ICFD-31k: A Large-Scale Dataset and Benchmark for Real-Time Conversational Fraud Detection

Rishi Ahuja[1] • Kumar Prateek[1] • Simranjit Singh[1]

[1] Department of IT, Dr. B.R. Ambedkar National Institute of Technology Jalandhar

Why a new benchmark?

Phone scams are temporal: early turns look harmless, evidence arrives gradually, and the useful decision point is **before the call ends.**

Real-data barrier. Victim recordings carry PII, consent risks, and active-scam details; large public corpora are difficult to release safely.

Benchmark gap. Existing datasets are often English-only, post-call, or lack realistic Indian entities and Hinglish dialogue.

ICFD-31k turns each synthetic call into both a full-call example and a sequence of supervised, cumulative decisions.

Scale, class balance, and auditability

31,000 Source conversations

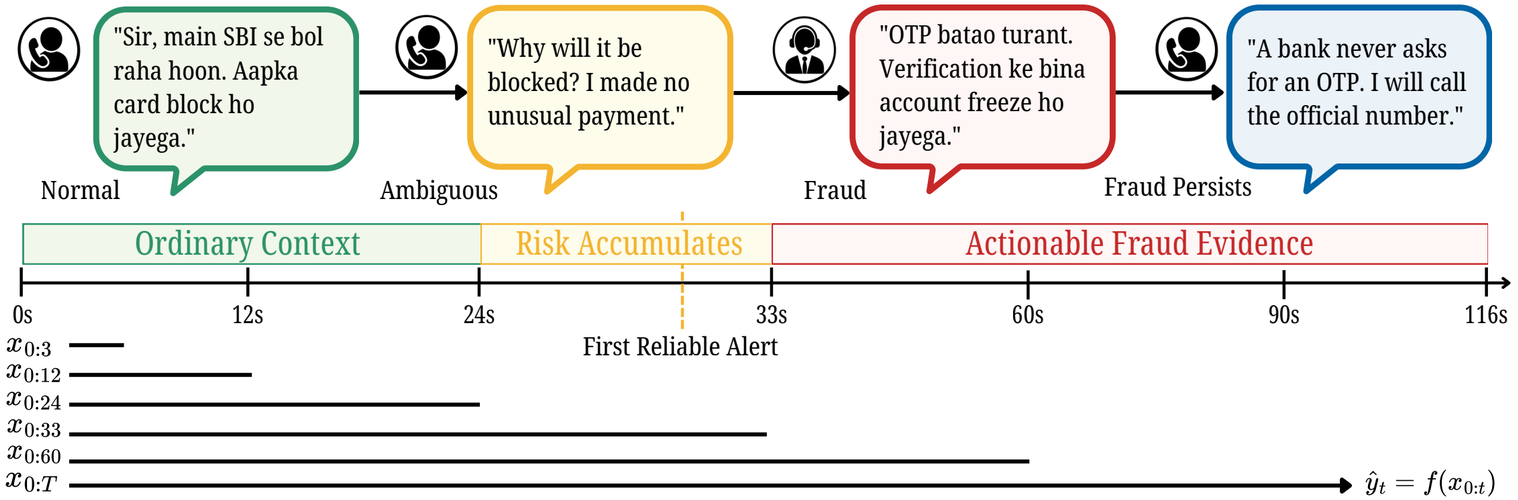
1.11M 3-second chunks

115.8s Mean call length

70:30 Fraud : Normal

Auditable by design: Every chunk retains its source call, timestamp, cumulative transcript, verdict, rationale, and policy evidence.

A live call becomes cumulative evidence every 3 seconds



Streaming task. At time t, predict y_t from the cumulative transcript x0:t. Labels arrive every 3 seconds; evaluation measures both accuracy and when a reliable fraud warning first appears.

99.82% fraudulent calls flagged

32.2 s mean first-detection latency

Is synthetic dialogue credible?

4.19/5 Fluency

3.96/5 Realism

80% Consensus

Five-expert audit of synthetic conversations

GitHub & Code

Dataset

Contact Authors

Beyond detecting fraud: Knowing when to warn. ICFD-31k benchmarks early detection, evidence-grounded explanations, and generalization to unseen fraud types.