Explainable Artificial Intelligence (XAI)

Opening the Black Box – Making AI Understandable for Everyone

Keynote Speaker

Asso. Prof. Dr. El Arbi Abdellaoui ALAOUI *(IEEE Member)*

M2IP Team, ENS, Moulay Ismaïl University of Meknes, Morocco

e.abdellaoui@ieee.org

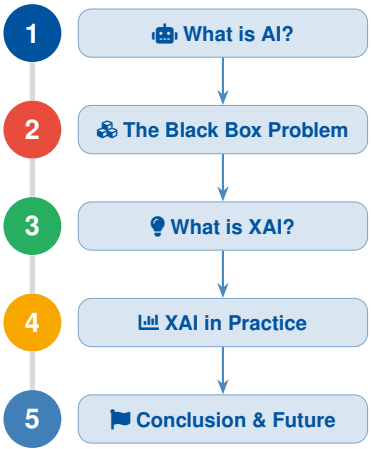
5th International Conference on Engineering and Technology

ICOEAT 2026

Universitas Alma Ata – Indonesia

February 6th, 2026

Outline



- AI in daily life
- How machines learn

- Why we can't trust what we can't understand
- Real-world consequences

- XAI definition & intuition
- Key methods (SHAP, LIME, . . .)

- Case study: Banking
- SHAP visualizations

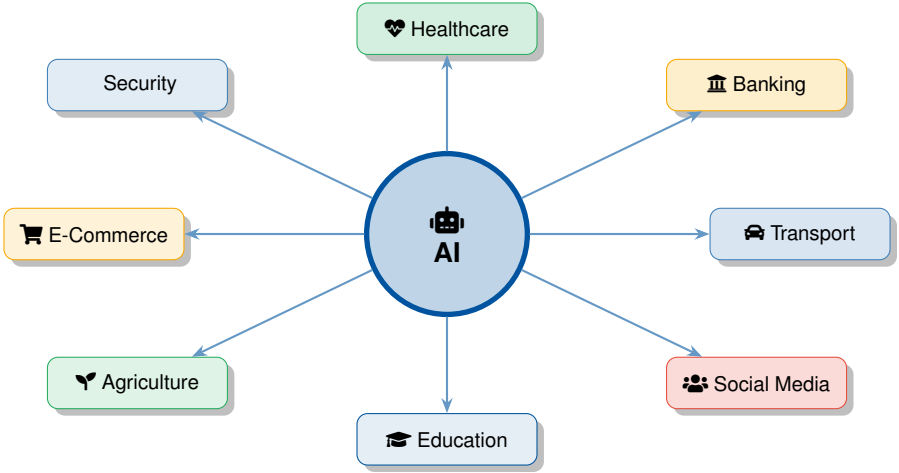
- Key takeaways
- Open challenges

PART 1: WHAT IS AI?

Key Points

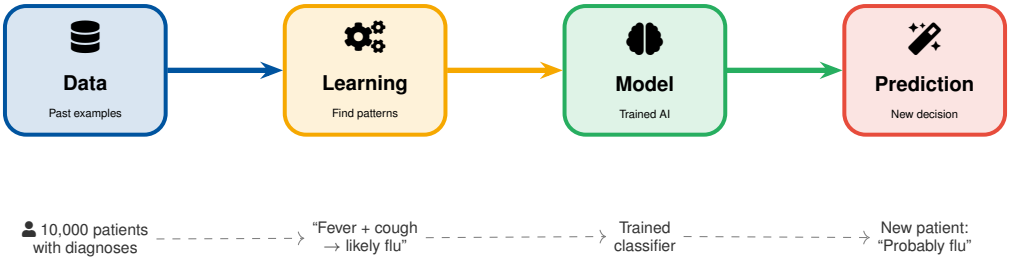
- AI is everywhere: phones, cars, hospitals, banks
- Machines learn patterns from data to make decisions
- But how do they actually decide?

AI is Everywhere Around Us



⚠ AI makes **thousands of decisions** about us every day – but do we understand **why**?

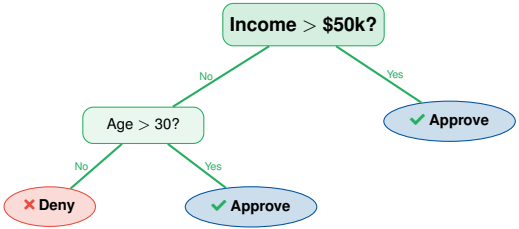
How Does AI Learn? – A Simple View

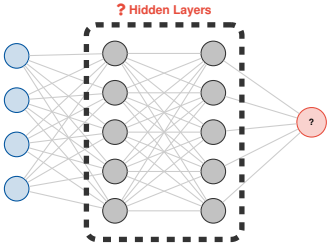


Key Idea

AI learns from **past examples** (Data) to make **future predictions** – just like we learn from experience!

Simple vs. Complex AI Models





Interpretable Model

Easy to understand

We can read the rules!

Black-Box Model

Very accurate but...

Nobody knows WHY!

PART 2: THE BLACK BOX PROBLEM

Key Points

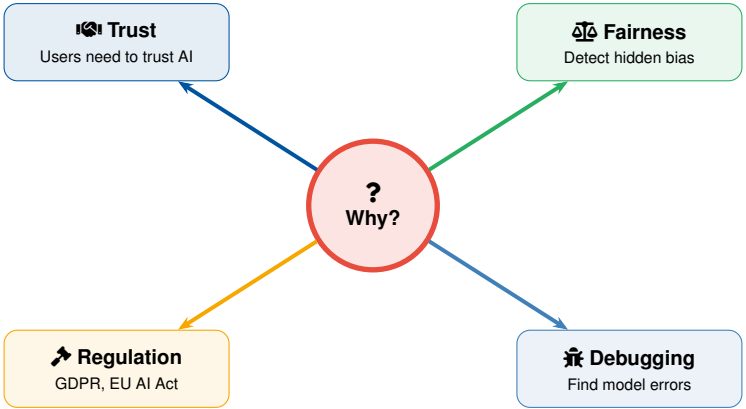
- Complex AI models are opaque – we can't see inside
- Real consequences: denied loans, wrong diagnoses, unfair decisions
- Trust requires understanding

The Black Box Problem

The Core Problem

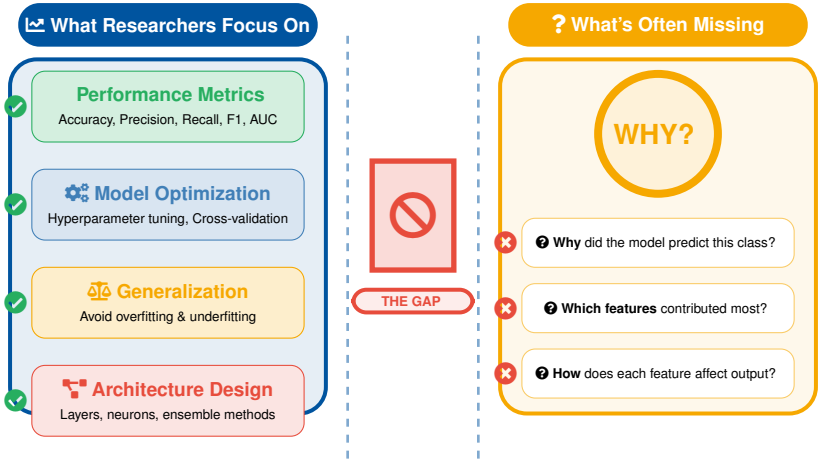
The model says “High Risk” but **cannot explain** its reasoning. Would you trust this for your health? Your loan? Your freedom?

Why Does Explainability Matter?



The EU AI Act (2024) **requires** explanations for high-risk AI systems in healthcare, finance, and justice.

The Research Gap: Performance vs. Understanding



High accuracy is NOT enough! We need to understand **WHY** and **HOW** models make decisions.

When Black Boxes Go Wrong – Real Examples

Healthcare

An AI system used by hospitals was found to **systematically underestimate** health risks for Black patients – affecting millions.

Source: Science, 2019

Criminal Justice

COMPAS recidivism tool was shown to be **biased against minorities** – with no way to audit why.

Source: ProPublica, 2016

Finance

Credit scoring AI denied loans to qualified applicants based on **zip code correlations** with race.

Source: The Markup, 2021

Hiring

Amazon's AI recruiting tool **penalized women's resumes** – learning bias from historical data.

Source: Reuters, 2018

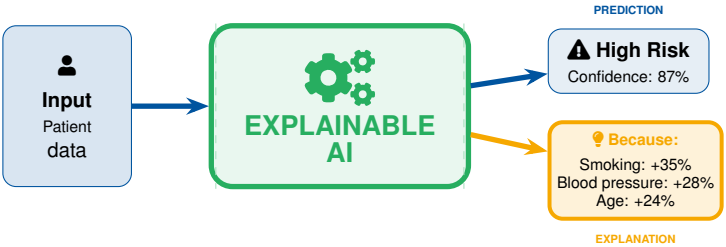
Without explanations, we cannot detect, debug, or fix these problems!

PART 3: WHAT IS XAI?

Key Points

- XAI = Methods that make AI decisions understandable to humans
- Two types: Global (overall behavior) and Local (single prediction)
- Key methods: SHAP, LIME, Attention, Grad-CAM, ...

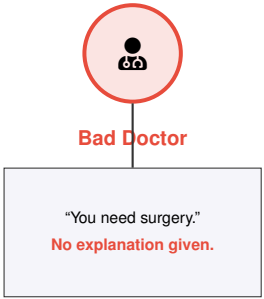
Explainable AI (XAI) – The Solution

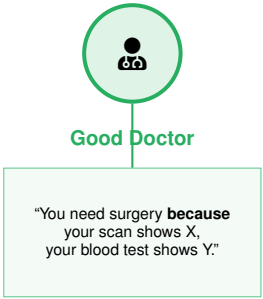


XAI in One Sentence

XAI = AI that can **explain its reasoning** in a way humans can understand, verify, and trust.

XAI Analogy – Think Like a Doctor





XAI transforms AI from a "bad doctor" into a "good doctor" – providing **reasons** alongside decisions.

Two Types of Explanations

GLOBAL Explanation

“What matters most overall?”

- Which features are most important?
- How does each feature generally affect predictions?
- Are there patterns across all customers?

Like understanding a doctor’s general approach

LOCAL Explanation

“Why THIS specific decision?”

- Why was this customer approved/denied?
- What pushed the decision for/against?
- Which factors were decisive?

Like asking “why did you prescribe ME this?”

Local vs Global XAI: Understanding the Difference

Local XAI

Individual Predictions

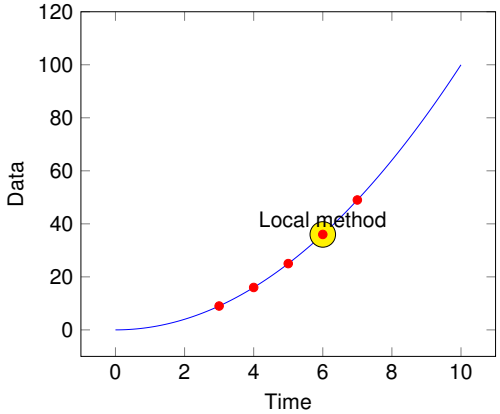
1

Single Client

Why one specific customer might subscribe to term deposit?

- Contract type: +0.3
- Monthly charges: +0.2
- Tenure: -0.1
- Internet service: +0.15

Use Case: Personalized marketing approach



Local vs Global XAI: Understanding the Difference

Global XAI

Overall Model Behavior

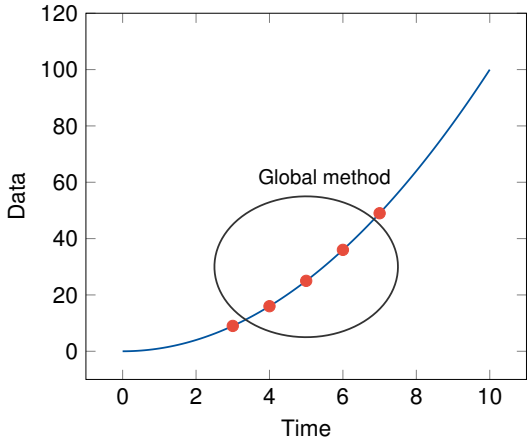
ALL

All Customers

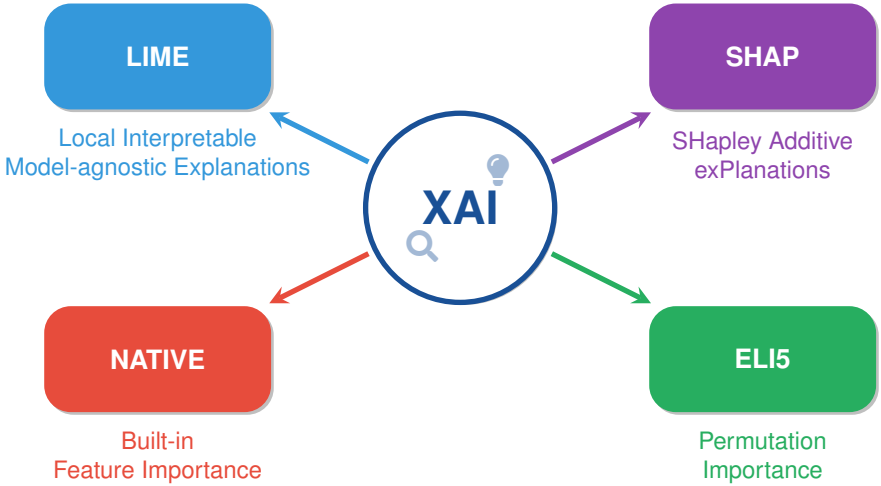
What features matter most across all predictions?

1. Contract type (0.25)
2. Tenure (0.18)
3. Monthly charges (0.15)
4. Internet service (0.12)

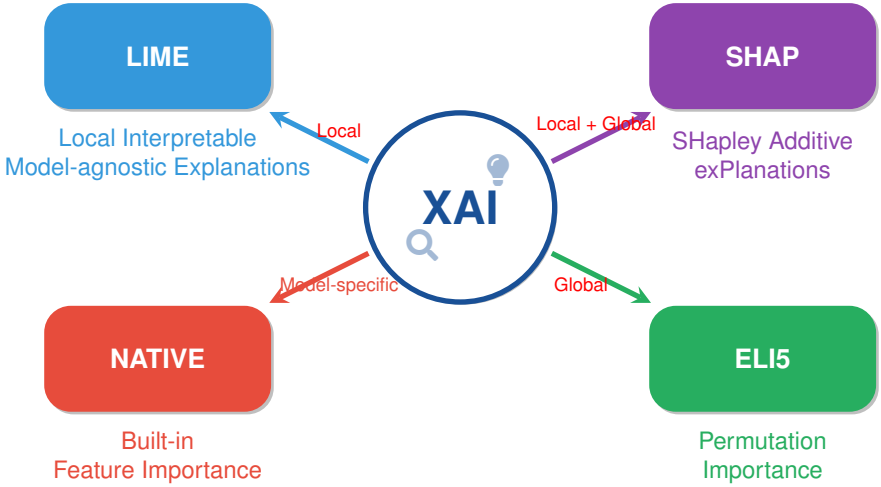
Use Case: Business strategy planning



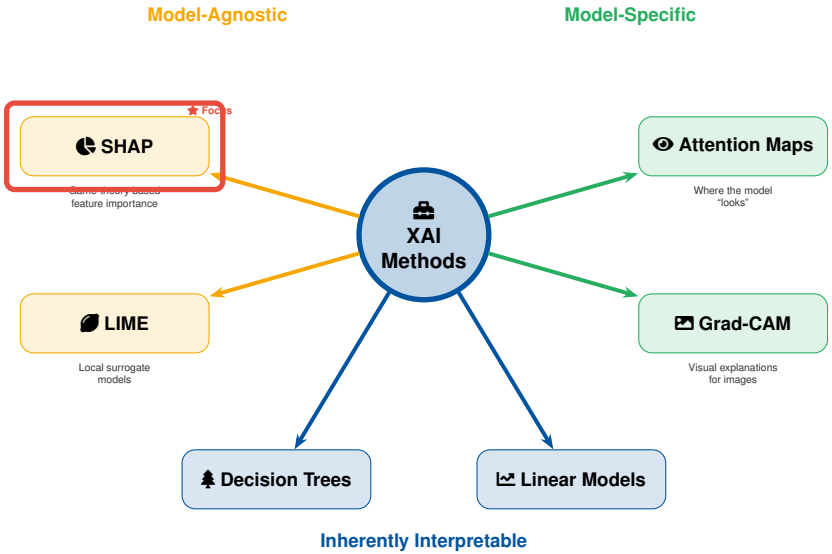
XAI Methods – The Toolbox



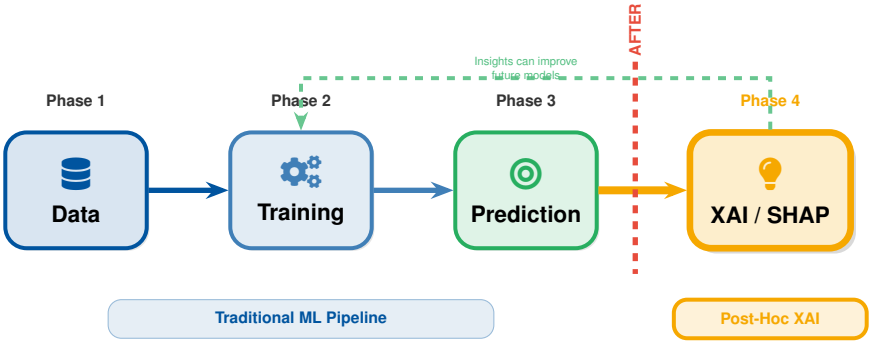
Explainable AI Methods



XAI Methods – The Toolbox



Where Does XAI Fit? – Post-Hoc Explanation

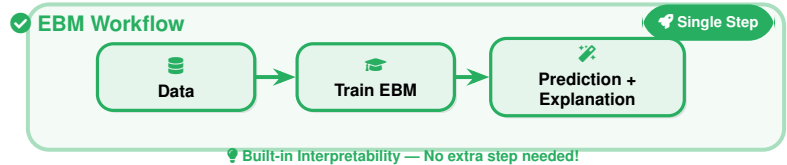


Key Point: XAI is a Post-Hoc Method

SHAP and most XAI techniques are applied **after** the model is trained and predictions are made.

They explain **existing predictions** — they don't change how the model works.

Implementation Workflows: EBM vs SHAP



EBM: Faster, simpler, but limited to specific models | **SHAP:** Flexible, powerful, works with CatBoost, XGBoost, Neural Networks...

EBM = Explainable Boosting Machine

SHAP – Explained Simply

The Idea Behind SHAP

Imagine a team of players scoring a goal. **SHAP** answers: “How much did **each player** contribute to winning?”

→ **Green** = pushes toward YES → **Red** = pushes toward NO

$$f(x) = \text{Base Value} + \phi_{\text{Duration}} + \phi_{\text{Housing}} + \phi_{\text{Age}} + \dots \quad \text{(Each } \phi \text{ is a SHAP value)}$$

SHAP – Explained Simply

Shapley Value from Game Theory

Each feature is a “player” contributing to the prediction outcome.

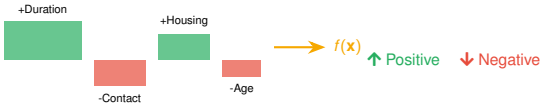
Additive Property

$$f(\mathbf{x}) = \phi_0 + \sum_{i=1}^{d} \phi_i$$

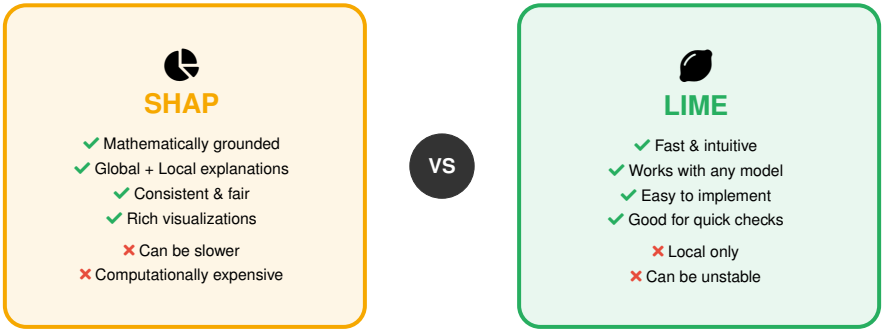
where $\phi_0 = \mathbb{E}[f(X)]$

Shapley Value

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \Delta_i(S)$$



SHAP vs. LIME – Quick Comparison



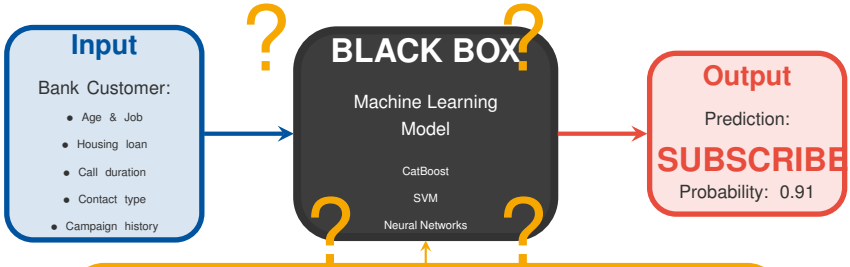
In practice, **SHAP** is often preferred for thorough analysis; **LIME** for quick prototyping.

PART 4: XAI IN PRACTICE – A CASE STUDY

Key Points

- Real example: Predicting bank term deposit subscriptions
- CatBoost + SHAP: Accuracy meets transparency
- 7 SHAP visualization types for full understanding

Why XAI? The Black Box Problem in Banking ML



The Problem:
We get accurate predictions but **NO** understanding of **WHY** the customer will subscribe

Trust
Build banker confidence

Compliance
Meet financial regulations

XAI Solution: Opening the Black Box

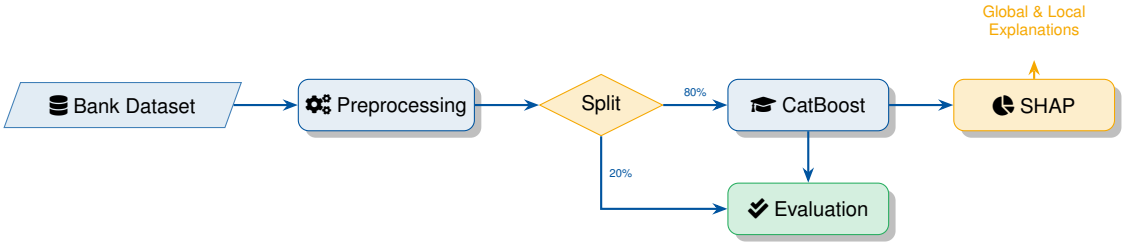
XAI methods provide explanations that reveal:

- Which features influenced the subscription decision most
- How each feature contributed to the final prediction
- Why this customer is likely to subscribe to term deposit

Targeting
Enable focused marketing

Optimization
Improve model performance

Case Study: Bank Marketing Prediction



Data

45,211 customers
Portuguese bank
Marketing campaign

Model

CatBoost classifier
Gradient boosting
Handles categories

XAI

Full SHAP analysis
7 visualization types
Global + Local

Model Performance – Strong Foundation

Model	Accuracy	Precision	F1-Score	AUC
CatBoost	**0.8646**	0.8517	0.8670	**0.9326**

Strong model performance = **reliable foundation** for meaningful SHAP explanations.

SHAP Analysis – 7 Views

7 complementary views to understand predictions at **global** and **local** levels

GLOBAL Explanations

- **Beeswarm** – Feature importance + distribution
- **Bar Plot** – Ranked mean |SHAP|
- **Heatmap** – Patterns across instances
- **Dependence** – Feature interactions

LOCAL Explanations

- **Waterfall** – Step-by-step breakdown
- **Force Plot** – Push/pull forces
- **Decision** – Cumulative path

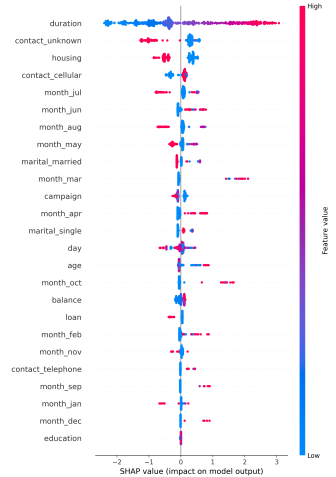
Explain **individual** customer decisions

GLOBAL EXPLANATION

How does the model behave **overall**?

Beeswarm • Bar Plot • Dependence • Heatmap

Global: Beeswarm Plot

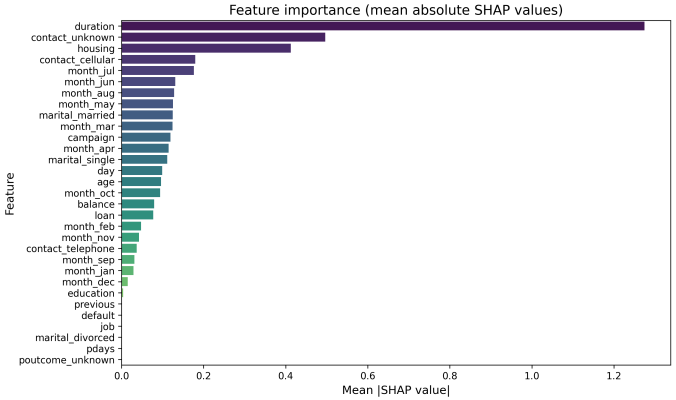


Key Observations

- **Duration** has largest impact
- Contact method is critical
- Housing status matters

High = Red
Low = Blue

Global: Feature Importance (Bar Plot)



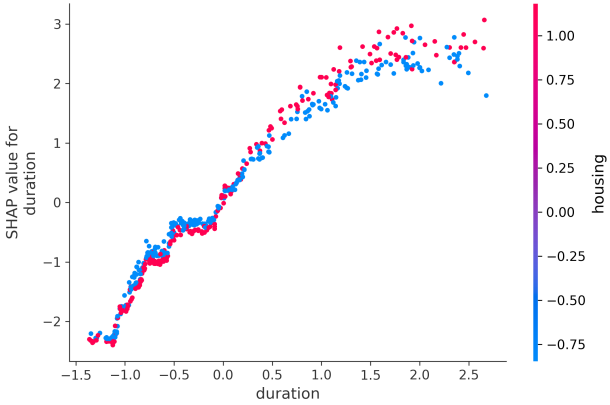
Top 3 Features

1. **Duration** (≈1.25)
2. contact_unknown (≈0.5)
3. housing (≈0.4)

Insight

Duration dominates all other features in predictive importance.

Global: Dependence Plot – Duration



Observation

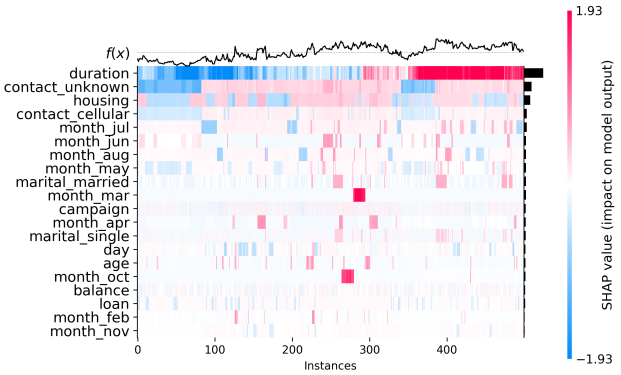
Strong **positive correlation** between duration and SHAP value.

Business Insight

Longer calls ⇒ Higher subscription probability

Color shows **housing** interaction effect.

Global: Heatmap



What It Shows

SHAP values across **500 instances**

Red = positive impact
Blue = negative impact

Pattern

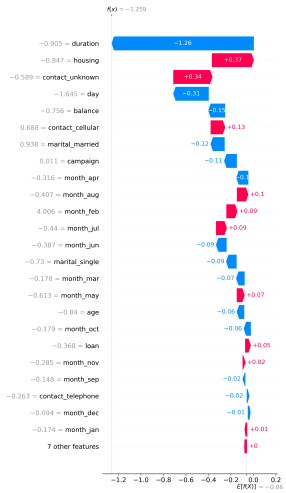
Duration shows consistent strong influence across all customers.

LOCAL EXPLANATION

Why this prediction for **one customer**?

Waterfall • Force Plot • Decision Plot

Local: Waterfall Plot – Single Prediction



Interpretation

For this customer:

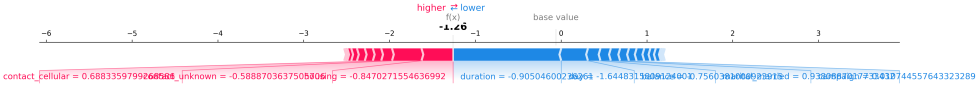
- Duration: **-1.26** (negative)
- Housing: **+0.37** (positive)
- Contact: **+0.34** (positive)

Final Prediction

$f(x) = -1.259$

Short call duration strongly decreased subscription probability.

Local: Force Plot



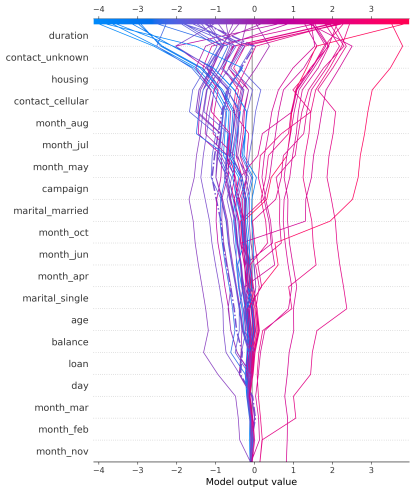
← Pushes Lower

Features in red push prediction toward subscription (YES)

→ Pushes Higher

Features in blue push prediction away from subscription (NO)

Local: Decision Plot



Cumulative Impact

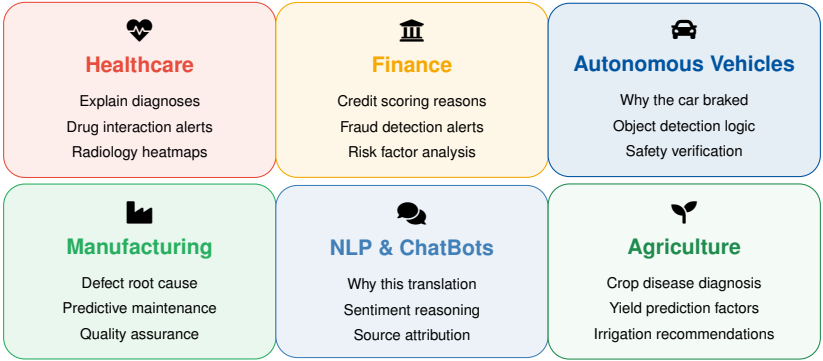
Each line = one customer's prediction path through features.

Key Finding

Duration causes the largest initial spread in decision paths.

Pink → YES
Blue → NO

XAI Applications Across Domains



PART 5: CONCLUSION & FUTURE DIRECTIONS

Key Points

- XAI bridges the gap between accuracy and understanding
- Open challenges: scalability, evaluation, user trust
- The future: XAI as a standard requirement for all AI systems

Key Takeaways

AI without explanation is not trustworthy. Trust requires transparency.

XAI methods like SHAP make black boxes transparent. Grounded in math.

Global + Local explanations give the full picture. Both matter.

Regulation is coming – XAI is not optional. EU AI Act, GDPR.

XAI benefits everyone: developers, users, and society. Better AI for all.

Open Challenges in XAI

Scalability
SHAP can be slow for very large models (billions of parameters).

Adversarial Attacks
Can explanations be manipulated to hide bias?

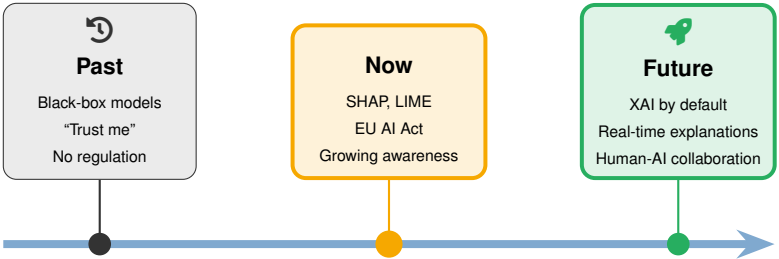
Evaluation
How do we measure if an explanation is actually "good"?

Multimodal XAI
Explaining AI that processes text, images, and audio together.

User Studies
Do non-experts actually understand these explanations?

LLM Explainability
How to explain ChatGPT-like models with billions of parameters?

The Future of XAI



Vision

The goal is not just AI that **works** – but AI that **works and explains itself**.
Explainability should be a **feature, not an afterthought**.

My Research Contributions in XAI

NEW – Applied Soft Computing (Elsevier)

"Evolving Explainable Neural Architectures: Dynamic Behavioral-Novelty Weighting in Multi-Objective Neuroevolution for Multimodal Spam Detection"

Information Sciences (Elsevier) – Make XAI more faster for different Applications

IEEE Access – Predicting Student Performance Using SHAP Explanations

Taylor & Francis – Explainable Web Attack Detection for Industrial IoT

Springer – SHAP for Parking Occupancy & Fake Account Detection

XAI Application Domains:

10+
XAI Papers

A Final Thought

> *“If you can’t explain it simply,*
> *you don’t understand it well enough.”*
>
> – Often attributed to **Albert Einstein**

The same applies to AI. If AI can’t explain its decisions, we shouldn’t blindly trust them.

Thank you:

- Any question ?

Institutions:

Speaker

El Arbi ABDELLAOUI ALAOUI

ENS of Meknes – UMI, Morocco

e.abdellaoui@ieee.org

www.abdellaoui.net