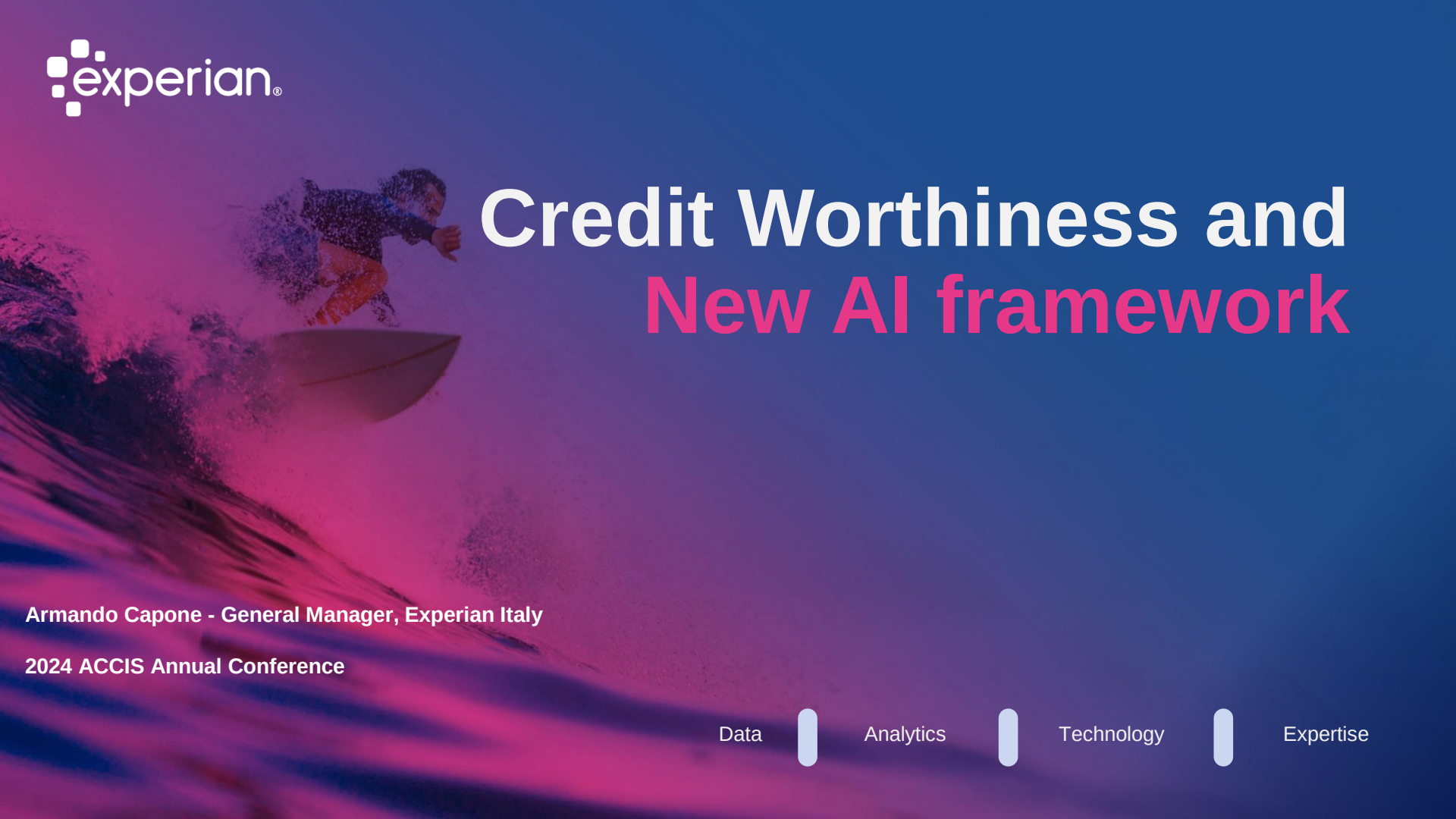


Credit Worthiness and New AI framework

Armando Capone - General Manager, Experian Italy

2024 ACCIS Annual Conference

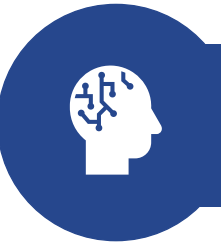
Data | Analytics | Technology | Expertise

Agenda

- The definition of AI and the risk-based approach
- From AI to GenAI
- How Experian uses AI?
- A new frontier: Synthetic Data Modelling
- Main Case Study – Synthetic Data for Acquisitions
- Q&A

The definition of AI and the risk-based approach

(By the European Commission, aligned with the OECD proposal)



The Act defines an “**AI system**” as a "machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments”.

Minimal risk	Applications such as spam filtering or video games are deemed to carry a minimal risk and as such, they are not subject to further regulatory requirements	**Code of conduct**
Limited risk	Some examples are image and video processing, recommender systems, and chatbots.	**Transparency**
High risk	Services directly affecting citizens’ lives (e.g., evaluating creditworthiness or educational opportunities, applications applied to critical infrastructure).	**Conformity Assessment**
Unacceptable risk	Some AI applications such as social scoring systems or manipulative systems potentially leading to harm are outlawed completely.	**Prohibited**

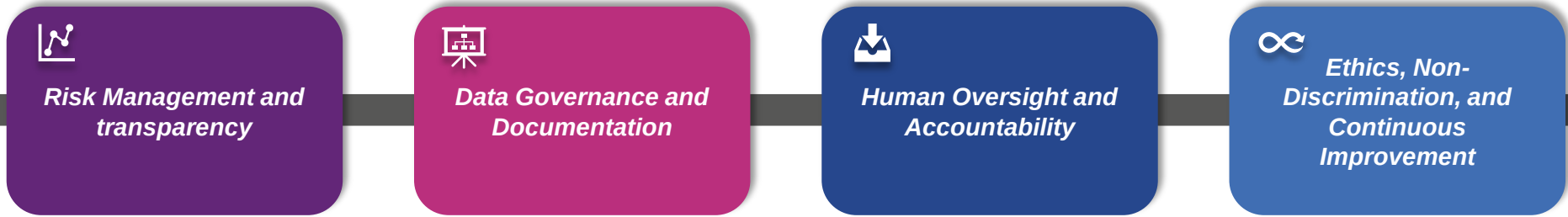
Credit Worthiness solutions will be high risk

Today, **credit scoring models** - based on AI or simpler mathematical techniques - are subject to key regulations:

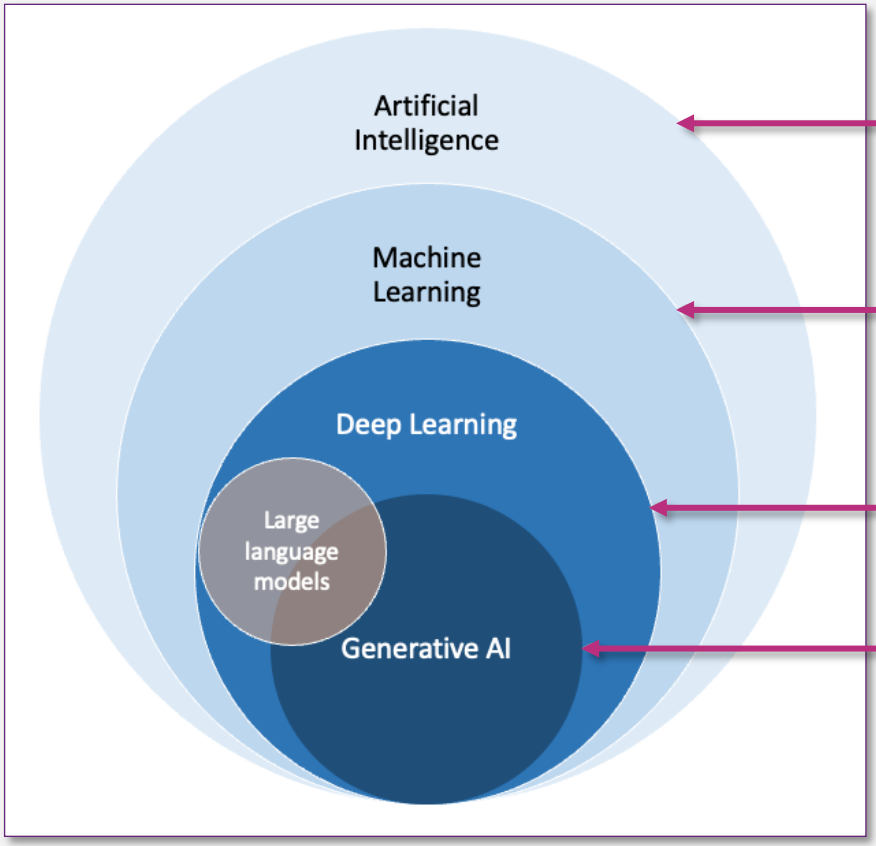
- ***Sectorial financial regulations***
- ***Horizontal legislation (GDPR)***
- ***National level regulations***

All these provide a **clear legal framework** to ensure that credit scoring models produce accurate, and avoid biased or discriminatory, outcomes and for consumers to ensure a **high level of protection of including their fundamental rights**

✓ *The new **EU AI Act**, aligned with ethical and responsible AI use, emphasizes **4 main principles**:*



From AI to GenAI



Broader field, that includes traditional models

Is a subfield of AI

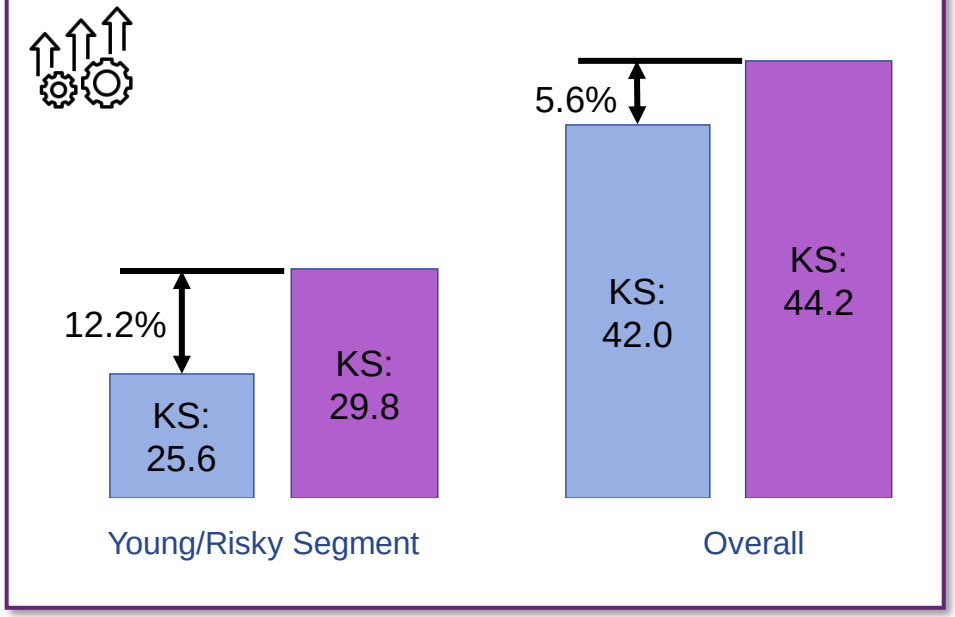
Is a subfield of ML

Is a subfield of DL and a particular application of AI

How Experian uses AI?

Improving over traditional logistic regression/scorecard-based approach

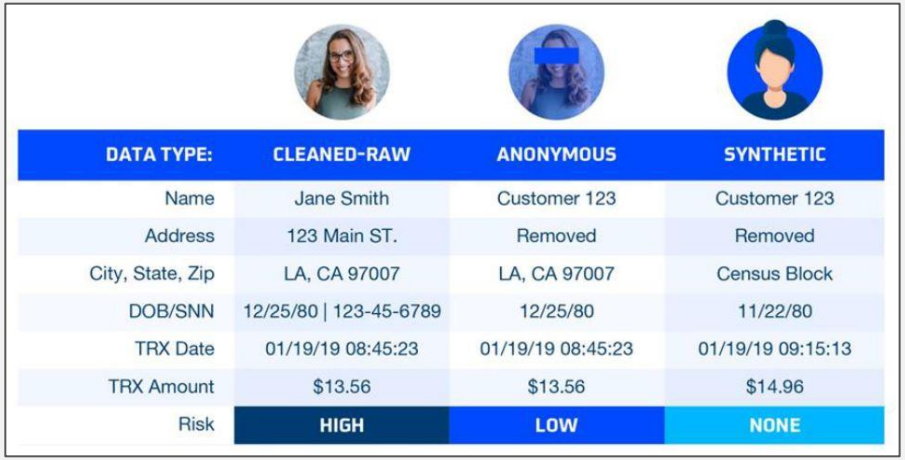
- Credit scoring has traditionally been scorecard-based or logistic-regression based
- ML solution can better ensure regulatory requirements (e.g. monotonic requirements)
- Demonstrated performance gains:
  - **12.2% more accurate scoring in young/risky segment segments contributing to financial inclusion**
  - **5.6% more accurate scoring in overall database**



A new frontier: Synthetic Data Modelling

Synthetic data:

- mimics the original data
- removes the personal or other sensitive information
- retains the accuracy and statistical significance
- ideal for testing, modeling, projecting trends, and more



DATA TYPE:	CLEANED-RAW	ANONYMOUS	SYNTHETIC
Name	Jane Smith	Customer 123	Customer 123
Address	123 Main ST.	Removed	Removed
City, State, Zip	LA, CA 97007	LA, CA 97007	Census Block
DOB/SNN	12/25/80 | 123-45-6789	12/25/80	11/22/80
TRX Date	01/19/19 08:45:23	01/19/19 08:45:23	01/19/19 09:15:13
TRX Amount	$13.56	$13.56	$14.96
Risk	HIGH	LOW	NONE

Source: Facteus: Changing the Game With Synthetic Data, June 2021

Why Synthetic Data is important for the future of AI?

*”Such data can become **unifying bridge** between policy support computational models … and become **the key enabler of artificial intelligence in business** and policy applications in Europe. Synthetic data have potential help **controlling unevenness and bias** in algorithmic governance and enable better targeted policies with **small regulatory footprint**” **

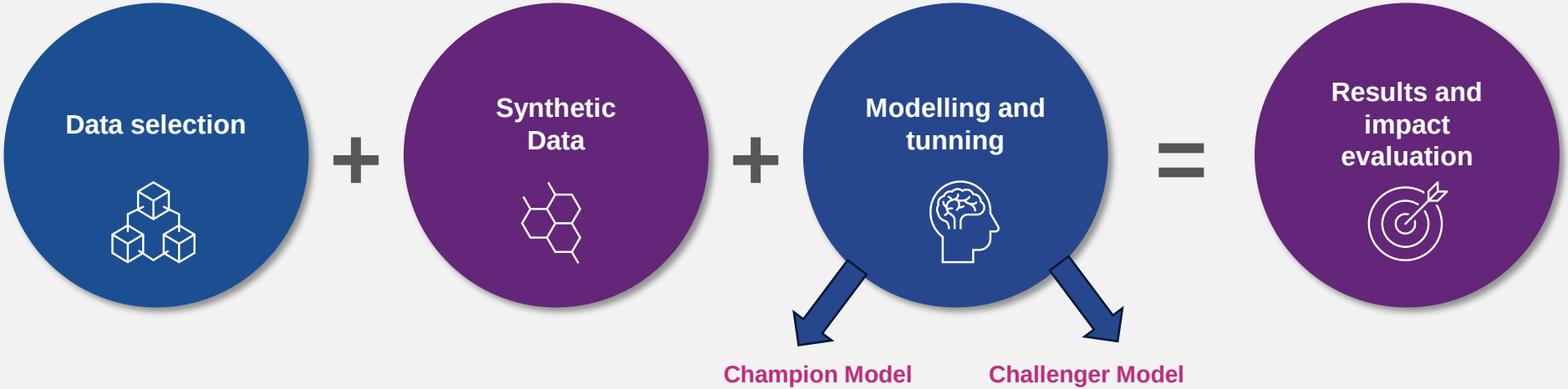
- It is disconnected from PII. Therefore, it could be stored for future purposes.
- It can be freely shared, monetized and used to create new revenue streams

- Incorporating domain knowledge into the training of AI models to improve the quality of the model’s predictions
- Generates optimal data, to increase fairness and reduce bias from models

- Augmented data sets: BNPL, transactional data
- Increasing consumers trust and following compliance guidelines for AI development

** European Commission's Joint Research Centre*
https://publications.jrc.ec.europa.eu/repository/handle/JRC128595

Main Case Study – Synthetic Data in Credit Risk



Benchmarking: Can an ML model built over synthetic data perform well?

Results: 15% reduction in losses due to false positive reduction and better acceptance

Evidence of fairer models with Synthetic Data

Traditional Model

Top 5 Features (% Contribution)

1. Age (29%)
2. Debt amount (10%)
3. Bureau Data A (9%)
4. Debt ratio (9%)
5. Bureau Data B (8%)

Total Contribution in top 5 = 65%

Discrepancies in feature importance between models may indicate potential biases

Synthetic Data Model

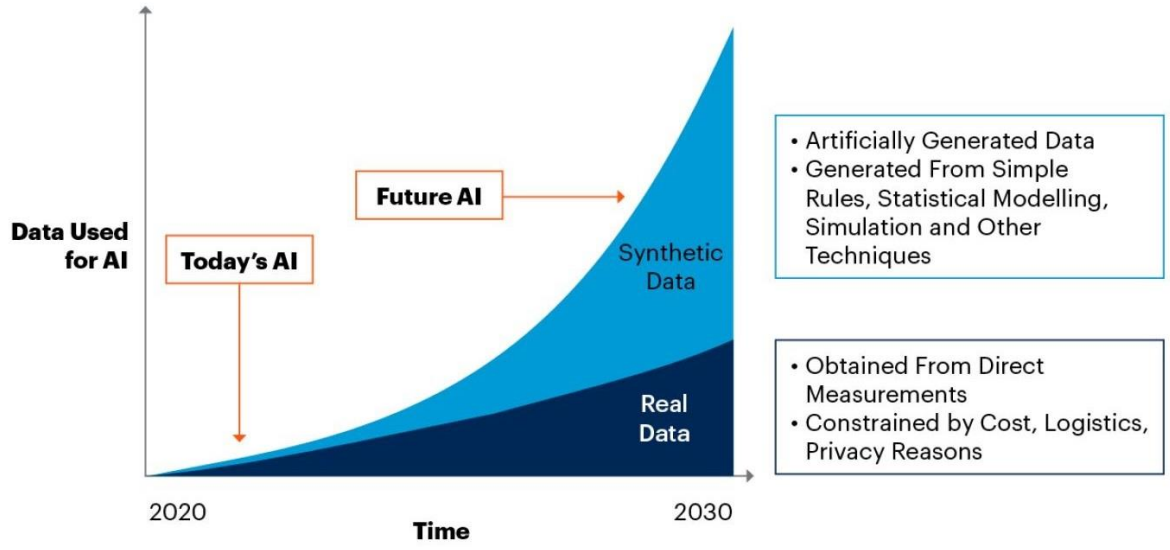
Top 5 Features (% Contribution)

1. Debt Amount (10%)
2. Debt ratio (9%)
3. Bureau Data A (8%)
4. Age (8%)
5. Bureau Data B (8%)

Total Contribution in top 5 = 43%

More balanced and fairer model

And much more can be done!



Source: Maverick Research: Forget About Your Real Data - Synthetic Data Is the Future of AI, June 2021*

The next leap in AI efficacy cannot be achieved without taking advantage of the synthetic data techniques into the training of AI models

Use cases

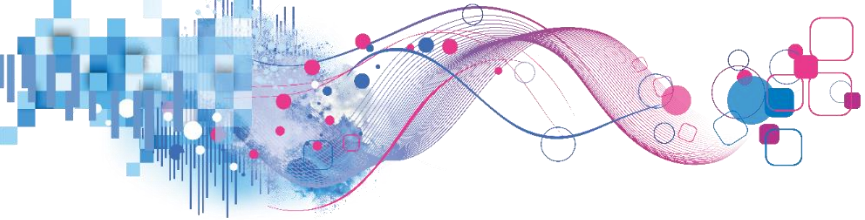
- Credit Risk modelling
- Fraud Detection
- Stress Testing
- AML/Regulatory Compliance

Jobs to be done

- Data Augmentation
- ML Training and Validation
- Scenario Analysis
- Anomaly Detection
- Data Privacy

Verticals

- Retail Banking
- BI
- Insurance
- Credit Card
- Telcos



experian™