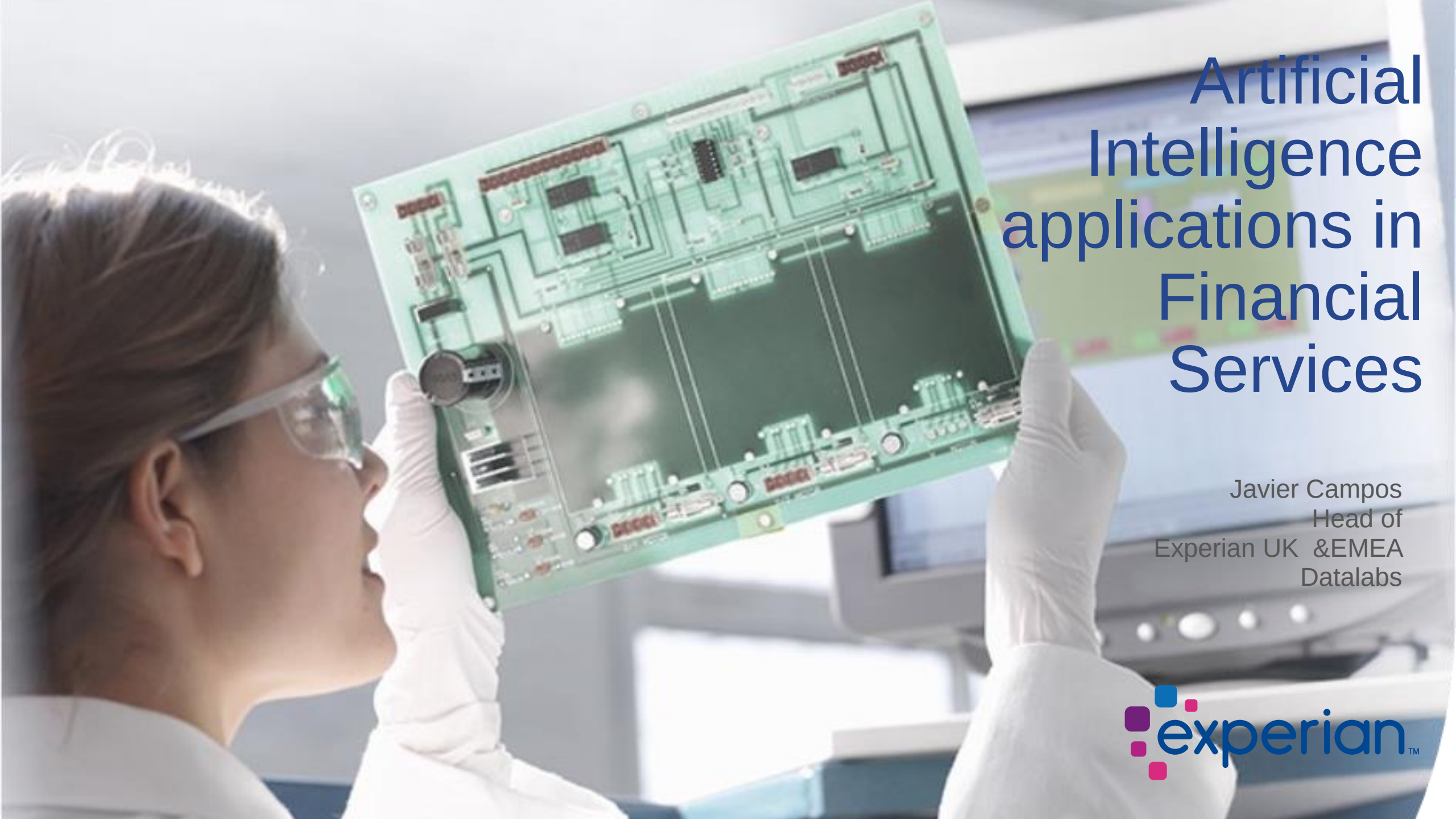


Artificial Intelligence applications in Financial Services

Javier Campos
Head of
Experian UK &EMEA
Datalabs

experian™

Experian – what do we do? We help consumers and companies with Financial Data and Analytics

INSURANCE
Facilitating insurance eligibility checks and claims

CREDIT MANAGEMENT
Facilitating real-time credit decisions

Alerting consumers about changes to their credit

ONLINE SHOPPING
Enabling relevant offers and messages

Verifying identities

HOME PURCHASES
Supporting mortgage applications

FINANCIAL ACCESS
Extending opportunities for the fair extension of credit and services

HEALTH VISITS
Qualifying patients for coverage / charitable benefits

AUTO PURCHASES
Ensuring visibility to vehicle history for safe purchase decisions

Experian - Helping people and organisations all around the world

17,800
EMPLOYEES

45
COUNTRIES

US **$5.2**bn
revenue

US **$403**m
reinvested in data, software, and technology

3bn credit decisions
supported, facilitating billions in loans

Credit data on:
>1.3bn people & **163**m businesses

>100,000
Diverse mix of clients

45
Countries worldwide

c.**20** milliseconds
real-time ingestion rate for data

54,500
Employee volunteer hours

We have offices in:

Argentina
Australia
Austria
Brazil
Botswana
Bulgaria
Canada
Chile
China
Colombia
Costa Rica
Denmark
France
Germany
Greece
India
Indonesia
Ireland
Italy
Japan
Lesotho
Malaysia
Mexico
Monaco
Mozambique
Namibia
Netherlands
New Zealand
Norway
Peru
Philippines
Poland
Russia
Singapore
South Africa
South Korea
Spain
Taiwan
Thailand
Turkey
Uganda
United Arab Emirates
United Kingdom
United States
Vietnam

experian™

Experian **DataLabs**

Research and development, innovation and incubation

Purpose

- Innovate with AI/ML, new data sources & new technologies
- Breakthrough research in a compliant, safe & secure environment
- Emphasis on financial services, retail, telecom and healthcare

Approach

- Hypothesis driven and rapid experimentation
- Global collaboration centers: San Diego, CA; London, U.K.; São Paulo, Brazil; Singapore
- New, currently unsolved industry and client issues

Previous Projects

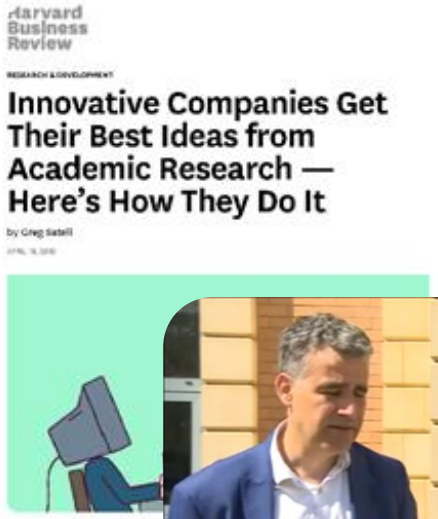
- Financial inclusion through AI and alternative data credit scoring
- Identity resolution using AI/ML and graph-based technology across digital and offline data
- Touchless application process using APIs and mobile technologies

~100 data scientists, technologists and industry experts focused on innovation

Core Capabilities

- Artificial Intelligence/ Machine Learning
- Explainable Artificial Intelligence
- Advanced data science/ non-linear modeling
- New technologies
- New data source integration
- Software

Driving thought leadership



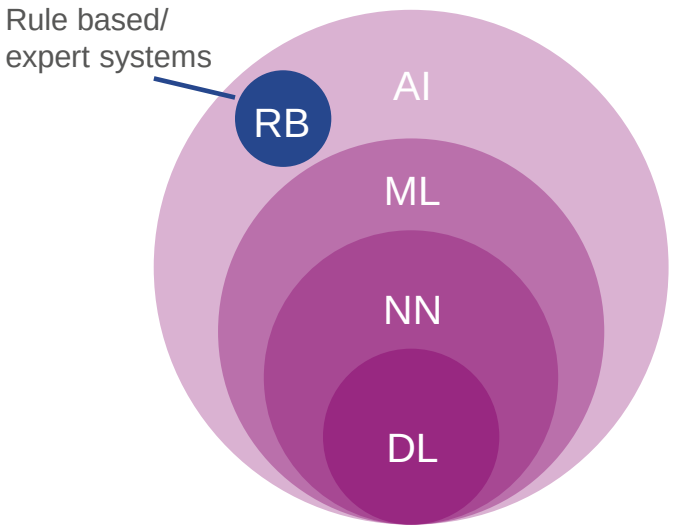
"**Experian** has become such an uncommonly successful innovator because it has been able to integrate both **disruptive and sustaining innovations** and power them both through its active involvement in the scientific community."

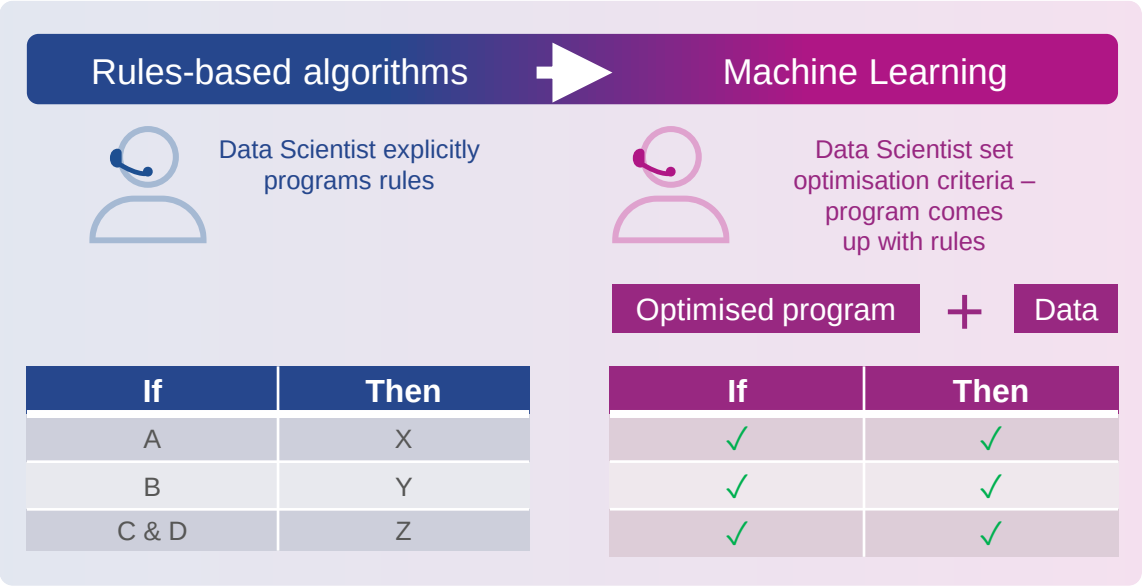
Source: "Mapping Innovation" by Greg Satell

What is AI and ML?

According to the Oxford dictionary AI is “the theory and development of **computer systems able to perform tasks normally requiring human intelligence,** such as visual perception, speech recognition, decision-making, and translation between languages”

AI can automate ‘human-like’ decisions at a local level at massive scale – without being explicitly programmed





If	Then
A	X
B	Y
C & D	Z

If	Then
✓	✓
✓	✓
✓	✓

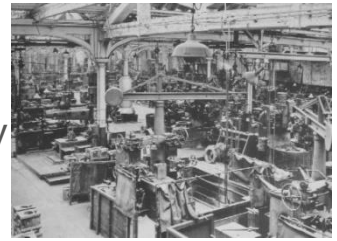
TESCO Bank

experian

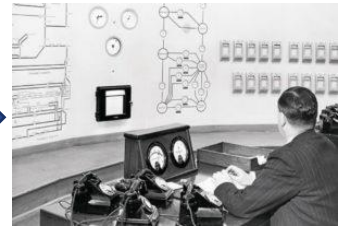
AI/ML – why it is important?

1. Electricity transformed Society 1900…

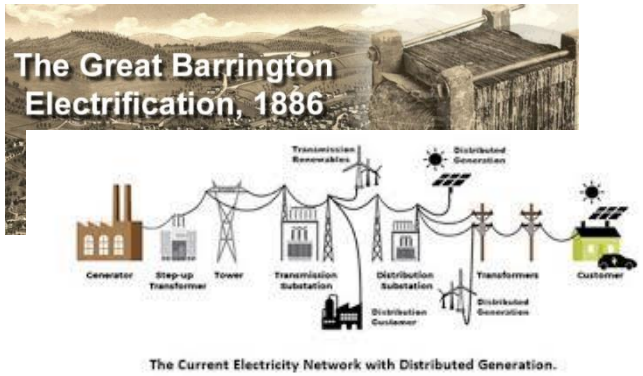
- Power concentrated around large factories -



- To distributed power-



Massive changes in society



2. AI, the new electricity, will be ubiquitous …

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

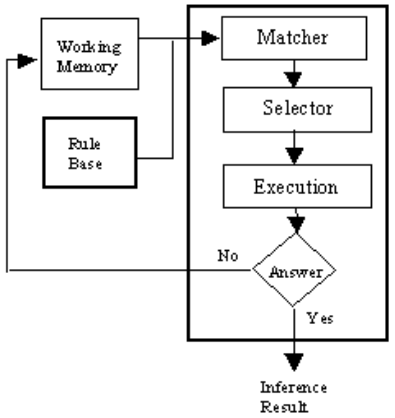
- From central operations with humans-

- To distributed intelligent decisions -

Jobs and society will change significantly…

AI/ML: Why it works

Most current legacy systems are mainly rule-based – Built over time adding 1,000's of manual rules



Machines learn by processing a valid training set that contains the features necessary to tune an algorithm. The algorithm enables machines to execute specific tasks:

AI/ML systems **exploits the underlying symmetries of the data**

WHY?

Physics	Machine learning
Hamiltonian	Surprisal $-\ln p$
Simple H	Cheap learning
Quadratic H	Gaussian p
Locality	Sparsity
Translationally symmetric H	Convnet
Computing p from H	Softmaxing
Spin	Bit
Free energy difference	KL-divergence
Effective theory	Nearly lossless data distillation
Irrelevant operator	Noise
Relevant operator	Feature

- **The SMALL PRINT:** Similar to what physics does with Math, in ML we could argue that the success of shallow neural networks hinges on symmetry, locality, and polynomial log-probability in data from or inspired by the natural world, which favours sparse low-order polynomial Hamiltonians that can be efficiently approximated.

MACHINE LEARNING USE CASES IN FINANCE

- Example initiatives Experian Datalabs -

Ideation

Spot Opportunity

- Bureau Based Fraud Detection (DA)
- AML KYC (Digital)
- Employment DB
- AI CLIMATE CHANGE Risk Model & Greenscore(DA)
- Synthetic Data
- Scoreboost v2 (ECS)
- P&M Engine Replacement

Design stage

- NHS COVID
- Macro-Aware Credit Score
- Impact Stability Credit Score
- Client– CC Risk Models
- MLOps

Incubation

- Fairness as Service (FaaS)
- AI ADDRESS MATCHING
- GLOBAL BOOST

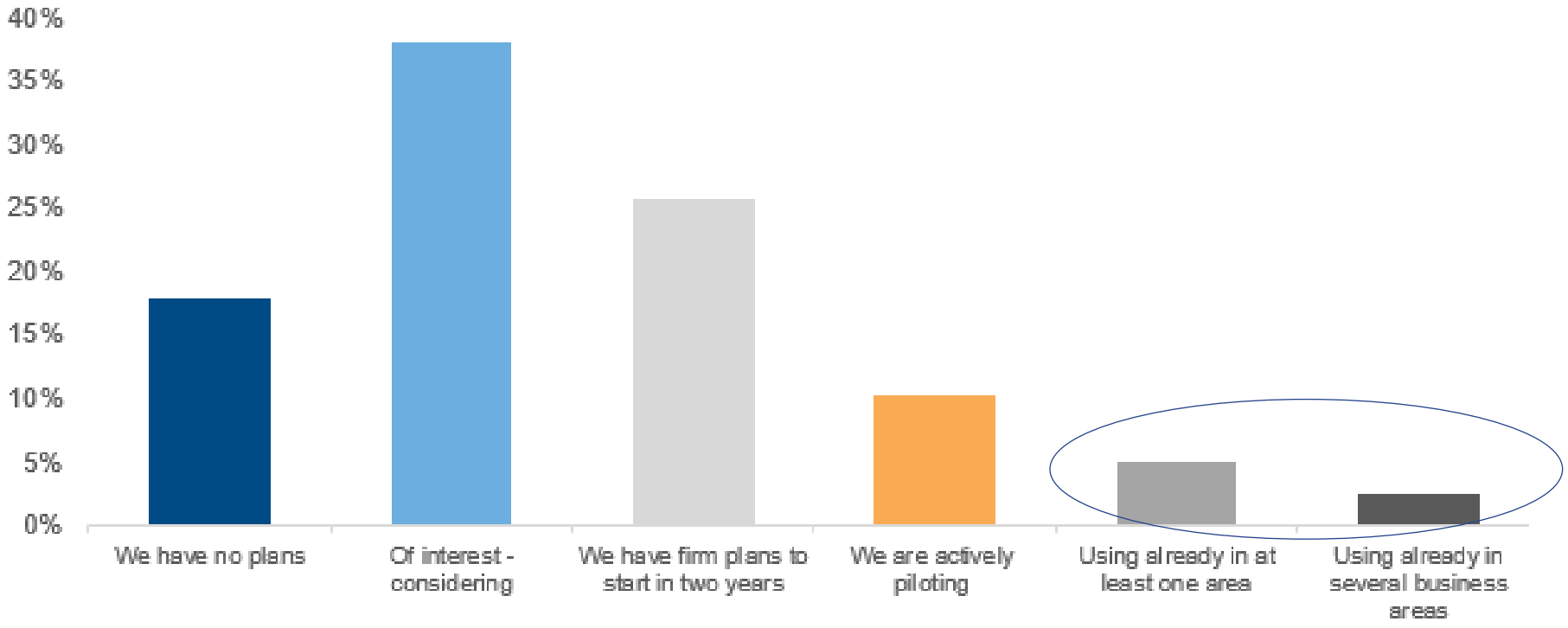
Transitioned to Scale

- Trusso
- HUNTER AI MODULE
- Trusso SME

experian

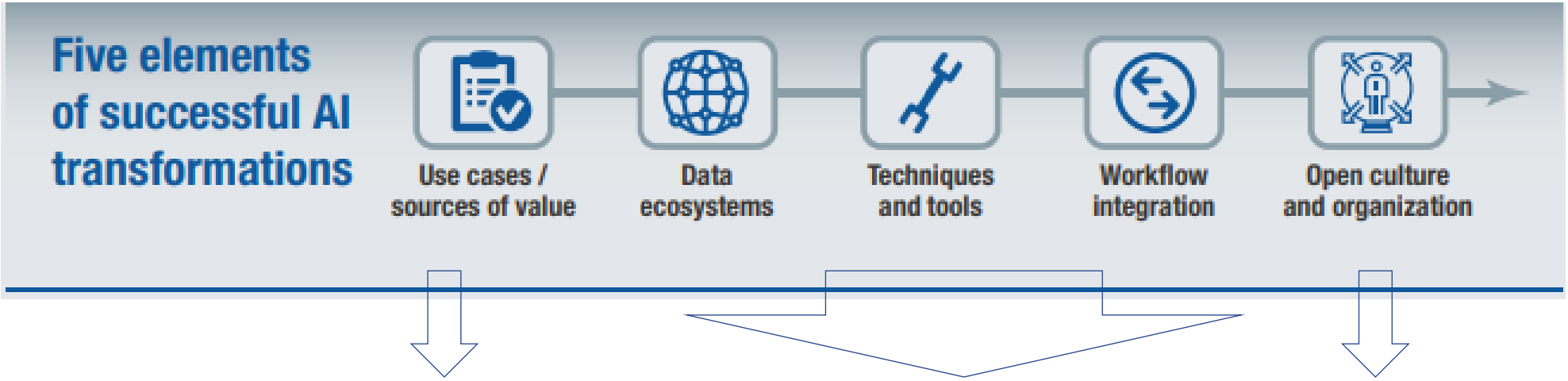
Despite the potential, still very few models in production

Current usage of AI Solutions in Europe:

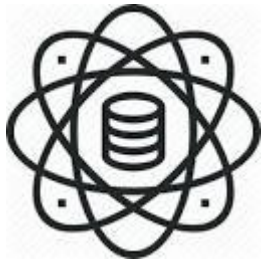


Source: IDC's *Western Europe AI/Cognitive Solutions Survey*, April 2017 (n = 350)

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

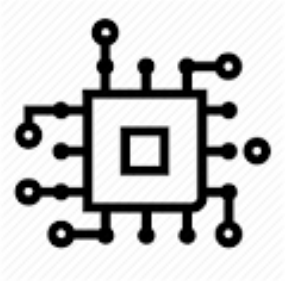


- Data Science -



- Having the right approach and Skillsets
- …

- Data & Technology -



- Manage new Data (IoT)
- AI platforms:
  - Uber: Michelangelo
  - Google: TFX
- …

- Governance -

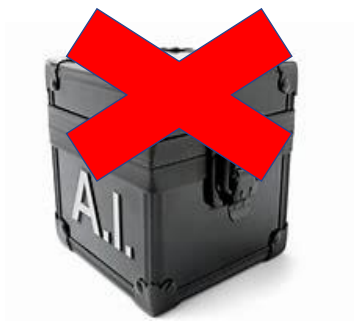
- Who's job it is? Chief AI?
- How to align different teams
- …

Little known secret of AI…

Many 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:



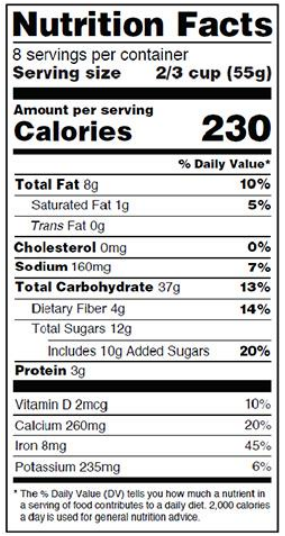
F - FAIRNESS & ETHICS

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?

ian™

Going deep in a few examples…

UK COVID: NHS Trust Model forecast and risk map

Our state of the art model is a Bayesian probabilistic model based on a well-principled Discrete SEIR model, ideal to estimate latent variables such as R

The labs is made up of **3 core skills** sets:

DATA SCIENTISTS

SOFTWARE ENGINEERS

PRODUCT MANAGERS

Covid19 ***Epidemiological*** *Simulator and Forecaster*

Input Data

- Death Rates: 13 EU Countries + UK
- Population statistics (density across age & location
- Death Rates: UK @ Regional Level (120 NHS Trusts)
- Actual history of social distancing policies (13 EU & UK)
- Mobility data (public)
- 111 calls

Bayesian Hierarchical Model

Advantages:

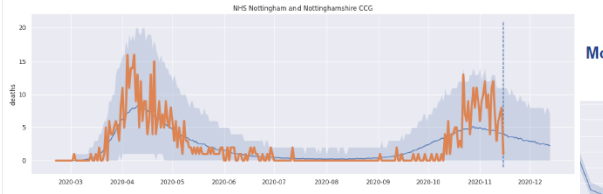
- ✓ *Daily time series*
- ✓ *Realistic data-driven uncertainties*
- ✓ *Based on 13 EU countries, the UK, 120 NHS Trusts*
- ✓ *Results can be per country / region / trust*
- ✓ *What-if simulation module*

Output Data

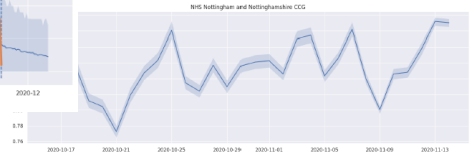
- Infected Fatality Rate (IFR)
- Infection rates
- Reproductive Ratio (R_t)
- Death rates

R prediction/Deaths

Covid19 **Mobility + 111 model calls deaths**



Mobility + 111 model results



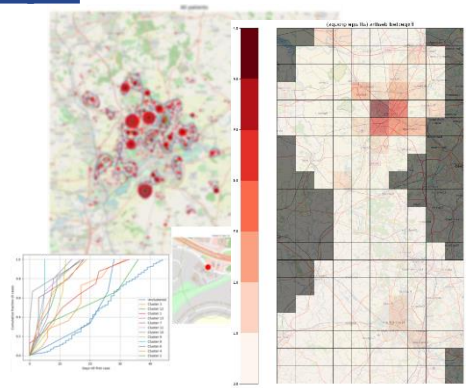
R = 0.80 – 0.91

Hot spots/Risk maps

Using the residential address of COVID-19 patients and utilising datasets available by Experian, we identified tight clusters of cases (hotspots), which we examined both temporally and spatially

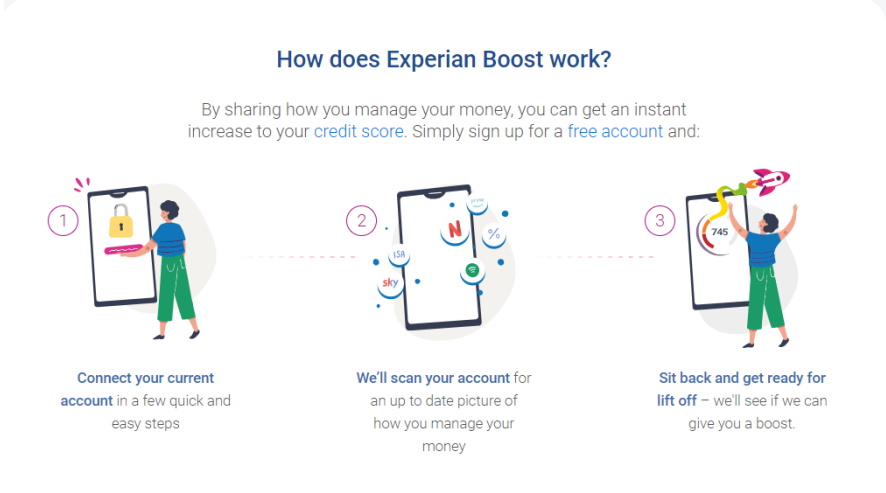
These results can be used to:

- ✓ Identify mature or young hotspots of infection to apply targeted testing and eventually "squash the fire before it expands"
- ✓ Understand nature of hotspots to inform lock-down lifting strategy.
- ✓ Extend forecasting models (use-case 1a) to account for the clustering tendency of COVID-19 infections.

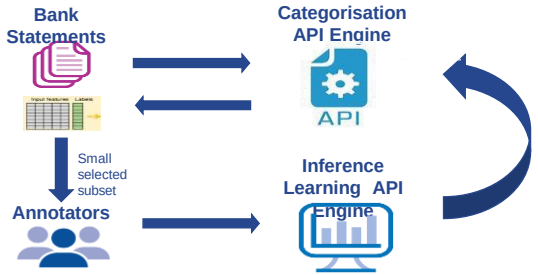


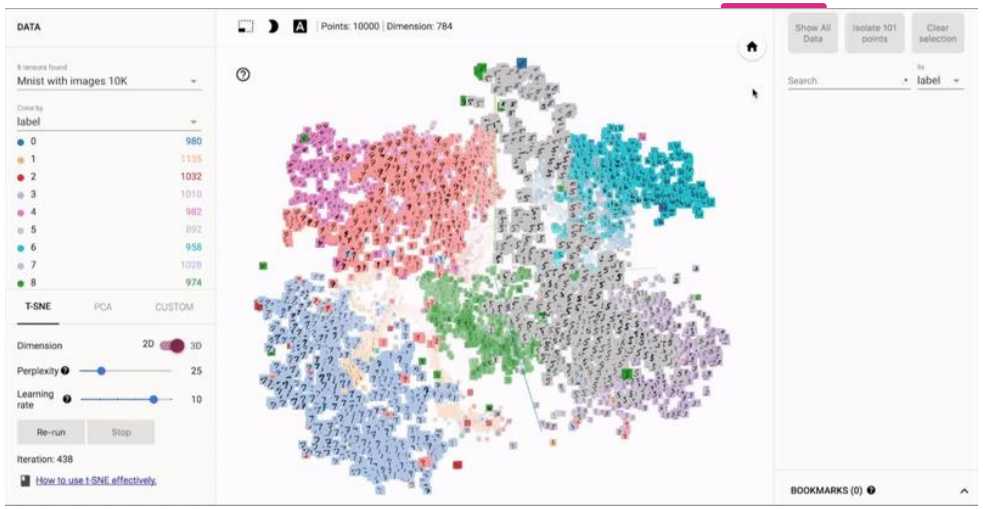
Experian Boost UK – Two major innovations came from Datalabs

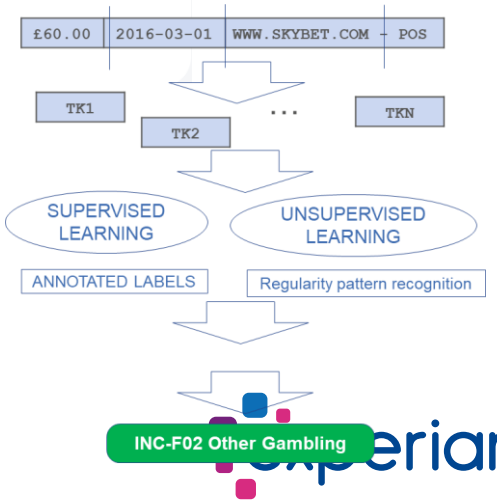
We have created a product that using the transactions from your bank account in real time, can help increasing your score and hence getting more credit or in better conditions.



- Categorisation Engine (Trusso) -







The power of ML – AI credit risk models

Dual Bureau ML model has **8 GINI point uplift** and **35.2% Bad rate reduction** compared to traditional model.

Traditional Modelling

Bank Logistic Regression

Single Bureau	Dual bureau
73.5 GINI	74.2 GINI

Machine Learning

Experian XGBoost Model

Single Bureau	Dual bureau
77.5 GINI	83.1 GINI

Credit Risk Models
Beta Calibration

Bank application (10 features)

Experian core & TransUnion core & TransUnion affordability features

Experian pre-processing

Machine learning

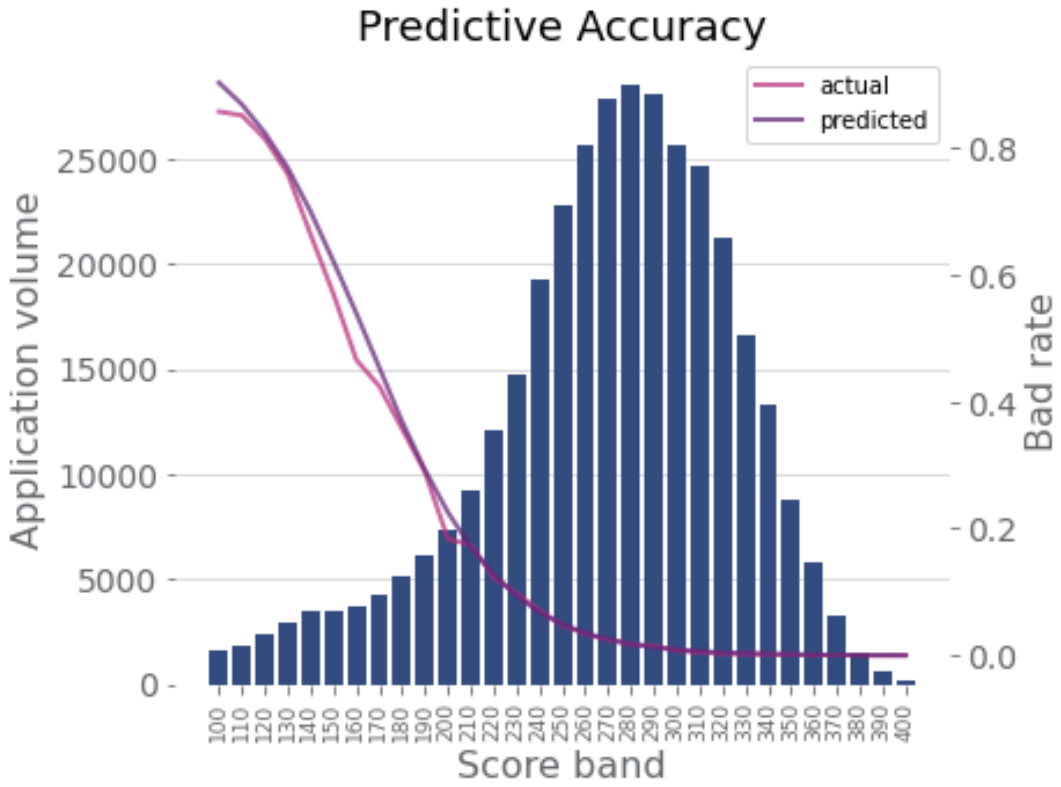
Post-modelling calibration

Dual Bureau ML model

Model Accuracy and Swapsets

Original Model*	New bad rate	Uplift
2.4%	1.5%	36%

Original decision	New decision	Applications	Bads	Test bad rate
accept	accept	221617	2206	<1%
accept	decline	31241	3881	12%
decline	accept	30210	1671	5%
decline	decline	71921	25013	34%

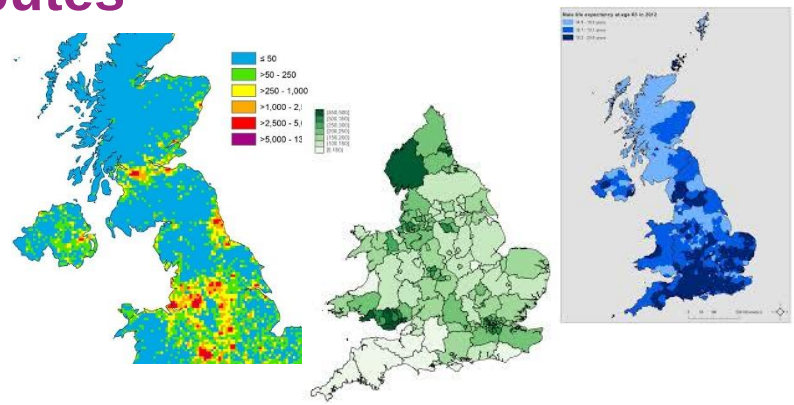


* Tesco Bank Dual Bureau Logistic Regression

Will ML improve Bias or not?

How are you helping financial inclusion/bias reduction?

Experian normative datasets => Statistical Distribution of protected attributes

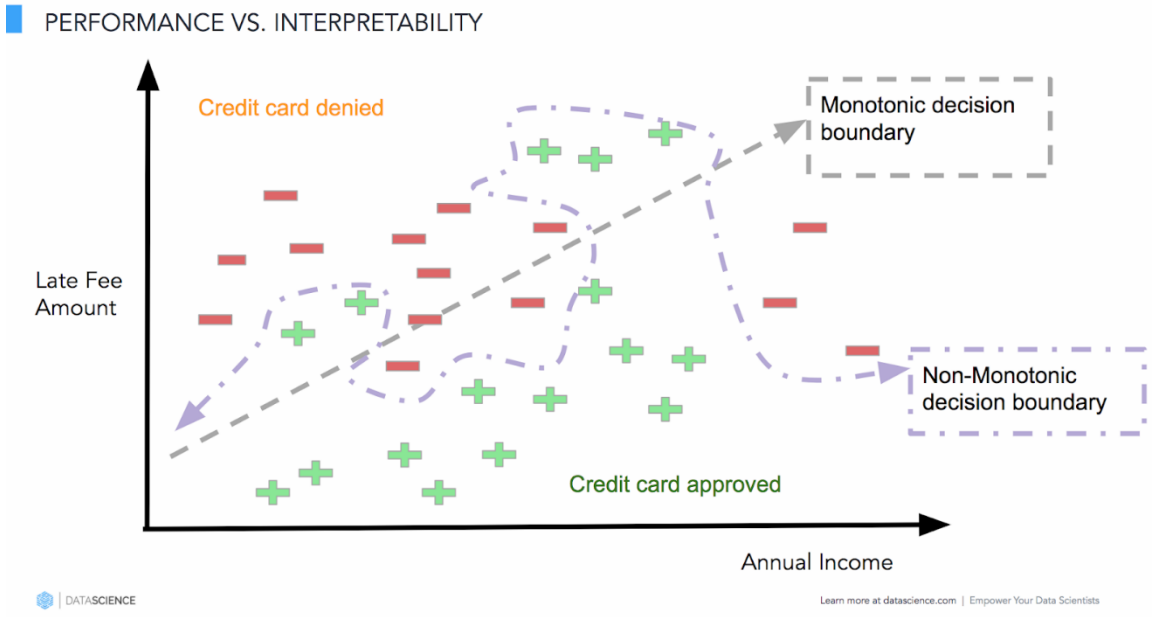


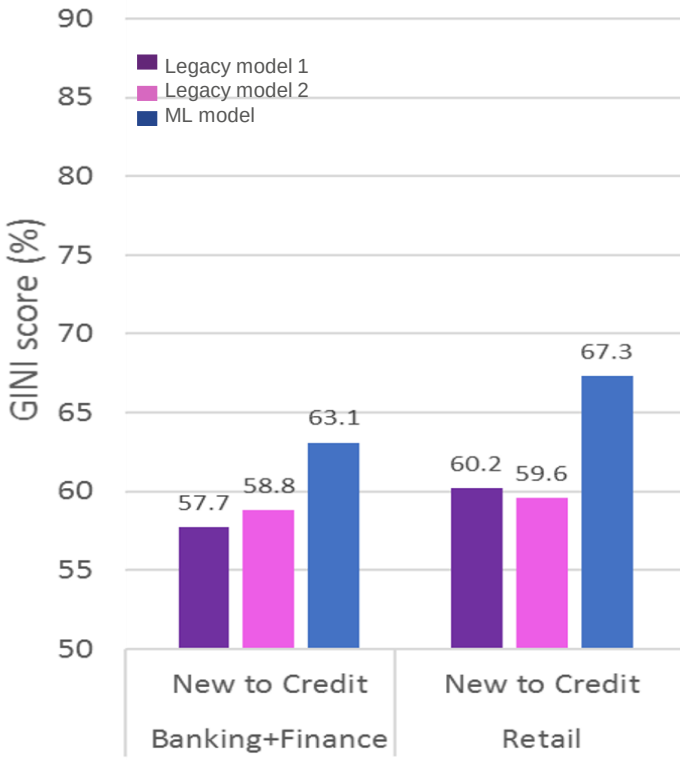
YES ! ML improves bias reduction over Logistic Regression - bank

Metric	Attribure	Improvement
Disparate Impact	Nationality	4%
	Ethnicity	22%
	Religion	22%
Eq. Opportunity	Nationality	3%
	Ethnicity	16%
	Religion	22%
Equalised Odds	Nationality	1%
	Ethnicity	8%
	Religion	19%

It is all about the customer

If you are using the past to predict the future, are you creating a barrier to entry?





experian™

Machine Learning Use case 1 – Anomaly detection
Fraud Attributes

We are applying new and innovative techniques to enhance transaction fraud detection systems

Transactional fraud detection

Traditional approach

Learn fraud patterns from **PAST** fraud tags

Limited ability to identify new fraud patterns

New thinking

Identify behavior different from what's normally done

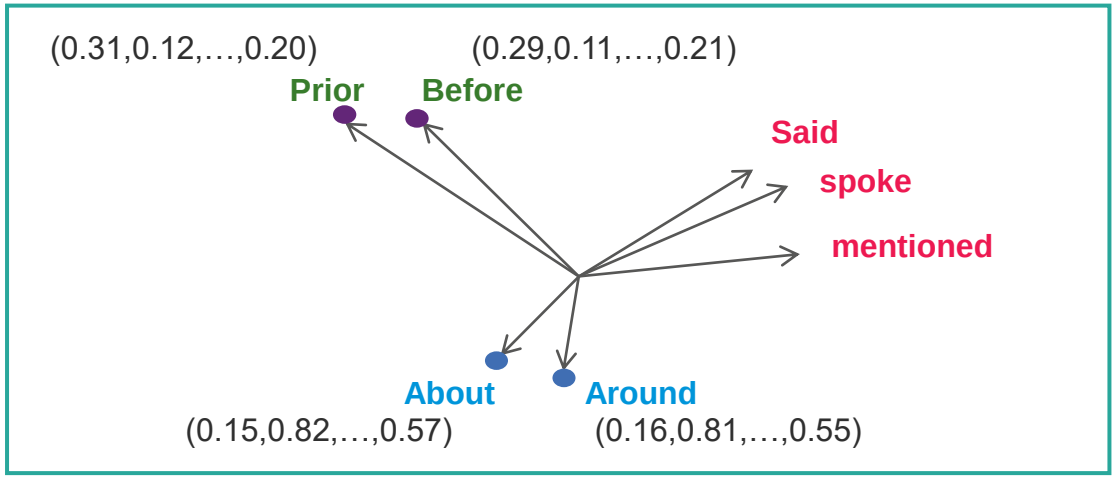
- **What is the normal behavior?**
- **How do we represent normal behavior in models?**

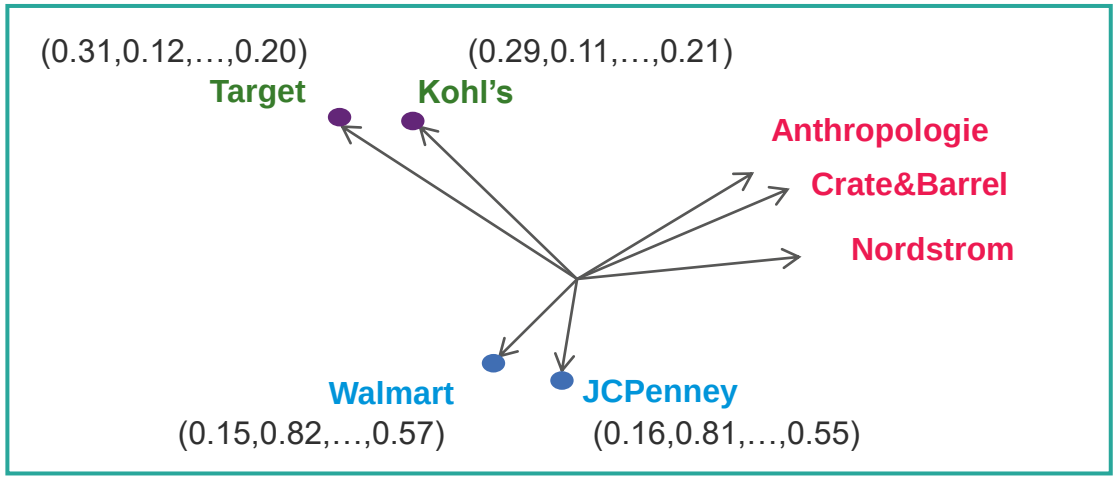
Neural Embedding
Fraud attribute technology

- Deep learning / language model technique
- Each word is associated with a point in a high dimension space similar words ➔ close vectors

- Similarity measured by cosine similarity

$$\text{similarity} = \cos(q) = \frac{A \times B}{\|A\| \times \|B\|}$$





Transaction Data Insights
Fraud Attributes – Spend Characteristics

Neural embedding algorithms are applied on multiple aspects of spend to identify behaviors that are consistent with elevated fraud risk

What did you buy?
- Merchants frequented
- Merchant category

When did you buy?
- Time of day / Day of Week spending habits

How much did you buy?
- Average transaction amount
- Cumulative spend
- Overall and within different spending categories

Where did you buy?
- Online vs. Offline
- Geo-location analysis (highly concentrated vs. dispersed)

+

+

+

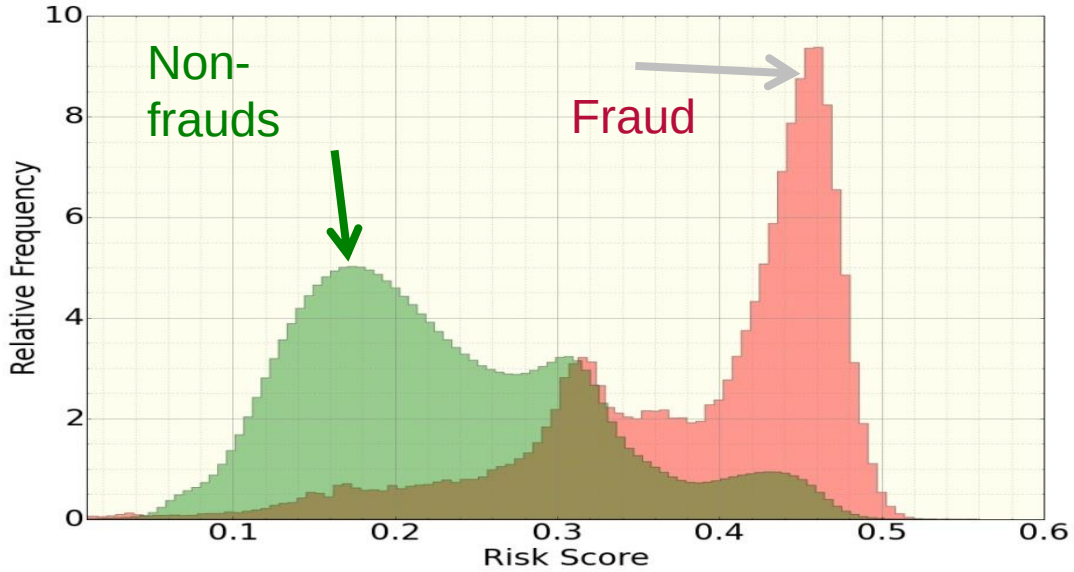
=

experian™

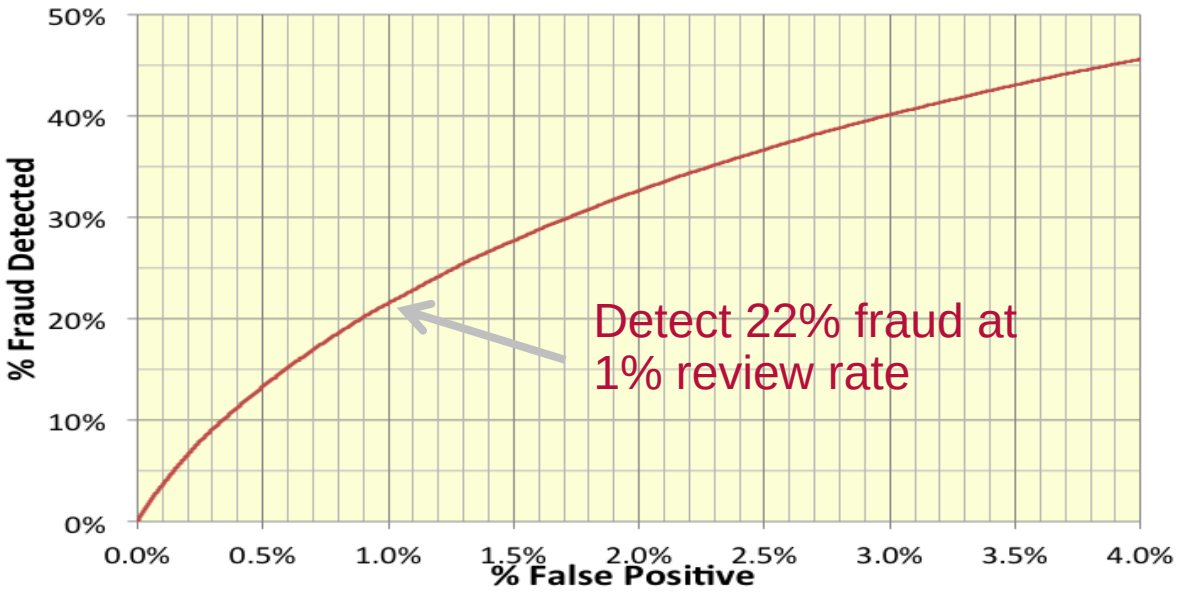
Transaction Data Insights
Fraud Attributes – Performance

Strong incremental predictive power in assessing fraud transaction risk

- Top **single** attribute achieves ~22% fraud detection rate at 1% review rate
- Lowering customer friction by reducing false positive



Score distribution

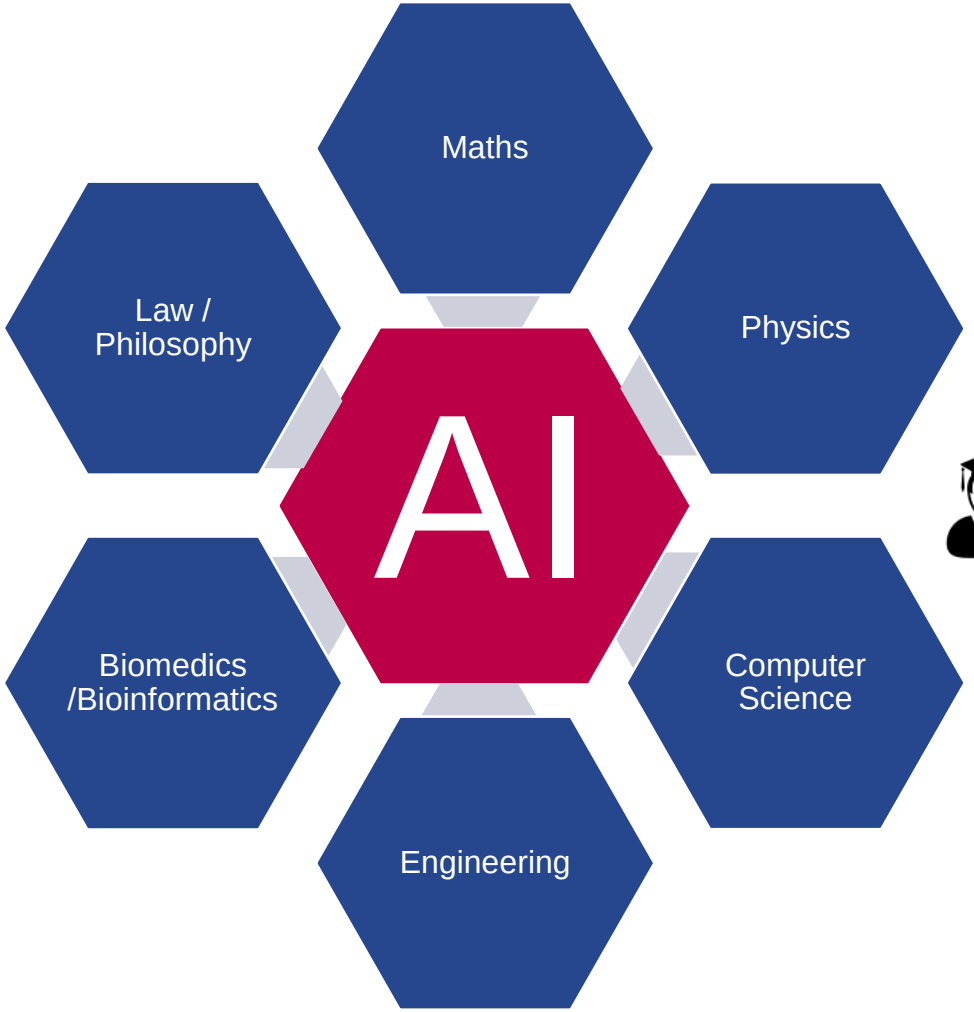


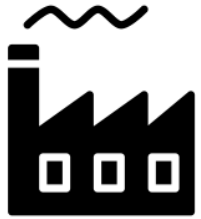
Single attribute performance

How to get a career in AI

How to get a career in AI

Many different degrees will provide the right skillsets and foundations to work in the field. Once the degree is completed, also several options to be a practitioner:





Industry (Applied research or implementation: Healthcare, Finance, manufacturing, retail, etc)

Academia (Research on Research)

AI Start ups

Some AI jobs

Researcher AI

Business Analyst

Chief Ethics Officer

Data Analyst

Big Data Engineer

Scala Developer

AI product manager

Machine Learning Engineer

AI Cyber-security

Data Scientist

AI Forensics

Robotics Engineer

AI General Manager

DevOps AI

AI Engineer

AI Lawyer

Data Science Tooling Engineer

Source: Oxford University careers, Indeed.com

experian

AI Long term – Myths vs Facts

Myth:
Superintelligence by 2100 is inevitable

Myth:
Superintelligence by 2100 is impossible

Fact:
It may happen in decades, centuries or never: AI experts disagree & we simply don't know

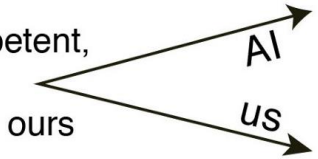
Myth:
Only Luddites worry about AI

Fact:
Many top AI researchers are concerned

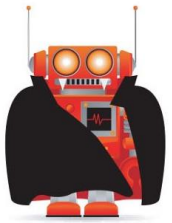
Mythical worry:
AI turning evil

Mythical worry:
AI turning conscious

Actual worry:
AI turning competent, with goals misaligned with ours



Myth:
Robots are the main concern



Fact:
Misaligned intelligence is the main concern: it needs no body, only an internet connection

Myth:
AI can't control humans

Fact:
Intelligence enables control: we control tigers by being smarter

Myth:
Machines can't have goals

Fact:
A heat-seeking missile has a goal

Mythical worry:
Superintelligence is just years away

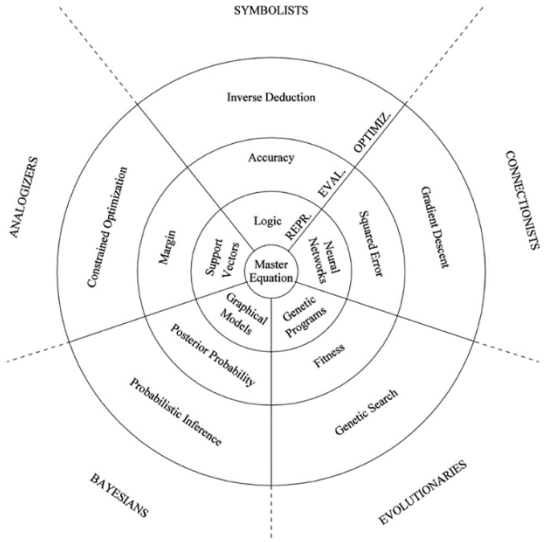
Actual worry:
It's at least decades away, but it may take that long to make it safe

experian™

The key Take aways…

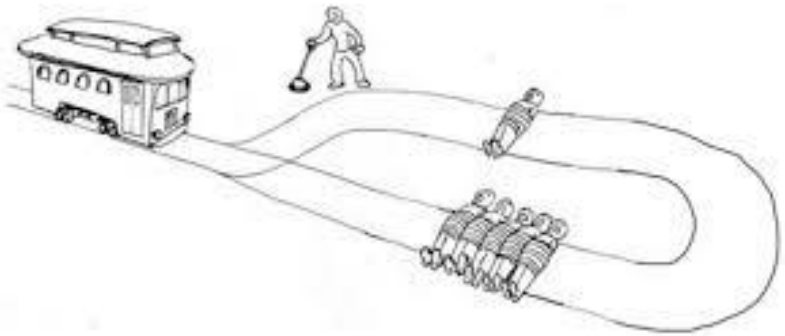
Take away Point 1: Where are we in the AI journey?

Despite awesome recent progress (Winning Go, protein folding, etc), many areas still greenfield ! Many different ways to get into AI, not just computing science.



Human-level AI: The roadmap

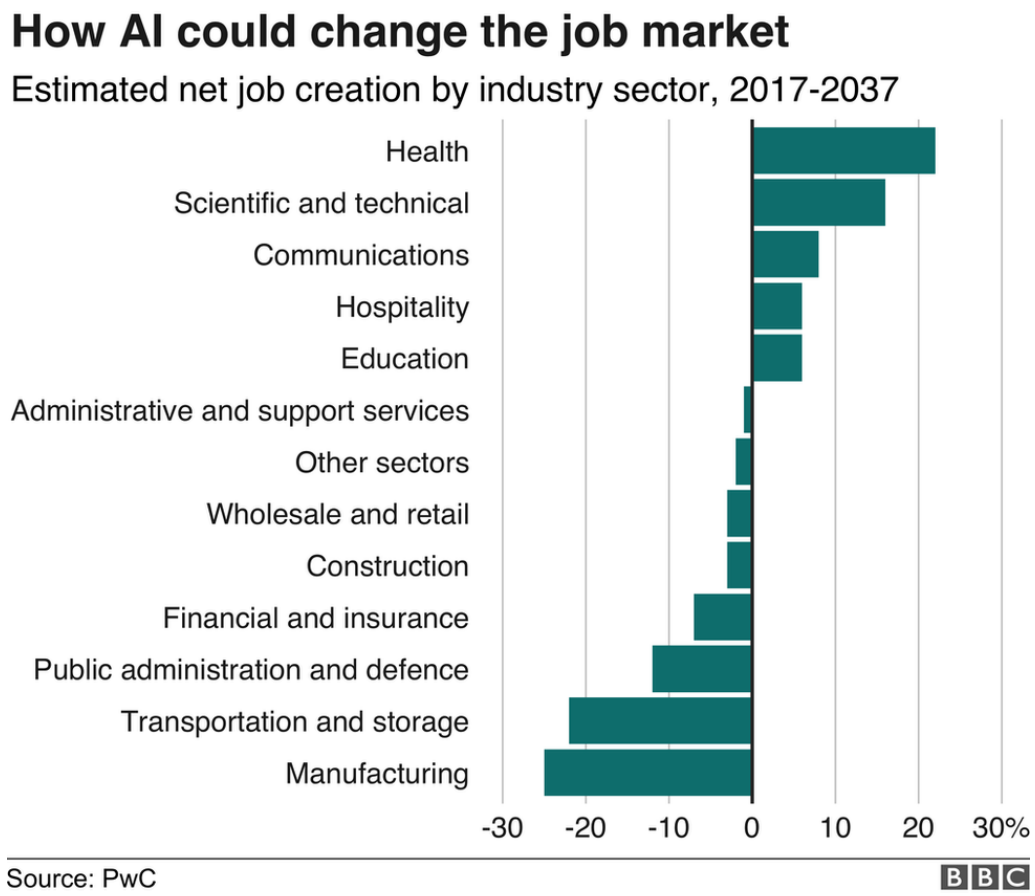
Need to Solve for	State of Play
Generality	Solved ✓
Learning without being taught	Solved ✓
Transfer learning	Not yet
Common sense	Not yet
Self-awareness	Still mysterious

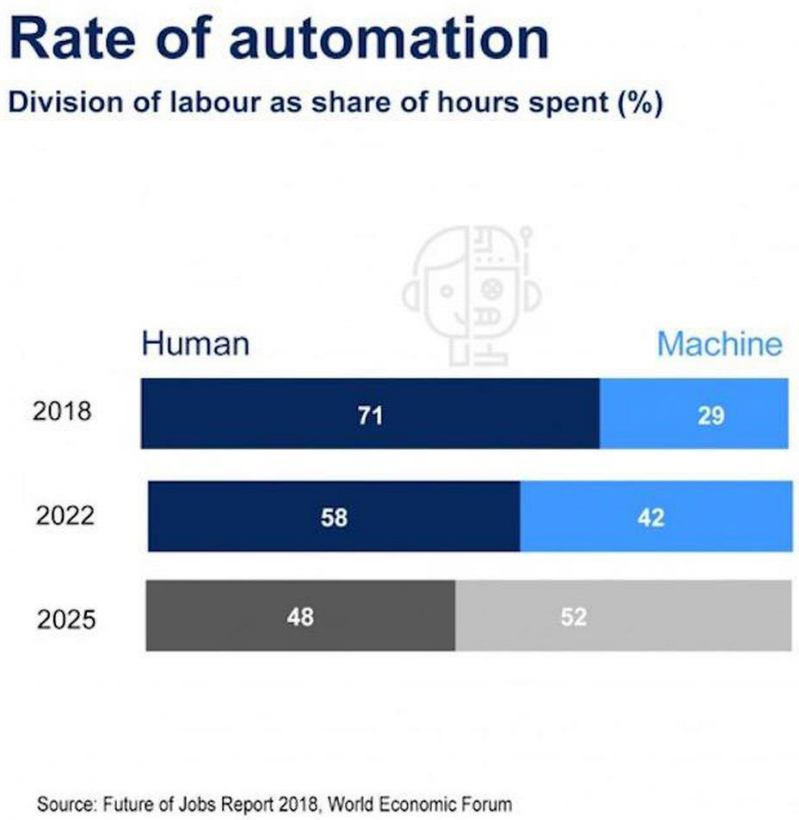


experian™

Take away Point 2: AI will change your future jobs

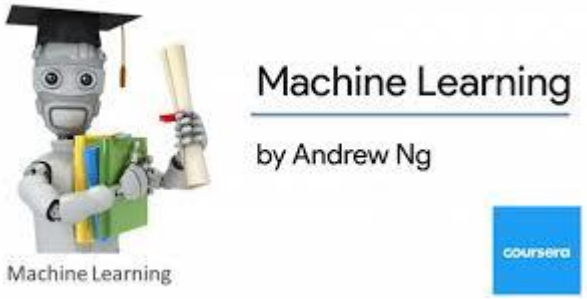
Whether you go into the field or not, it is very likely you will interact with AI in your future jobs





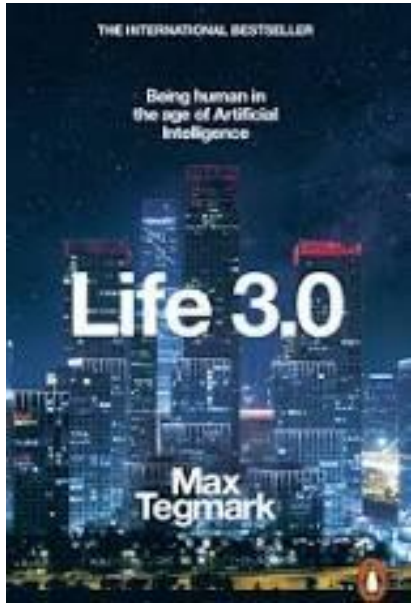
Take away Point 3: How to get started now!

Below there are some good books and links to get started !

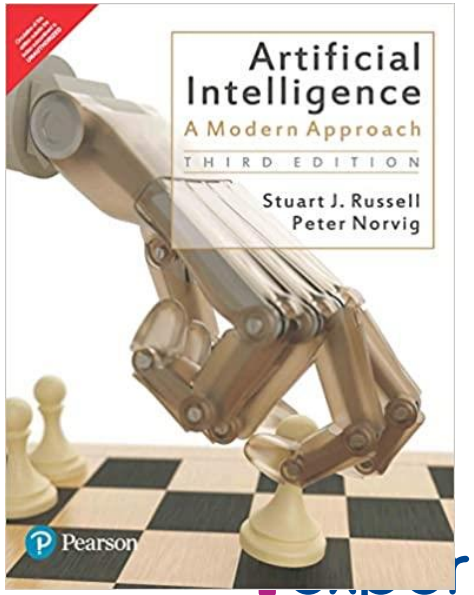


- Free interactive online course: https://www.coursera.org/learn/machine-learning -

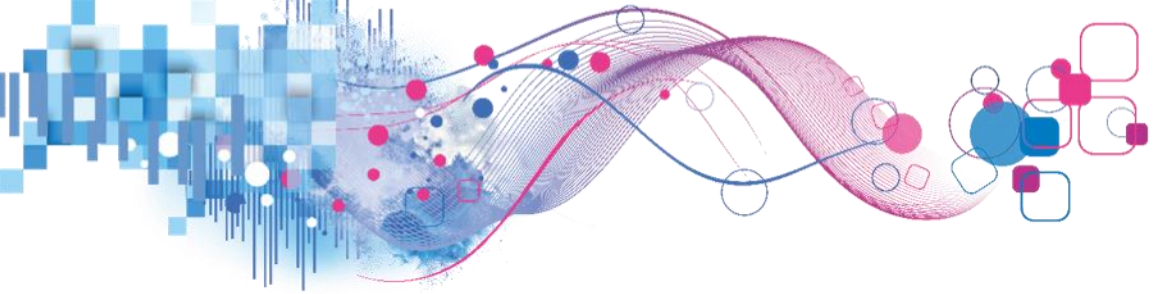
- Future of AI -



- Degree textbook, Advanced (loads of math !)-



- https://www.kaggle.com/alexisbcook/titanic-tutorial -



experian™

AI applications in Financial services

Artificial Intelligence has been defined as the “new electricity “. It is now approaching an inflection point that it could be compared to the impact of electricity back in the 1800s. Before electricity, any factory requiring [steam] power had to be located close to a coal mine. But electricity allowed power to be decentralized and distributed remotely, far away from the source. This, in turn, changed the way industry and then society was able to locate and organize. The same is happening with the transformational potential of AI. It will enable distributed decision-making in every part of an organization as AI-supported insights are pushed to the frontline

And indeed within Financial services, AI is enabling incumbent financial institutions to deliver smarter and more secure services to their clients and customers. Firms are also using AI solutions to create robust fraud detection and prevention systems, accelerate risk calculations and fraud detection

experian™