Formal Explainability for Artificial Intelligence in Dynamic Environments

Jaime Cuartas Granada

26-February-2026

Department of Data Science and Artificial Intelligence (DSAI), Monash University, Australia

Introduction & Motivation

The Need for Formal Explainability

- **The Rise of AI:** Increasing deployment of AI in dynamic, high-stakes environments necessitates rigorous eXplainable AI (XAI).
- **The Current Gap:**
  - Most XAI methods rely on *heuristic* approaches.
  - They lack rigorous mathematical guarantees about the explanations generated [**?**].
  - There is a critical need for **formal approaches** that provide exact, provable reasons for AI decisions.
- **The Challenge:** Explaining *sequential decision-making* (where current actions depend on an arbitrary number of previous inputs).

Motivation: Explaining Sequential Decisions

- **Our Approach:** We model sequential processes using **Automata** and **Formal Languages** (FA, CFG).
    - Widely used in software verification and syntax parsing.
    - Provide a symbolic, tractable representation of decision functions.
- **The Problem with Current Parsers:**
    - Modern tools (e.g., Tree-sitter, ANTLR) prioritize *performance* and low-latency error recovery.
- **Our Objective:** Shift the focus from purely performant parsing to **formal interpretability** by extracting the minimal reasons and corrections for sequential failures.

Research Problem Formulation

Research Problem Formulation

Overarching Goal: To provide formal, mathematically guaranteed explanations for sequential decision-making across increasingly expressive computational models.

1. **Explaining Finite Automata (Completed)**
   - How can we provide formal explanations for finite automata decisions?
2. **Explaining Context-Free Grammars (Current Focus)**
   - How can we provide formal explanations for context-free grammar (CFG) decisions?
3. **Explaining Stochastic Models (Future Work)**
   - How can we provide formal explanations for decisions made by stochastic models?
   - *Focus:* Probabilistic paths and "most likely" explanations (e.g., Probabilistic FA or Markov Models).

Completed Work: Explaining Finite Automata

Explaining FA Decisions: The Framework

- **The Setup:** We view a deterministic Finite Automaton (FA) as a classifier mapping an input word $w \in \Sigma^*$ to a class $\mathcal{K} = \{\text{accept}, \text{reject}\}$.
- **Explanation Language:** We use regular expressions formed by substituting selected characters in w with a wildcard token ($\star$).
- **Two Types of Explanations:**
  - **Abductive Explanations (AXps):** *"Why does the FA accept/reject w?"*
    A set of indices that, if **fixed**, guarantees the prediction remains unchanged regardless of the other tokens.
  - **Contrastive Explanations (CXps):** *"How can w be modified to alter the response?"*
    A minimal set of indices that, if **freed** (replaced by $\star$), makes it possible to flip the prediction.

Duality and Milestones

- **The Minimal Hitting Set Duality:**
  - Abductive and contrastive explanations share a profound structural relationship [**?**].
  - Every AXp is a minimal hitting set of the complete set of CXps, and vice versa.
  - *Significance:* This duality forms the basis for the algorithms we developed to efficiently enumerate formal explanations for FAs.

Research Milestone 1

This foundational work on extracting and enumerating formal explanations for regular languages has been completed and submitted to **ICALP 2026**.

Motivation

- ADD CONTENT

Separated content, e.g., variable explanations

Frame without a title

- Frames do not need to have a title...

Math Expressions

Integrals and Other Expressions

$$\iint_{\partial\Omega} f(x)\mathrm{d}x \in \mathbb{C} \tag{1}$$

$$E = mc^2 \tag{2}$$

$$F = ma \tag{3}$$

---

m	Mass
c	Speed of light

Theorem

The following statement is correct

$$\frac{\partial f(\vec{x})}{\partial x_i} = \sum_{l=1}^{L} \cos\left(l\frac{2\pi}{L} + 0\right) \tag{4}$$

Elements

Typography

```
The theme provides sensible defaults to
^^I^^I\emph{emphasize} text, \alert{accent} parts
^^I^^Ior show \textbf{bold} results.
```

becomes

The theme provides sensible defaults to *emphasize* text, accent parts or show **bold** results.

Font feature test

- Regular
- *Italic*
- Small Caps
- **Bold**
- ***Bold Italic***
- **Bold Small Caps**
- `Monospace`
- *`Monospace Italic`*
- **`Monospace Bold`**
- ***`Monospace Bold Italic`***

Lists

Items

- Milk
- Eggs
- Potatoes

Enumerations

1. First,
2. Second and
3. Last.

Descriptions

PowerPoint Meeh.

Beamer Yeeeha.

Table 1: Largest cities in the world (source: Wikipedia)

City	Population
Mexico City	20,116,842
Shanghai	19,210,000
Peking	15,796,450
Istanbul	14,160,467

Blocks

Three different block environments are pre-defined and may be styled with an optional background color.

Default

Block content.

Alert

Block content.

Example

Block content.

Default

Block content.

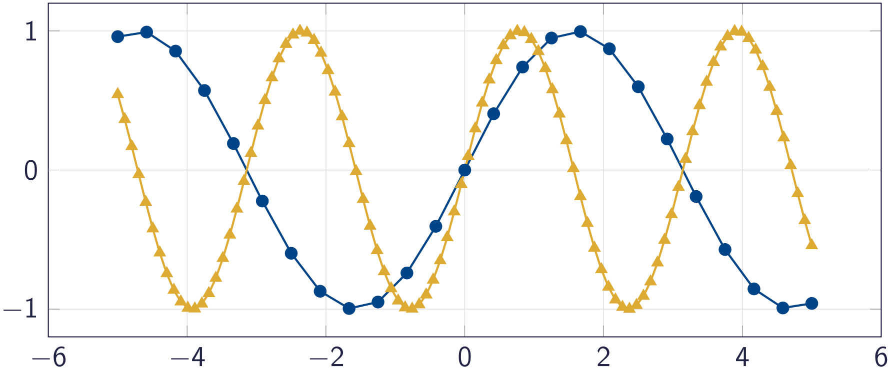
Alert

Block content.

Example

Block content.

Line plots



Standout Frame!

Backup slides

Sometimes, it is useful to add slides at the end of your presentation to refer to during audience questions.

The best way to do this is to include the `appendixnumberbeamer` package in your preamble and call `\appendix` before your backup slides.

The theme will automatically turn off slide numbering and progress bars for slides in the appendix.