

Trust Calibration



Appropriate Trust

Balancing between over-reliance and under-utilization



Performance Awareness

Communicating model strengths and limitations clearly



Failure Cases

Highlighting scenarios where model may be unreliable



Confidence Alignment

Ensuring user trust matches actual model performance