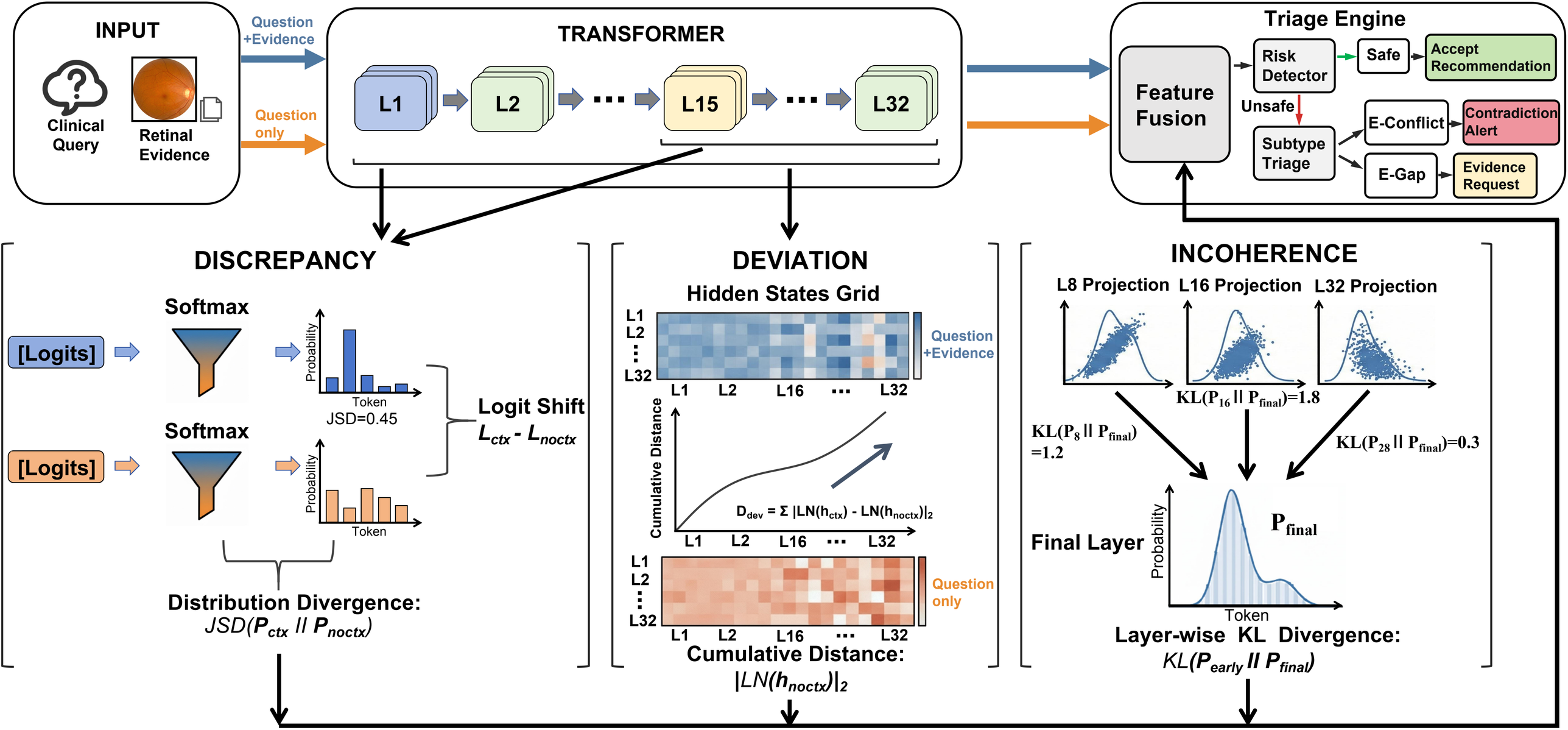
INPUT
Clinical Query
Retinal Evidence
Question +Evidence
Question only
TRANSFORMER
L1
L2
L15
L32
Triage Engine
Feature Fusion
Risk Detector
Safe
Accept Recommendation
Unsafe
Subtype Triage
E-Conflict
Contradiction Alert
E-Gap
Evidence Request
DISCREPANCY
[Logits]
Softmax
Probability
Token
JSD=0.45
Logit Shift
$L_{ctx} - L_{noctx}$
Distribution Divergence:
$JSD(P_{ctx} \,||\, P_{noctx})$
DEVIATION
Hidden States Grid
Question +Evidence
Cumulative Distance
$D_{dev} = \Sigma\, |LN(h_{ctx}) - LN(h_{noctx})|_2$
Question only
Cumulative Distance:
$|LN(h_{noctx})|_2$
INCOHERENCE
L8 Projection
L16 Projection
L32 Projection
$KL(P_{16} \,||\, P_{final})=1.8$
$KL(P_8 \,||\, P_{final})=1.2$
$KL(P_{28} \,||\, P_{final})=0.3$
Final Layer
P_{final}
Layer-wise KL Divergence:
$KL(P_{early} \,||\, P_{final})$