

Model Transparency and Explainability

- **Why it matters:** Users and developers need to trust and understand model decisions.
- **Challenges:**
 - Transformer models are complex and act like black boxes
 - Hard to trace how visual cues influence outputs
- **Research Directions:**
 - Attention maps, saliency heatmaps, token attribution
 - Explanation-by-example, interactive debugging tools
- **Real-World Need:** Healthcare, law, education—where decisions must be justified.