

XAI: EXPLAINING HOW AI MAKES DECISIONS

Have you ever unlocked your phone with just a glance? Or scrolled through a TikTok feed that seems to read your mind? Maybe you've used Google Maps to navigate a new city or seen an eerily perfect product suggestion on Amazon.

Each of these moments is powered by Artificial Intelligence (AI), a silent partner in our digital lives. AI works in the background, learning from data to make predictions, recommendations, and decisions. You've probably even had a direct conversation with ChatGPT to explain tricky concepts for a homework assignment. How does it craft a complete answer without you writing even a single line of code? What's really happening behind the screen?

From Movie Picks to Life Changing Decisions

Sometimes, an AI's decision is low-stakes, like recommending a funny cat video. But increasingly, AI is involved in decisions with serious consequences.

Imagine you apply for a competitive summer internship. A week later, you get the email: *"After careful consideration, we regret to inform you that we will not be moving forward with your application."*

You're left wondering: *Why?* Was it my GPA? My college? My lack of prior experience?

If a human recruiter made the decision, you could potentially ask for feedback. But if an automated AI screening system rejects your application, you're met with a digital wall of silence. You can't ask why unless the AI is designed to talk back.

The Problem of the "Black Box"

Without the ability to explain itself, an AI can feel like a "black box." We see the input (your application) and the output (your application getting rejected) but the reasoning inside is a complete mystery. Similarly, in other domains:

- Movie Recommender: "You should watch *The Matrix*." (But *Why?*)
- Scholarship Application: "Not selected." (Based on what criteria?)

This lack of transparency can lead to mistrust, frustration, and most importantly, unfair or biased outcomes. That is where the field of Explainable AI (XAI) comes in.

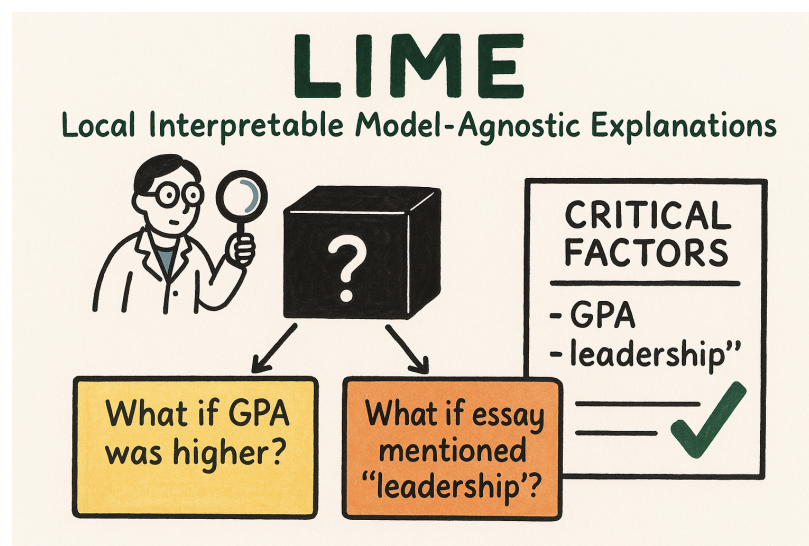
How XAI Shows its Work:

The goal of XAI is to open the black box. Think of it like your math teacher insisting that you “show your work” and the final answer is not enough, i.e. the process of deriving the answer is equally important.

AI researchers have developed clever detective tools to do this. Instead of a single method, they have a whole toolkit. Two of the most foundational tools in the toolkit are called LIME and SHAP:

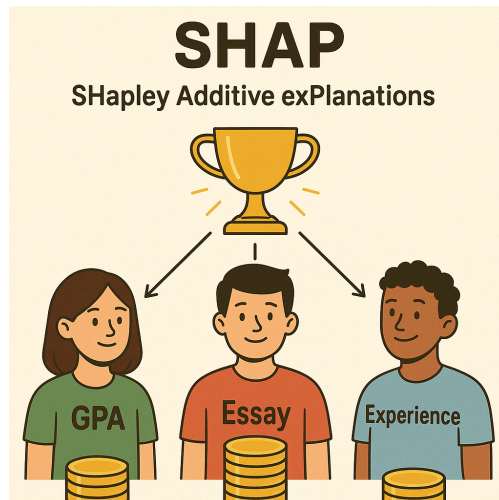
1) LIME (Local Interpretable Model - Agnostic Explanations)

LIME works by acting like a scientist running a focused experiment. To understand one specific decision (like your internship rejection), it slightly changes the inputs and watches what happens. What if your GPA was a bit higher? What if your essay mentioned “leadership” instead of “collaboration”? By observing how the AI’s verdict wobbles with these small tweaks, LIME figures out which factors were the most critical for “your” specific case. It’s called “local” because it explains just one decision at a time, not the AI’s entire decision system.



2) SHAP (SHapley Additive exPlanations)

SHAP borrows a Nobel Prize-winning idea from game theory. Imagine a team of gamers wins a tournament prize. The question is how do you divide the prize money fairly based on each person’s contribution? SHAP does this for AI. It treats every piece of your input information (GPA, essay keywords, prior experience etc.) as a “player” of the team. It then calculates the precise contribution of each “player” to the final outcome. This gives a holistic and mathematically sound view of what mattered most.



While their approaches are different, both tools turn a complex decision into a simple, human-readable report card.

Now let's revisit that movie recommendation but explained by one of these tools:

We recommend *The Matrix* because:

- You loved the movie *Inception* (major positive influence)
- You consistently rate action and sci-fi films highly (positive influence)
- You tend to dislike romantic comedies (minor negative influence, pushing away other options)

Suddenly, the decision makes sense. The mystery is gone, replaced by logic we can understand and evaluate.

Try It Yourself

Imagine an interactive tool where you could tell an AI your favorite movie genres, get a recommendation, and instantly see a bar chart explaining *why*: green bars show the factors pushing for the recommendation and red bars show factors pushing against it. If you change your favorite genre from “Action” to “Comedy,” you’d see the chart and the recommendation update in real-time.

This is exactly what AI researchers and engineers do. They use XAI tools to audit their models, making sure they’re making decisions for the right reasons and not based on flawed or biased patterns.

Why This Matters for Everyone

As AI is integrated into automated essay grading, scholarship awards, and hiring, the stakes become incredibly high. If we can't understand how AI reaches its conclusions, we can't trust it with decisions that shape our lives. And if it makes a mistake, we have no way to challenge it.

XAI helps by:

- **Creating Transparency:** It shines a light on the decision-making process.
- **Catching Bias:** It helps us spot when an AI is making unfair decisions based on factors like zip code, gender, or race.
- **Building Trust:** It builds a bridge of understanding between humans and the machines we rely on.

A Quick Reality Check

Explanations aren't a silver bullet. They can sometimes oversimplify the AI's complex process. More critically, a biased AI can produce an explanation that sounds logical but still justifies an unfair outcome - like a politician spinning a bad decision.

Because of this, XAI is best seen as a tool for interrogation, not a guarantee of fairness. It empowers us to ask better questions.

The Takeaway

When AI can "talk back," it gives us the power to understand, question, and ultimately improve its decisions. The next time you get a recommendation from a machine, ask yourself: Why did it decide that?

Imagine a future where every important automated decision could explain itself in plain English. Wouldn't that create a smarter, fairer world for everyone?



XAI Explained: How AI Makes Decisions

Understanding AI Through Rules, Trees, and Feature Importance

Sarah's Internship Application

GPA
3.2 / 4.0

Experience
6 months

Skills
Python, SQL

Projects
2 projects

Cover Letter
Generic



Black Box AI

"I process your data through complex algorithms..."

❌ REJECTED
(No explanation given)

CLICK TO REVEAL



HOW AI DECIDES



XAI - Explainable AI

"Let me show you exactly how I make decisions using:

- Decision Rules (If-Then logic)
- Decision Trees (Step-by-step choices)
- Feature Importance (What matters most)
- Attention Weights (Where I focus)

Decision Rules (If-Then Logic)

How AI thinks in simple rules:

IF GPA \geq 3.5:
ADD 25 points
✅ Sarah: 3.2 \rightarrow 0 points

IF Experience \geq 1 year:
ADD 20 points
❌ Sarah: 6 months \rightarrow 0 points

IF Skills include (Python OR Java):
ADD 15 points
✅ Sarah: Python \rightarrow 15 points

IF Projects \geq 3:
ADD 10 points
❌ Sarah: 2 projects \rightarrow 0 points

Sarah's Total: 15/70 points \rightarrow REJECTED

Decision Tree (Step-by-Step)

AI follows this decision path:

GPA \geq 3.5?

❑ NO (Sarah: 3.2)

Experience \geq 1 year?

❑ NO (Sarah: 6 months)

Skills score \geq 15?

❑ YES (Sarah: Python+SQL)

Projects \geq 3?

❑ NO (Sarah: 2)

REJECT

Path taken:
GPA \rightarrow Experience \rightarrow Skills \rightarrow Projects \rightarrow REJECT

Feature Importance (What Matters Most)

AI ranks features by importance:

GPA	90% Important
Experience	80% Important
Skills	60% Important
Projects	50% Important
Cover Letter	40% Important

💡 Tip: Focus on improving GPA and gaining experience first!

Attention Weights (Where AI Focuses)

AI attention on Sarah's application:

"Computer Science major with 3.2 GPA" 0.85

"6 months internship experience at local company" 0.78

"Proficient in Python programming and SQL databases" 0.65

"Completed 2 personal coding projects" 0.52

"I am interested in this internship opportunity..." 0.23

🔍 AI focuses most on GPA and experience, least on generic cover letter text



FINAL DECISION: REJECTED

Overall Score: 35/100 (Minimum required: 65)



Clear Path to Improvement:

- Raise GPA to 3.5+ (biggest impact)
- Gain more experience through internships/jobs
- Complete more projects to showcase skills
- Write a personalized cover letter



Reveal How AI Decides



Download as Image



Download as PDF