.

Many companies are trying to leverage AI by applying the Data to the algorithm(s), but …

… A better way is to start from the Decisions – this is where the value lies.

Focus on Data science

Alchemy
vs
Science

Balanced Team

Critical roles – “**Translator**” (business to DS), “right” Data Scientist, Tech and UX

The fantastic Four

The Translator

The Tech

The Scientist

The right part of brain
(UX creative)

Micro DS sprints
2-3 weeks

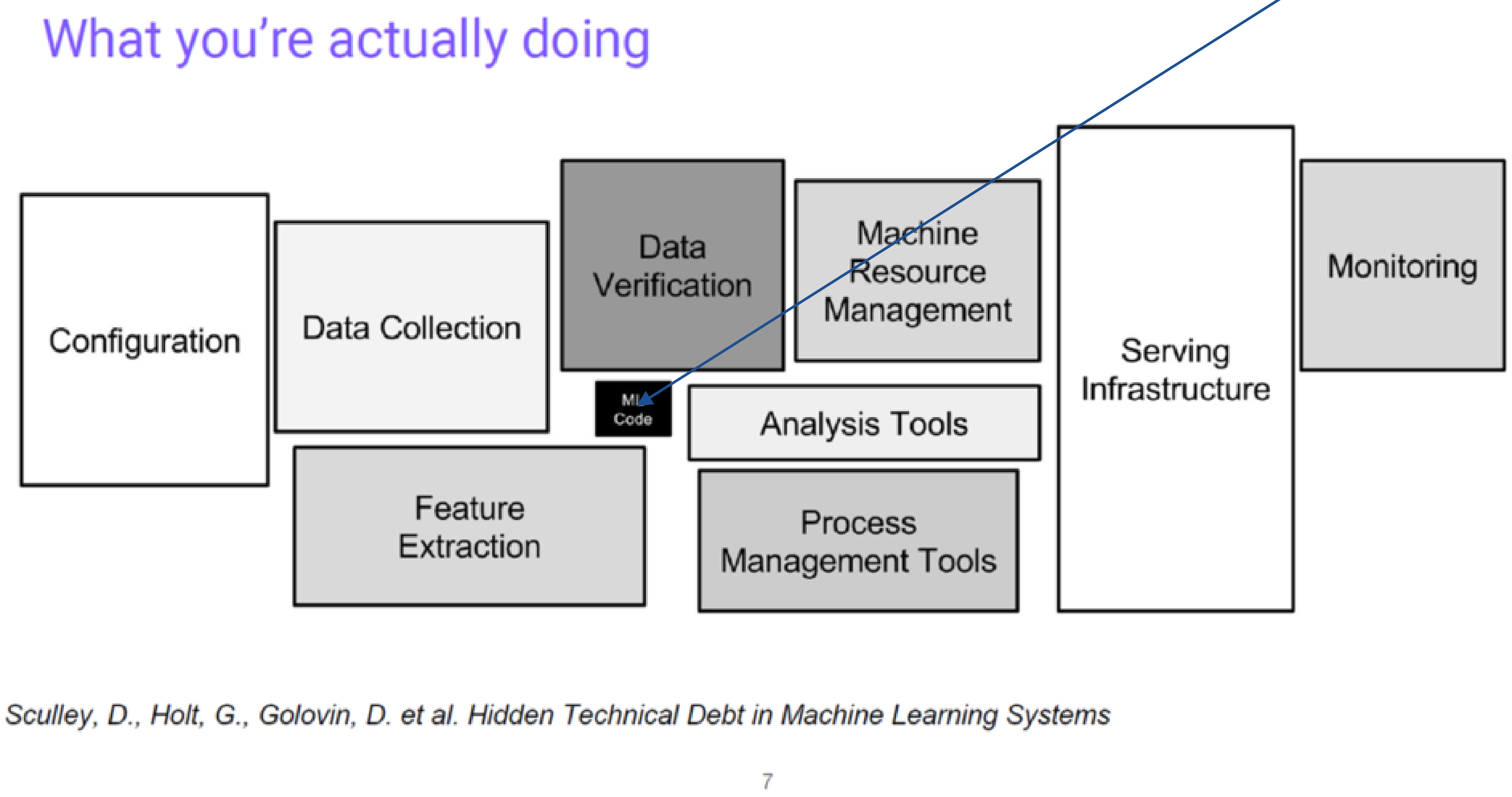
RAPID INCUBATION ➡ **GO TO MARKET** ➡ **PRODUCT EXPLOTATION**

The WHAT …

What a ML system looks today …

Trying to fit a ML system within existing infrastructure and methodologies is likely to introduce a system with very high technical debt …

This is the actual AI algorithm!

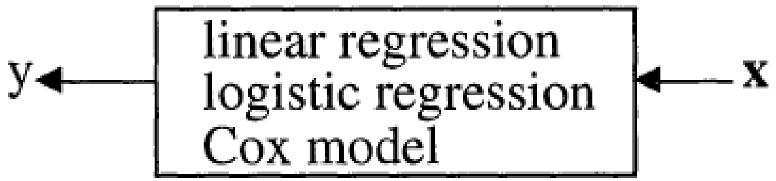


Machine Learning vs Statistics: The two cultures

The current analytical platforms are also not best suited, starting with the differences of methodology…

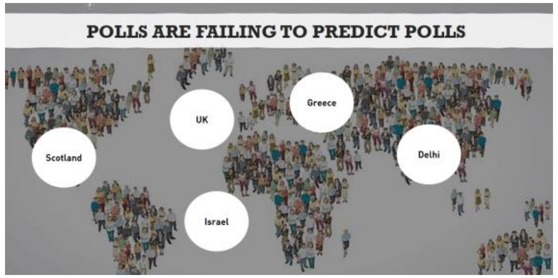
The Data modelling culture

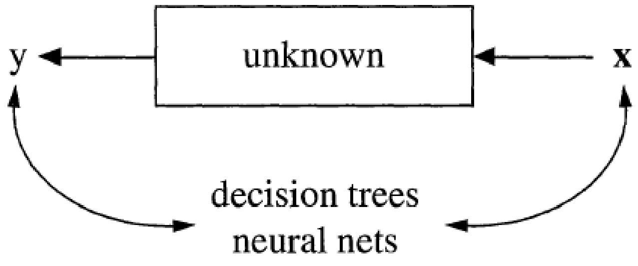
Focus on **the model**. For example, a common data model is that data are generated by independent draws from response variables = f(predictor variables, random noise, parameters



The Algorithmic culture

All about **the data** – optimisation problem. The analysis in this culture considers the inside of the box complex and unknown. Their approach is to find a function f(x)-an algorithm that operates on x to predict the responses y.





A better way: Data driven architecture vs Software driven

To implement the next gen of AI platforms, we need data driven architectures. Three key factors – co-evolution across subsystems (lack of decomposability), focus on data and outcomes, continuous deployment AND operation

- ML systems are **NOT decomposable** unlike traditional Software (e.g microservices)
- ML systems present **CO-EVOLUTION** – same as many elements in nature - more efficient way to evolve in dynamic environments
- ML systems should focus on **data outcomes** – not input or software

Data culture

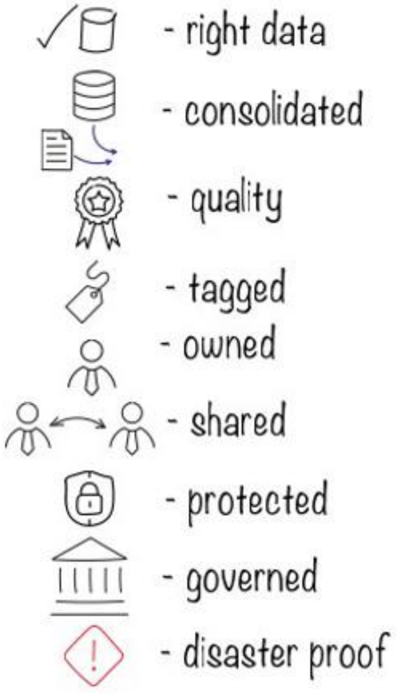
A cultural fit is required to transform data from today's byproduct of software systems to an primary asset that needs different focus, starting with ownership, governance and protection

Software module – all build around it

Data as by-product / exhaust

Software module – build around data

Data as main asset – focus on outcomes



experian

The WHY …

Operating a ML environment: Example Finance

Actually

- 170 models developed and deployed by year
- Mean model lifetime: 2 years

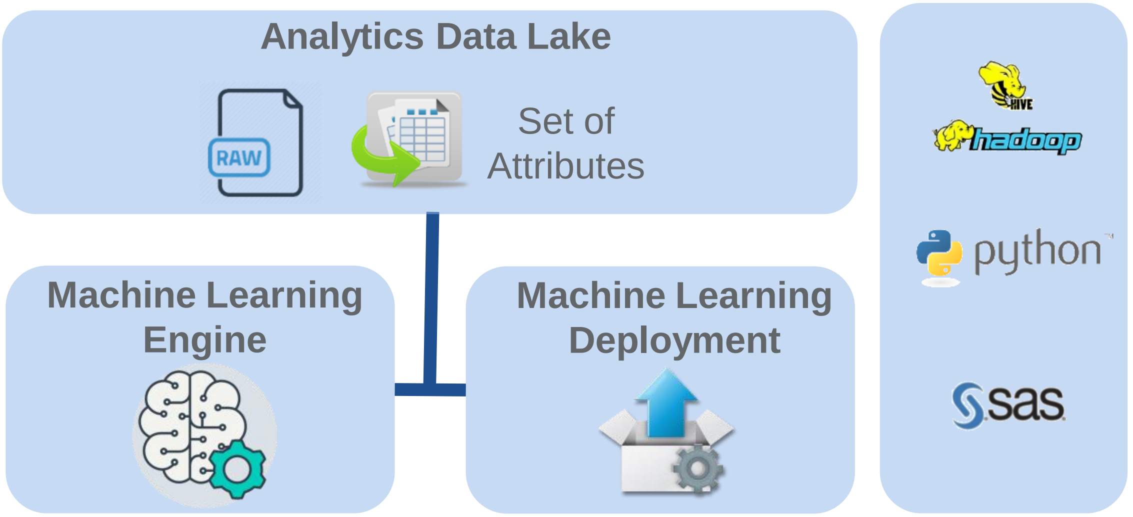
Online Learning

- 680 models developed and deployed by year
- Model lifetime (strict): 3 months

A new modelling environment & Operations becomes necessary

Machine Learning Environment

- Lowers development time (30 to 10 days)
- Lowers deployment time (30 to 5 days)
- Makes machine learning models a reality
- Model and deployment interfaces automatically connected



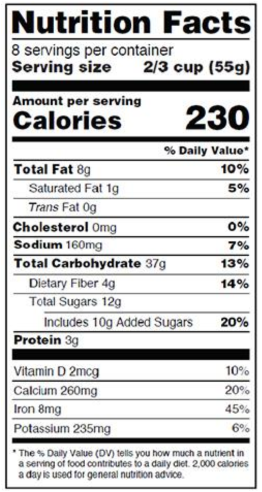
Little known secret of AI…

Current AI platforms needs large number of people tagging (e.g. labelling) the data in the background and overseeing production exception (e.g. Alexa call centre)

Future of AI… Trusted AI

Several dimensions need to be address when designing an AI system – a framework for trusted AI:

F - FAIRNESS & ETHICS
A - ACCOUNTABILITY
C - CUSTOMER
T - TRANSPARENCY
S - SAFETY/SECURITY


Nutrition Facts
8 servings per container
Serving size 2/3 cup (55g)
Amount per serving
Calories 230

	% Daily Value*
Total Fat 8g	**10%**
Saturated Fat 1g	**5%**
Trans Fat 0g	
Cholesterol 0mg	**0%**
Sodium 160mg	**7%**
Total Carbohydrate 37g	**13%**
Dietary Fiber 4g	**14%**
Total Sugars 12g	
Includes 10g Added Sugars	**20%**
Protein 3g	
Vitamin D 2mcg	10%
Calcium 260mg	20%
Iron 8mg	45%
Potassium 235mg	6%

* The % Daily Value (DV) tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.

Data Science Facts

Fairness

1. Was the dataset and model checked for biases?
2. Was any bias mitigation performed on the dataset?

Accountability

1. Does the algorithm has a clear definition of who is accountable for all their outcomes – direct and indirect?

Customer

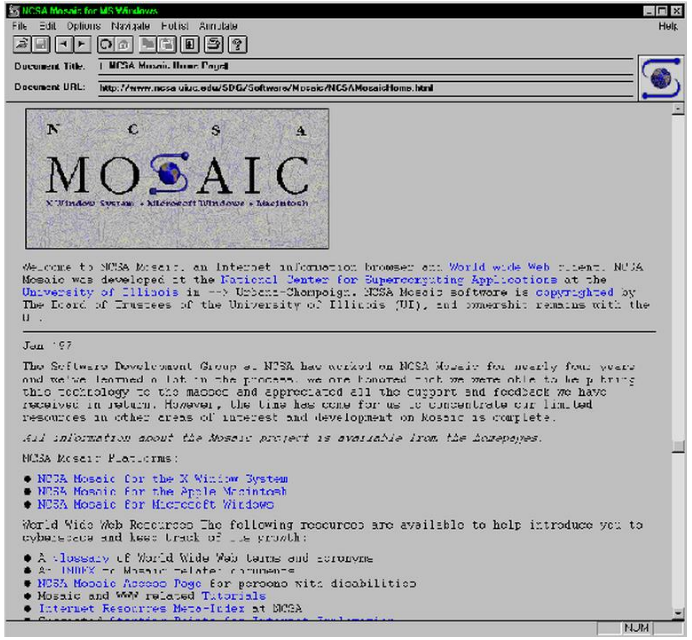
1. Was the service checked for robustness against adversarial attacks?
2. Is usage data from service operations retained/stored/kept?
3. What will be expected behavior if the input deviates from training/testing data?
4. What kind of governance is employed to track the overall workflow of data to AI service?

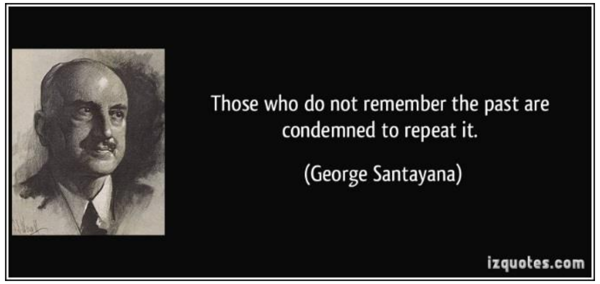
Transparency

1. Are algorithm outputs explainable/interpretable
2. Who is the target user of the explanation (ML expert, domain expert, general consumer, regulator, etc.)
3. Was the service tested on any additional datasets? Do they have a datasheet or data statement?

The key Take aways…

Take away Point 1: When are we in the AI journey?





Software Engineering key stages

- 1945 to 1965: The origins
- **1965 to 1985: The software crisis**
- 1985 to 1989: "No Silver Bullet“ Software projects
- 1990 to 1999: Prominence of the Internet
- 2000 to Now: Lightweight methodologies

Software crisis(1965 – 1985):
- Many projects *ran over budget and schedule*. Some caused property damage. A few caused loss of life. Due to: productivity, quality and lack of qualified programmers

Are we about to enter a “**Data Crisis stage” in the AI journey?**
- Large gap on skillsets
- Impact on fairness – Fake news, violent use of Media
- Few deaths Tesla autopilot, industrial robots
- AI goes through hype periods followed by ‘Winters’ After the research peak in Deep Learning, a new generation of new algorithms will come which will make current platforms obsolete

There is no such a thing as a silver bullet AI platform. All giant Tech firms have resorted to custom builds but this requires very large pockets. Focus at this stage on specific short term business value, do not worry too much about standardization - just yet …

Take away point 2: Fairness and Governance framework:

"A.I. makes mistakes as humans do, only faster and at scale"

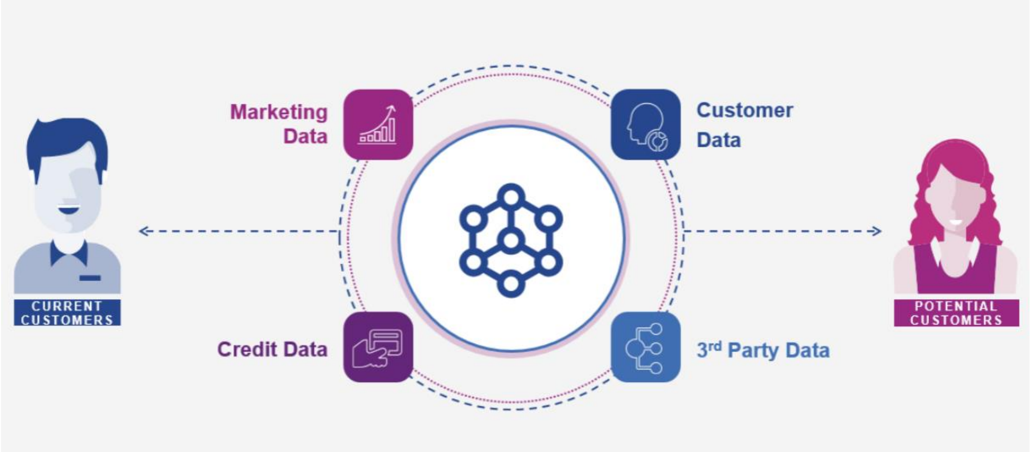
We can see a trend to start considering **Ethics as a quantifiable** (and measurable !) **business consideration** rather than as the current corporate social responsibility "soft" approach

As the machines make more and more critical decisions, we need to consider a wider AI framework: Privacy, Explainability are closely related to Fairness and rest of areas to consider when creating AI models. The regulators will go into this area so better prepare the organization.

Take away 3: Data, Data and data

We can see a trend to fusion data and platforms, instead of empty software (which is not that useful) outcome-data driven platforms

Example – Finance Experian Ascend platform – data and pre-define configurable outcomes



Foster culture around data… mix upskilling with bringing new talent – in particular ML product management is very hard. Set the right governance – Chief AI Officer? Among all, pragmatic based on business outcome – create an Operations capability that is data driven – continuous ops

EXPERIAN
INNOVATION
SUMMIT
2019

DATA ENGINEERING : NOUVEAU MOTEUR DE CROISSANCE

DATATECHNOLOGYANALYTICS