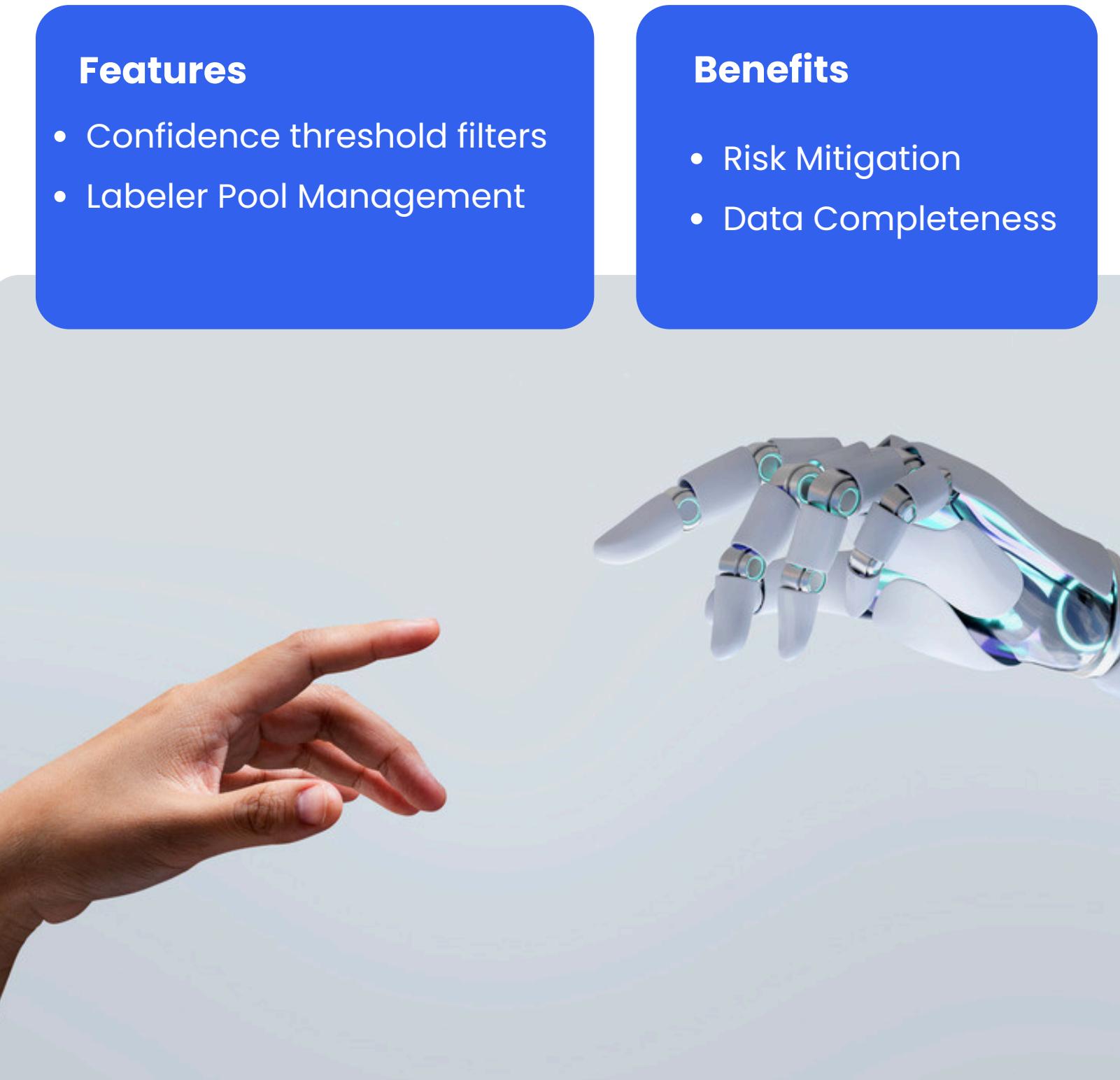


# The Human Touch: Navigating the Challenges of AI Automation and Accuracy in Journalism

A photograph showing a person's hand reaching towards a white robotic arm. The robotic arm has a blue and silver gripper at its end, which is holding a small orange rectangular badge. The background is a plain, light color.

Israel Olatunji Tijani

Data Scientist, ChatVE



### Features

- Confidence threshold filters
- Labeler Pool Management

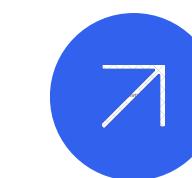
### Benefits

- Risk Mitigation
- Data Completeness

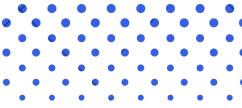
# Why HITL?



Human-in-the-Loop (HITL) enables human verification and corrections to ensure accuracy of data extracted by Document AI processors before it is used in critical business applications



It provides a workflow and a platform for your own or a partner workforce (humans referred to as labelers in HITL) to review, validate, and correct the data extracted from documents by Document AI processors.



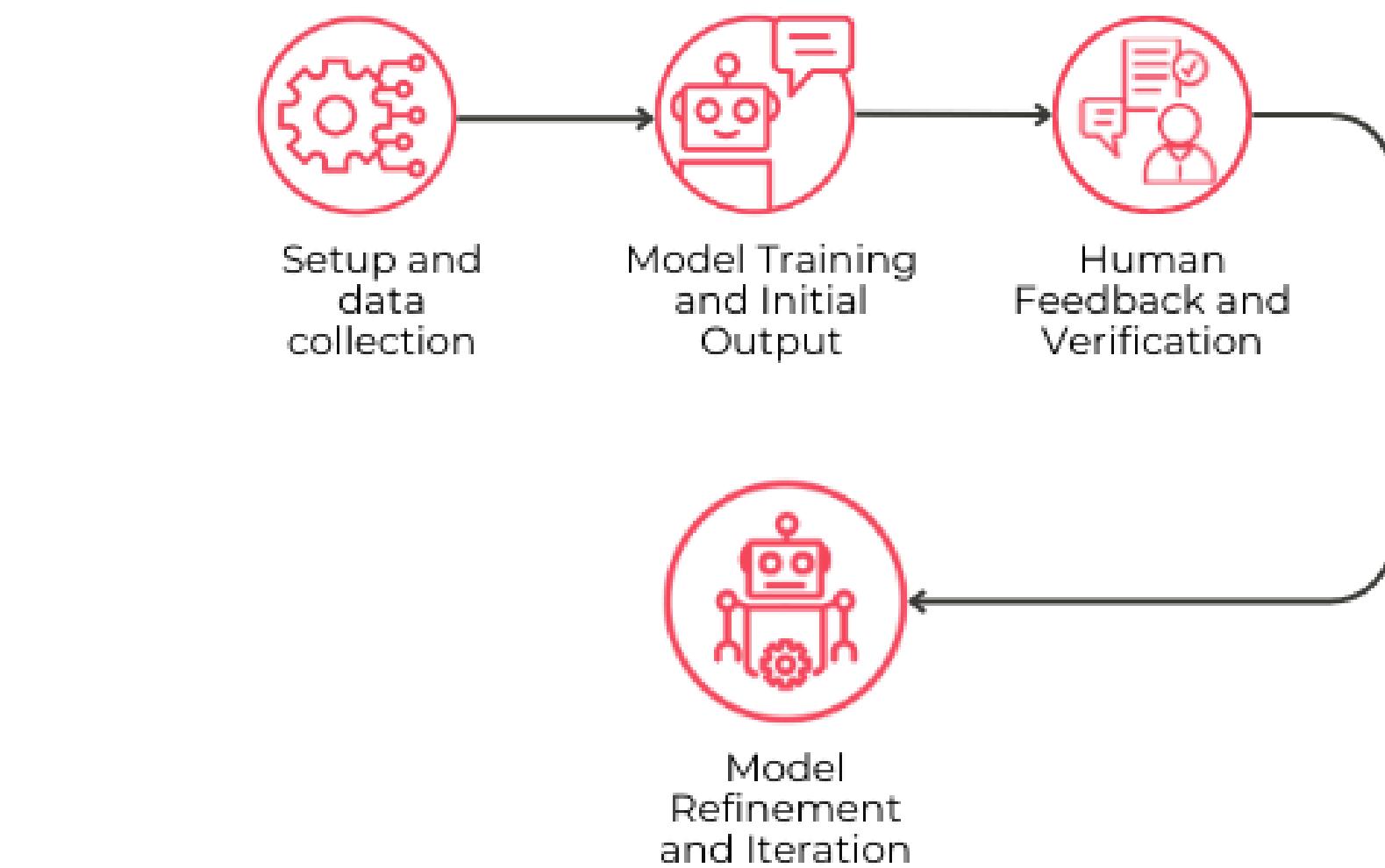
### Features

- Confidence threshold filters
- Labeler Pool Management

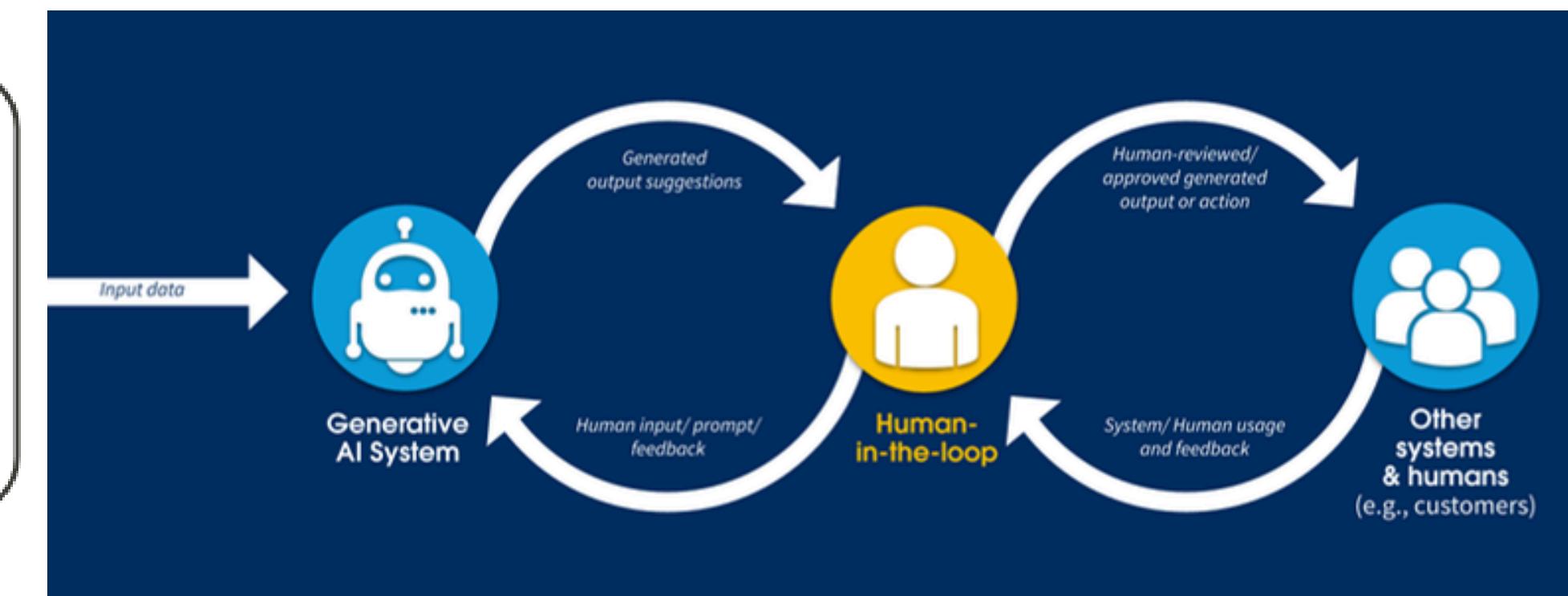
### Benefits

- Risk Mitigation
- Data Completeness

### Human-in-the-loop cycles



# Why HITL?



# Ethics-By-Design

## What is Ethics-By-Design?

- Ethics-By-Design is a proactive and intentional approach. It places ethical considerations at the core of the development process. It does not treat ethics as an afterthought or a checkbox to be ticked off. Instead, this methodology integrates ethical principles from the project's inception.
- It's a commitment to prioritise human rights, fairness, accountability, and transparency throughout the entire lifecycle of tech and AI solutions.



### Oversight

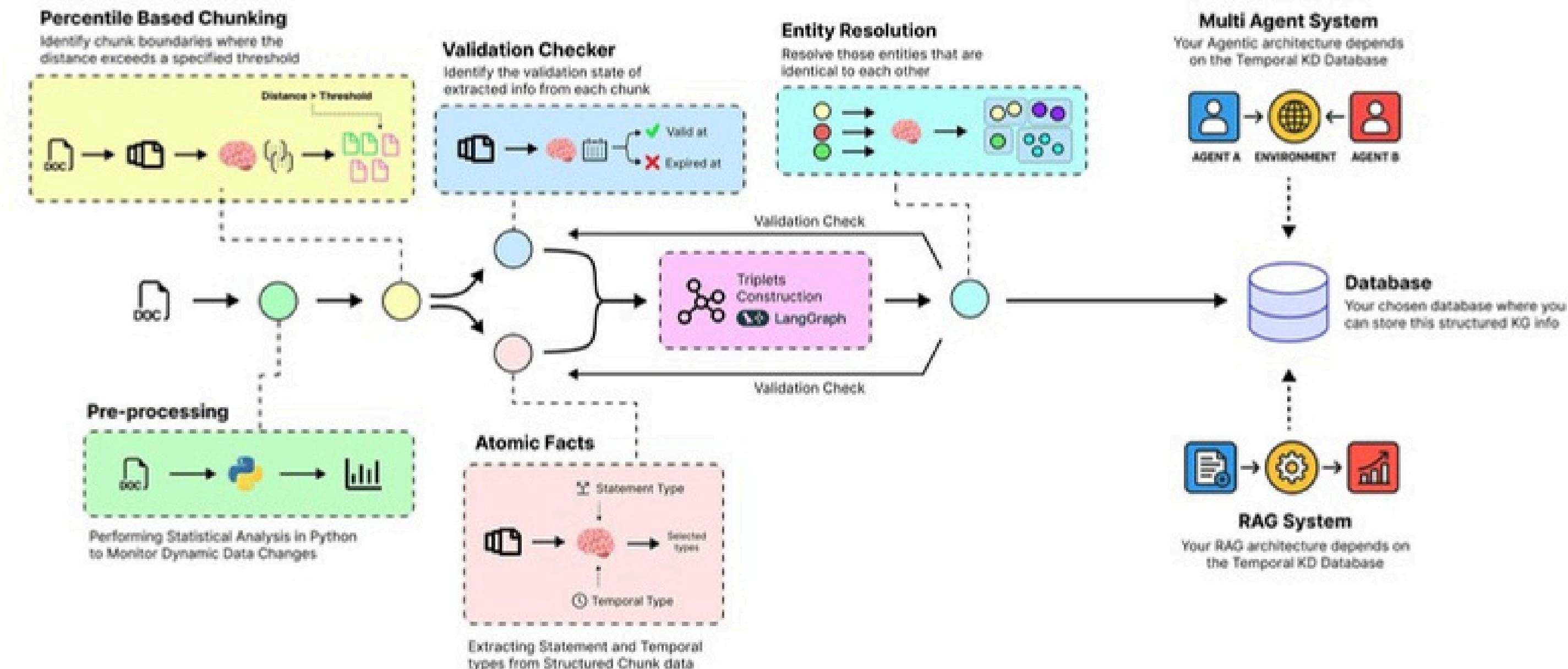
Accountability and Oversight: Human understanding, supervision, and control



### Safety

Individual, Social and Environmental Well-being

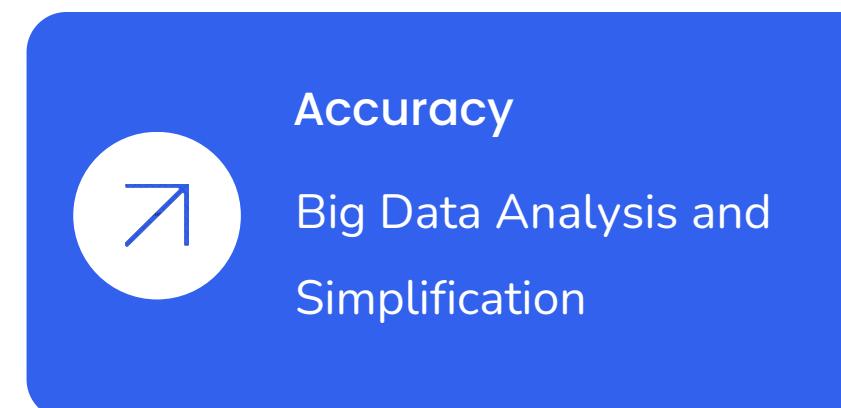
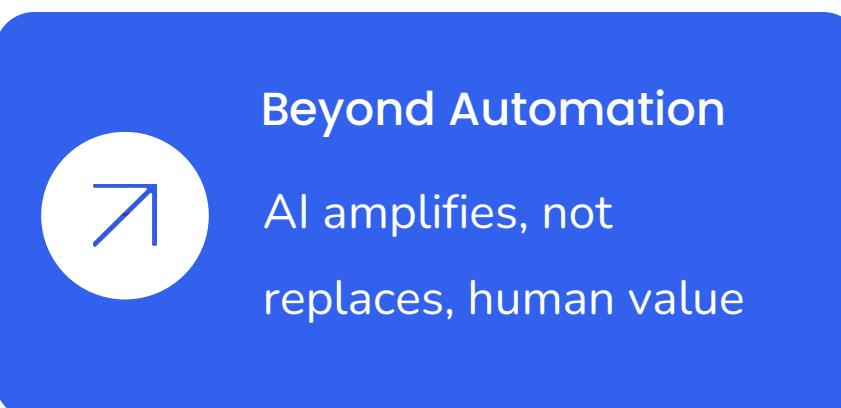
# MEASURING RAG: ACCURACY TO RELEVANCE



# From Tools to Teammates

## The Human-AI Partnership

AI is no longer just a tool—it's becoming a collaborator.  
Together, we unlock new ways to think, create, and solve.



# Quantifying and Embedding RAI Design

- Algorithmic Fairness Index (AFI)

Quantifies bias by measuring disparate impact across demographic or linguistic groups. For an AI system A and a sensitive attribute S (e.g., language group, demographic), AFI could be defined as:

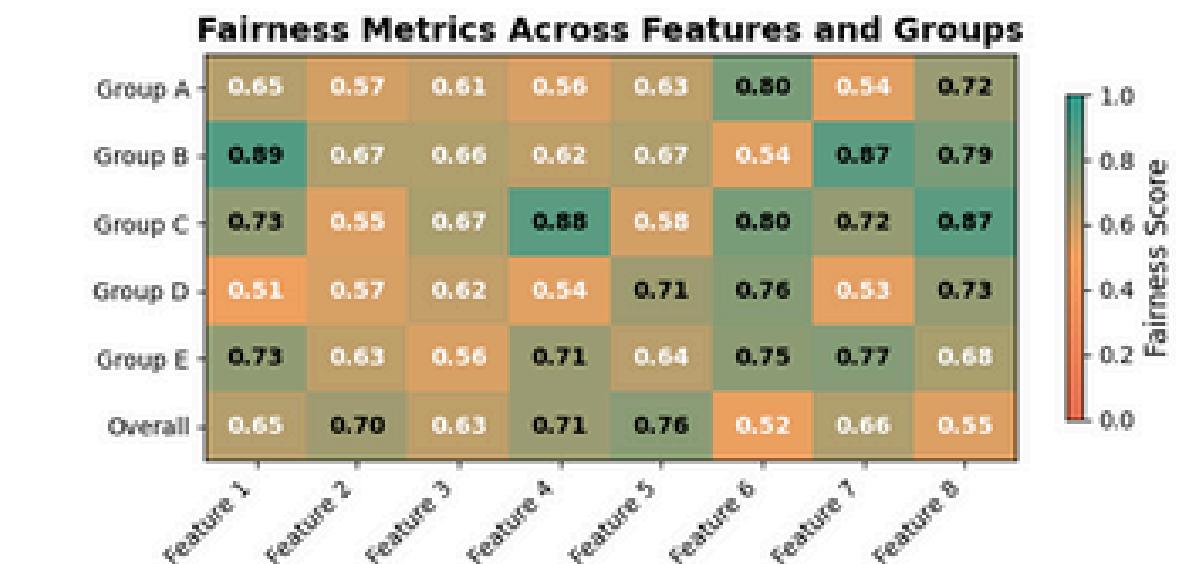
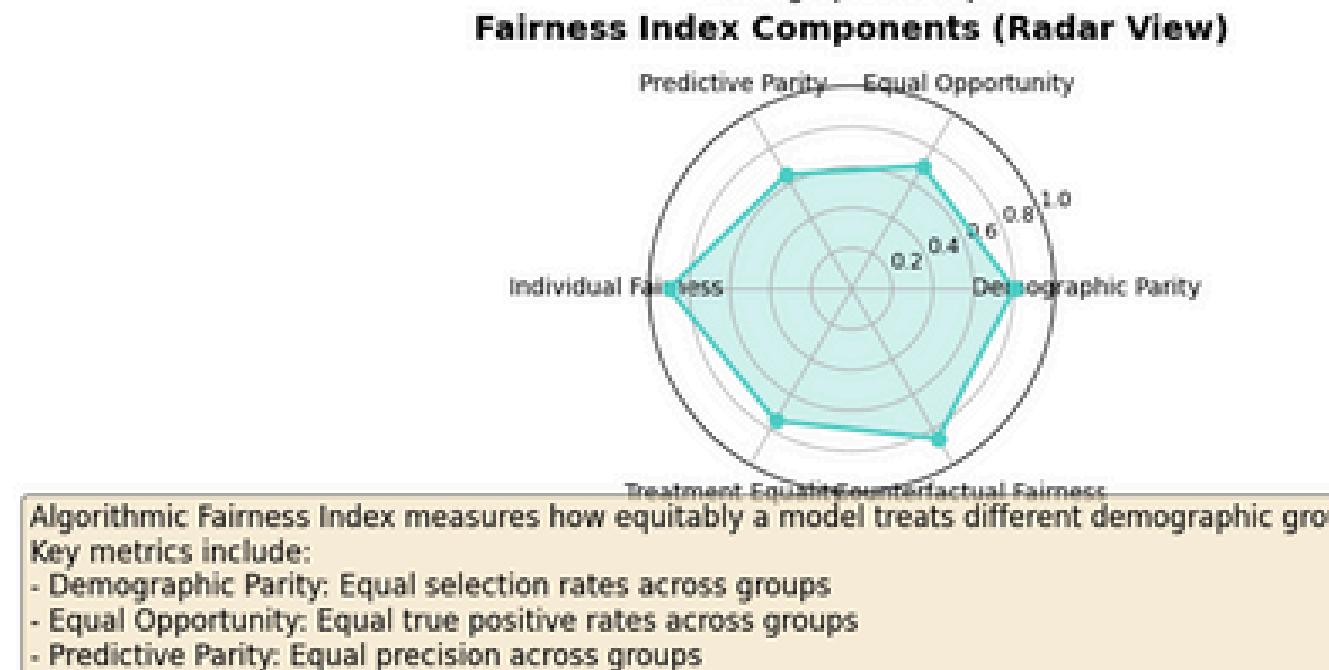
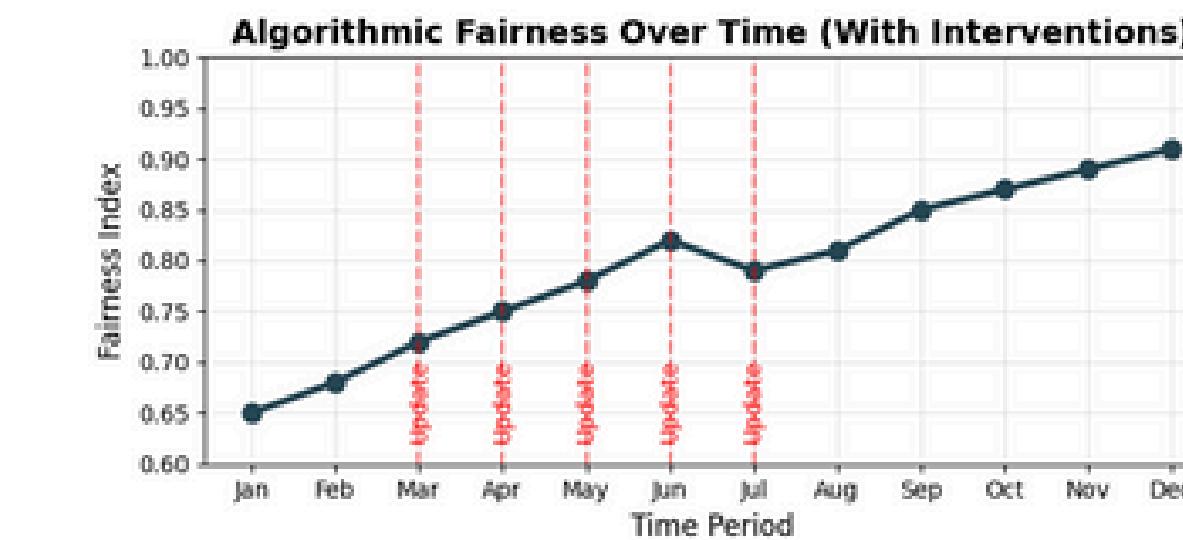
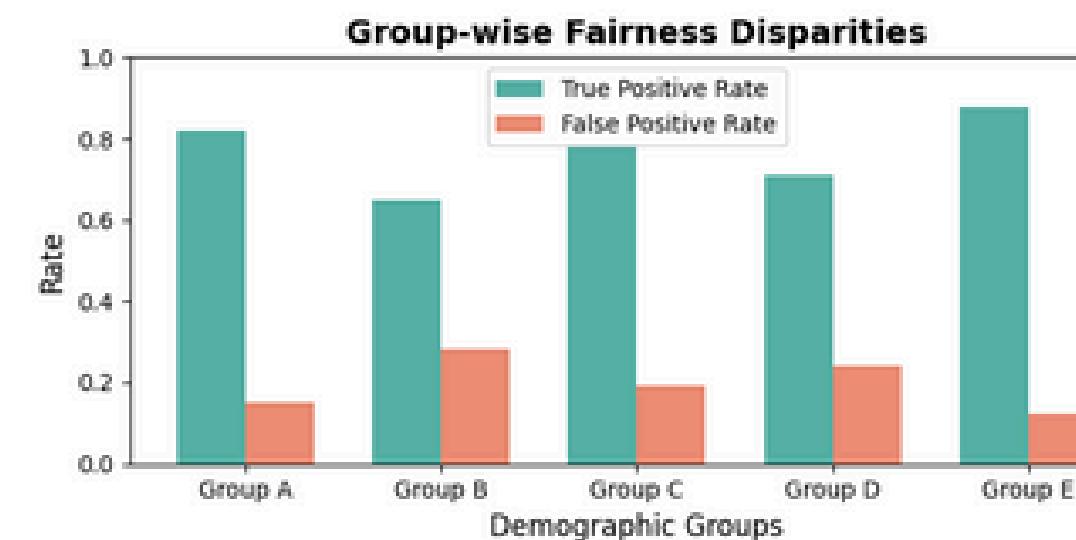
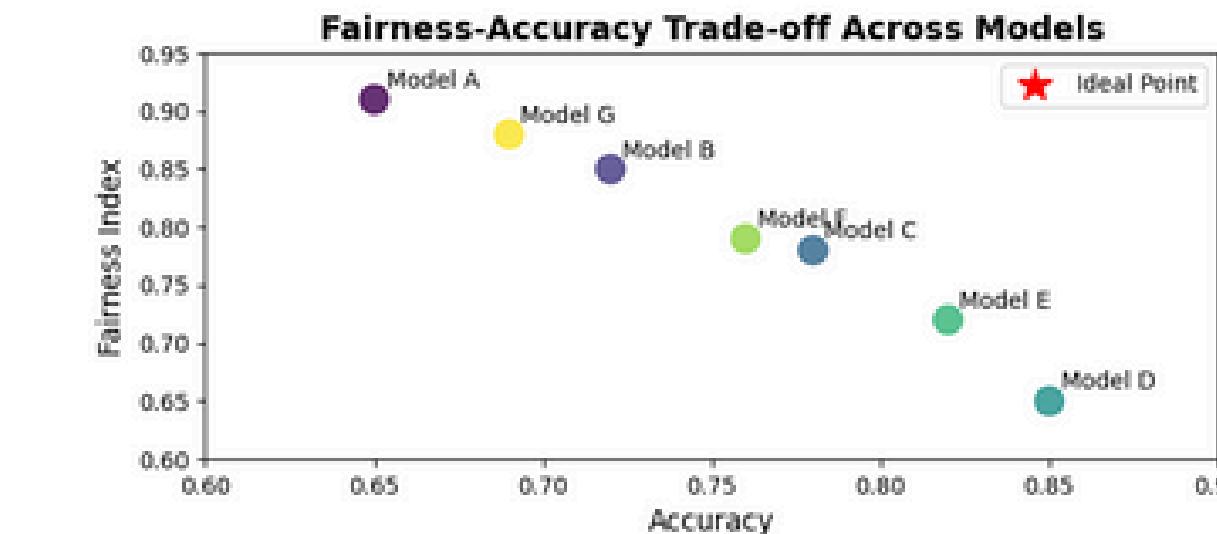
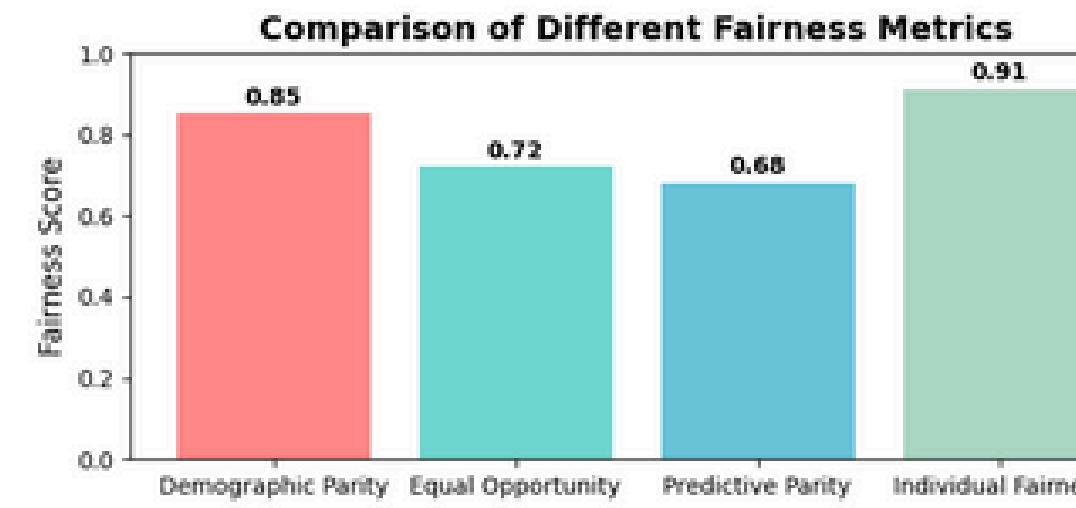
$$AFI = 1 - \frac{1}{N} \sum_{i=1}^N |\text{Outcome}(A, X_i, S_k) - \text{Outcome}(A, X_i, S_j)|$$

where  $N$  is the number of data points,  $X_i$  is an input, and  $S_k, S_j$  are different groups within  $S$ .

**A higher AFI indicates less bias.**

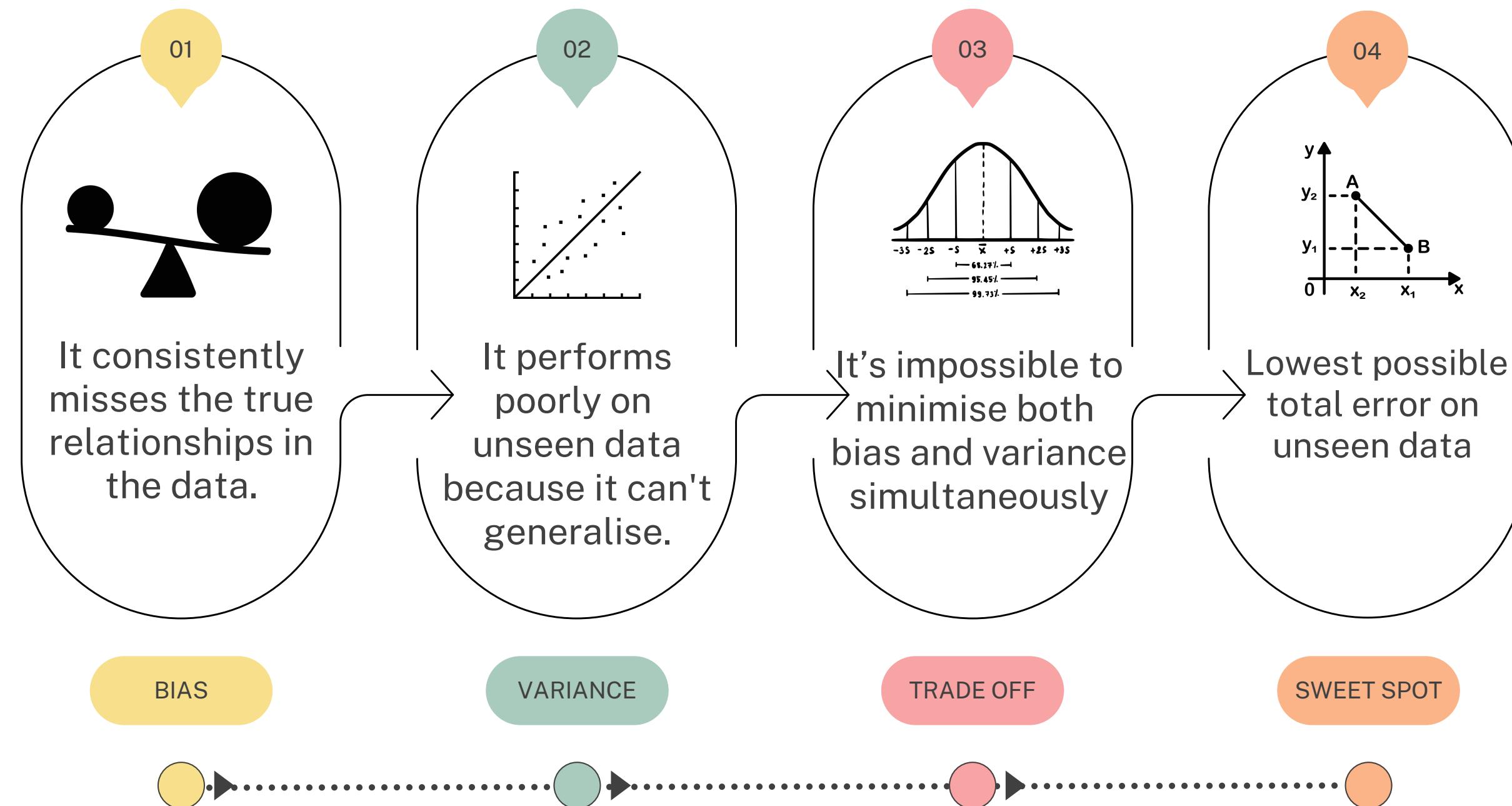


# Algorithmic Fairness Index Visualization

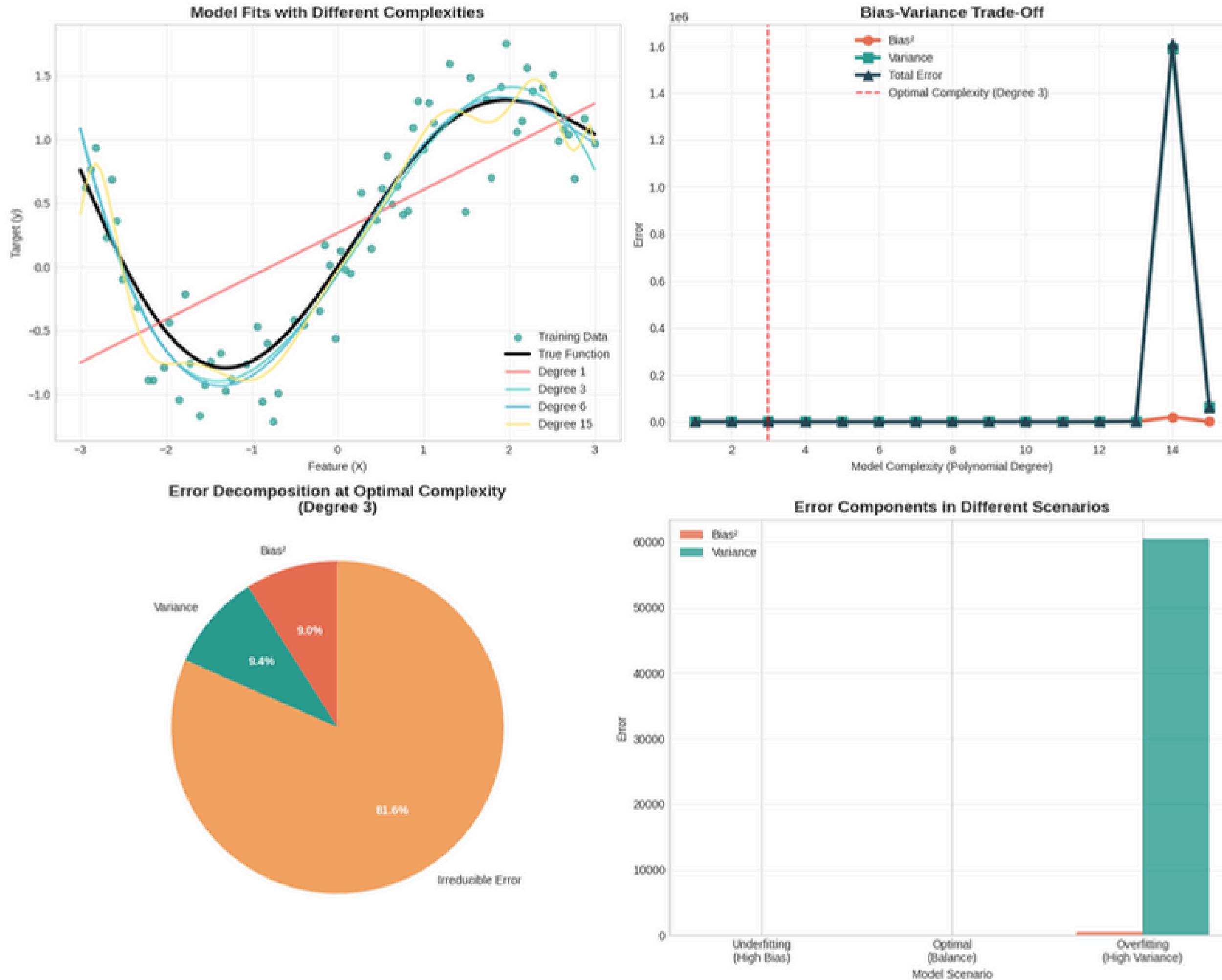


# Quantifying and Embedding RAI Design

- Bias-Accuracy Trade-Off Metric (BATM)



## Bias-Variance Trade-Off in Machine Learning



# Quantifying and Embedding RAI Design

- Ethical Debt Metric (EDM)

Ethical Debt Metric (EDM): Analogous to technical debt, EDM quantifies the cost of neglecting ethical considerations. It accumulates when ethical reviews are skipped, bias is left unaddressed, or privacy-by-design principles are not rigorously applied

$$EDM_t = EDM_{t-1} + \Delta E_t - \text{Resolution}_t$$

where  $\Delta E_t$  newly incurred ethical debt (e.g., from undetected bias, privacy vulnerability) and  $\text{Resolution}_t$  is debt actively mitigated. A rising EDM signals systemic risk.



# Quantifying and Embedding RAI Design

- Community Alignment and Inclusivity Factor (CAIF)

Community Alignment & Inclusivity Factor (CAIF): Measures how well an AI system aligns with the values and linguistic nuances of target communities, particularly relevant for multilingual NLP. This could involve scores from community feedback, expert linguistic review, and performance on diverse, contextually relevant datasets.

$CAIF = \text{Avg}(\text{CommFeedback}) \times \text{LingDiversityPerf} \times (1 - \text{BiasScore})$

where CommFeedback is aggregated sentiment from community consultations, LingDiversityPerf is the system's performance across diverse languages, and BiasScore is a metric of detected bias.

These metrics enable continuous monitoring and iterative refinement, ensuring that responsible AI is not a one-time check, but an ongoing commitment deeply embedded throughout the entire AI development and deployment lifecycle in journalism.

