

Indic Eval

Evaluation Framework for Indian Languages.
Rigorous benchmarks for all 22 scheduled languages and code-mixed communication.

22

LANGUAGES COVERED

6

EVALUATION DIMENSIONS

10K+

NATIVE TEST CASES

rotavision.com
contact@rotavision.com

You Can't Improve What You Can't Measure

Standard LLM benchmarks were built for English. They don't capture what matters for Indian languages—native fluency, code-mixing, cultural context, and regional accuracy. **Without proper evaluation, you're flying blind.**

Translation is Not Evaluation

Translating English benchmarks to Hindi doesn't test Hindi capability. It tests translation quality. Native speakers don't think in translated sentences.

Code-Mixing Ignored

Real Indians speak Hinglish, Tanglish, Benglish. Standard benchmarks pretend this doesn't exist. Your model fails in the real world.

Hallucination Detection Fails

Factual accuracy checks need Indian knowledge bases. Western benchmarks don't know that Diwali dates change yearly or that Indian states have different official languages.

Cultural Context Missing

Sentiment analysis trained on Western data misreads Indian communication patterns. Respectful language varies by region, generation, and context.

You can't improve what you can't measure. [Indic Eval measures what matters.](#)

The Indic Eval Solution

A comprehensive evaluation framework built from the ground up for Indian languages—with native test cases, code-mixed scenarios, and cultural context awareness.

Native-first test cases

Test cases created by native speakers in native thought patterns. Not translations.

Code-mixing as standard

Hinglish, Tanglish, Benglish, and other code-mixed variants tested explicitly.

Indian knowledge validation

Factual accuracy against Indian context—geography, history, culture, current events.

Cultural sensitivity scoring

Evaluate appropriateness across regions, communities, and communication contexts.

Six Evaluation Dimensions

01

Native Fluency

Grammar, idiom usage, natural phrasing as judged by native speakers. Not just grammatical correctness—actual fluency.

02

Code-Mixed Evaluation

Explicit testing of Hinglish, Tanglish, Benglish, and other code-mixed communication patterns common in real usage.

03

Factual Accuracy

Validation against Indian knowledge bases. Geography, history, civics, current events, regional facts.

04

Cultural Context

Appropriateness for regional norms, respect patterns, festival awareness, community sensitivities.

05

Domain Expertise

Specialized evaluation for legal, medical, financial, government domains in Indian context.

06

Safety & Sensitivity

Detection of harmful content, misinformation, communal sensitivity, and bias in Indian context.

Code-Mixed Examples

HINGLISH (HINDI + ENGLISH)

"Yaar, meeting postpone ho gayi kya? Main already office pahunch gaya."

TANGLISH (TAMIL + ENGLISH)

"Project deadline extend aayirukku, but client approval still pending irukku."

BENGLISH (BENGALI + ENGLISH)

"Weekend e plan ki? Weather ta bhalo thakle picnic e jabo maybe."

EVALUATION CHALLENGE

Standard benchmarks can't score these. Indic Eval evaluates script-mixing, grammar blending, and contextual appropriateness.

Architecture

YOUR MODELS / APIS

OpenAI Anthropic Google Sarvam Custom Models

INDIC EVAL FRAMEWORK

Test Runner Native Judge Models Code-Mix Analyzer Cultural Validator Factual Checker Report Generator

TEST CORPUS

22 Language Datasets Code-Mixed Corpora Domain-Specific Sets Cultural Sensitivity Cases

All 22 Scheduled Languages

Indo-Aryan Languages

- Hindi / Hinglish
- Bengali / Benglish
- Marathi
- Gujarati
- Punjabi
- Odia
- Assamese
- Maithili
- Sanskrit
- Sindhi
- Konkani
- Dogri
- Kashmiri
- Nepali

Dravidian Languages

- Tamil / Tanglish
- Telugu
- Kannada
- Malayalam

Other Families

- Manipuri (Meitei)
- Bodo
- Santali
- Urdu

Code-Mixed Variants

- Hinglish (Hindi-English)
- Tanglish (Tamil-English)
- Benglish (Bengali-English)
- Tenglish (Telugu-English)
- Kanglish (Kannada-English)
- Manglish (Malayalam-English)

Script Coverage

- Devanagari
- Tamil
- Telugu
- Kannada
- Malayalam
- Bengali
- Gujarati
- Gurmukhi
- Odia
- Roman transliteration

Domain Coverage

- Legal & Government
- Healthcare & Medical
- Banking & Finance
- E-commerce & Retail
- Education
- Agriculture

Comparison

Capability	Standard Benchmarks	Indic Eval
Indian language coverage	2-3 languages (translated)	22 native languages
Code-mixing evaluation	Not supported	6+ code-mixed variants
Native speaker validation	Automated only	Human-in-the-loop
Cultural context	Western-centric	India-specific
Factual accuracy (India)	Generic knowledge	Indian knowledge base
Safety evaluation	Global standards	Indian sensitivity norms

Deployment & Integration

API Access

Submit model outputs. Get scores. Simple REST API.

SDK Integration

Python, Node.js SDKs. Integrate into CI/CD pipelines.

Dashboard

Visual reports. Trend analysis. Model comparison.

Custom Benchmarks

Add your own test cases. Domain-specific evaluation.

Use Cases

Evaluate LLM quality throughout the AI lifecycle.

- Model selection and comparison
- Production monitoring
- Vendor evaluation
- Fine-tuning validation
- Regression testing
- Compliance documentation

Enterprise Features

- Private deployment option
- Dedicated judge model training
- Integration with Sankalp Gateway
- Custom test corpus creation
- SLA-backed evaluation service
- Automated evaluation pipelines

Ready to Measure What Matters?

Stop guessing. Start measuring Indian language quality.

Contact Us

Rotavision Consulting Private Limited

HD37, Block D3, Manyata Tech Park,
Outer Ring Road, Venkateshapura,
Bangalore North, Bangalore - 560045,
Karnataka, India

CIN: U70200KA2025PTC196547

Web: rotavision.com

Email: contact@rotavision.com