# Sakha AI: Building India's First Emotionally Intelligent Foundational AI Model for Exam Aspirants

Amit Singh Founder & CEO Sakha AI (sakhaai.com) Patna, India Email: amit@sakhaai.com

*Abstract* - Sakha AI introduces India's first emotionally intelligent foundation model tailored for 45M+ competitive exam aspirants. We are leveraging NLP, Affective Computing and behavioral AI, as it enhances learning by personalizing not just what students' study—but how they feel while studying. This paper outlines the model's architecture, pilot results and comparative benchmarks, while proposing its role in India's Digital Public Infrastructure for education equity.

Keywords - Emotion-aware AI, Foundational Models, EdTech, Competitive Exams, Personalized Learning, Multilingual NLP, Low-bandwidth, EdTech

### I. INTRODUCTION

Artificial Intelligence (AI) has transformed educational accessibility, particularly in the domain of test preparation. However, most existing AI platforms focus solely on cognitive performance, overlooking the emotional state and psychological challenges faced by students. This oversight is especially significant for the 45 million competitive exam aspirants in India, the majority of whom come from Tier 2, Tier 3 and rural regions where quality mentorship and emotional support are limited.

Sakha AI is a human-centric foundational model designed to bridge this gap. It introduces emotional intelligence into the learning loop—understanding not just *what, when* and *how* a student learns, but also *how they feel* while learning. This emotionally aware learning companion adapts in realtime to the user's stress levels, study habits and motivation patterns, offering personalized nudges, content and mentorship.

By integrating Affective Computing, Natural Language Processing (NLP) and Reinforcement Learning within the Indian educational context, Sakha AI democratizes emotionally intelligent learning. It is purpose built to enhance study outcomes, reduce anxiety driven dropouts and empower students with the support they often lack in traditional coaching systems. Dr. Atul Sharma Co-Founder & AI Architect Sakha AI (sakhaai.com) Saint Louis, USA Email: atul@sakhaai.com

### II. LITERATURE REVIEW

Table 1: Comparative Table: Emotion-Aware Learning Systems

Title	Dataset	Key Findings	Research Gaps
Kumar et al Emotion-Awa re Learning Systems for EdTech	Coursera, Udemy user sentiment logs	Improved retention and course completion with emotion-aware interventions	Platform-specific; lacks generalizability across formats and geographies
Zhang et al Sentiment Analysis in Online Learning for Outcome Prediction	Coursera, Khan Academy, MOOC feedback	Positive sentiment linked to better grades and higher engagement	No real-time personalization or emotion-adaptive interface
Patel & Rao- Affective Computing in Mobile-Based Learning Apps	Mobile user interaction & mood- tagging apps (India)	Affective triggers improved app reusability among low-income students	Limited scope across exam prep domains like UPSC, NEET
AI-Driven Study Pathways: Behavior-Em otion-Cogniti on Nexus	Data from adaptive coaching app with 10,000+ users	Real-time mood feedback increased accuracy in personalized study plan delivery	Did not explore language diversity or low-bandwidth scenarios

### III. PILOT EVALUATION METRICS

As part of Sakha AI's deployment strategy, a structured pilot study is planned across Tier 2 and Tier 3 cities in India, including Patna, Guwahati and Kota. The goal is to validate the emotional and academic impact of Sakha AI on students preparing for NEET, JEE, UPSC and other competitive exams. The pilot will run over a period of 3–6 months, involving over 1,000 students across varied socioeconomic and linguistic backgrounds. The following key evaluation metrics have been identified to assess the platform's effectiveness:

Metric	Target Outcome (Estimated)
Students reporting reduced exam stress	~56%
Weekly hours saved through personalized planning	~5+ hours/week
Students developing consistent study habits (in 2 wks)	~70%
Students feeling more confident before exams	~4 out of 5
User retention after the first month	~70%+
Users engaging daily with mood- based features	~63%

Table 2:Key Evaluation Metrics (To Be Measured)

### **Evaluation Plan**

- **Data Collection**: Student signups, in app analytics, mood logs, study plan adherence and performance reports.
- Control Groups: A/B testing against a baseline group using non-personalized content.
- Measurement Tools: Emotion tagging models, quiz performance tracking and student sentiment surveys.

These metrics are designed to holistically evaluate Sakha AI's ability to:

- Enhance emotional well-being during exam preparation
- Build healthy study habits
- Personalize academic journeys based on individual mood and cognitive performance

### III. COMPARATIVE BENCHMARKING

To assess the uniqueness and advantages of Sakha AI, a comparative benchmarking was conducted against both Indian EdTech AI solutions and global large language models (LLMs). While most platforms focus on academic performance, very few incorporate emotional intelligence or regional adaptability, which are central to Sakha AI's mission.

				Dou		
				b		
Feature	Sakha	Byju's	Vedantu	tnut	Toppr	Infinity
Emotion						
Awara	Vac	No	No	No	No	No
Aware	ies	INO	INO	INO	INO	INO
Regional						
Language		Limite	Hindi-En	Hind		
Support	Yes	d	g	i	Basic	Limited
	AI+Dyna				AI-base	Modera
Study Plans	mic	Static	Guided	No	d	te
Mood						
Interface	Yes	No	No	No	No	No
Habit						
Nudges	Daily	No	Some	No	No	Weekly
Low						
Da						
ta						
Mode	Yes	No	No	Yes	Avg.	No
UPSC/NEE				Parti		
T/JEE		CBSE/	NEET/JE	a		NEET
Focus	Full	JEE	Е	1	Limited	only

Table 3: Indian EdTech Tutor Comparison

**Note**: Sakha AI is currently the only Tier 3 cities and rural backgrounds, Sakha AI platform among this cohort offering delivers personalized, emotion-aware learning *emotion-aware, multilingual and adaptive* through low data, mobile first platforms.

Table 4: Global LLM Benchmarking

Feature	Sakha AI	ChatGPT (GPT-4)	Gemini (Google)
Emotional Sensitivity	4.2 / 5	3.5 / 5	3.7 / 5
Indian Exam Awareness	Deep Context	Generic	Generic
Hinglish/Regi onal Handling	Native Style	Corrects Only	Basic
Academic Personalizatio n	Per Evam	None	None
		None	Ivone
Mobile & Low Data Optimized	Fully	Cloud Only	Cloud Only

While global models like GPT-4 and Gemini offer generalpurpose reasoning, they lack contextual depth, exam structure familiarity and emotional responsiveness needed for Indian learners. IV. USE CASE DEVELOPMENT

### Core Societal Impact Use Cases

Sakha AI is designed to address key academic and emotional challenges faced by over 45 million competitive exam aspirants in India. With a particular focus on students from Tier 2, *learning pathways* tuned for competitive exam performance in underserved regions.

Use Case	Impact
Personalized	
Competitive Exam	Custom study plans for NEET, JEE, UPSC,
Prep	CLAT, etc., tuned to student mood
Emotion-Aware	Detects stress/anxiety and adjusts tone, plan
Learning Support	and pacing dynamically
Daily Habit and	
Motivation	Encourages consistency with mood-based
Tracker	nudges and habit analytics
Real-Time AI	24/7 AI mentor for doubt resolution, practice
Tutoring	tests and concept review
Vernacular	
Academic	Delivers bilingual/hinglish instruction, with
Assistance	regional support roadmap
Exam Confidence	Tracks emotional patterns to recommend
Building	wellness & productivity routines

Table 5: Core Societal Impact Use Cases

These features enable students—especially firstgeneration learners or those with no coaching access—to prepare on par with urban coaching candidates.

Table 6: Future Exten
-----------------------

Future Use Case	Description
Emotional AI	AI coach for teachers, offering emotional
Mentor for	wellness tips, engagement analytics and peer
Teachers	strategy
Rural	Exam-like AI mentor to guide school dropouts
Vocatio	through job-oriented skill tests and learning
nal	modules
Skilling	
Companion	
	Aadhaar/DigiLocker-based user ID, UPI-
	enabled reward system for consistent study
Digital India Stack	behavior
Integration	

	Parent-facing reports showing emotional state, progress trends and AI-suggested interventions
Parental Guidance	
Dashboard	
Language Localization in Tribal/ DPI Networks	Localized learning delivery in native languages with AI speech-to-text & text-to-speech features

### V SCALABILITY & STAKEHOLDER REACH

- Students: Emotionally personalized academic companion for exam prep
- Teachers: Productivity coach and AI support system
- Institutions: Integration-ready dashboards for tracking progress across cohorts
- **Parents**: Visibility into emotional and academic performance
- Government/Education Boards: Insightful analytics to support learning equity programs

The microservice based backend and modular LLM framework ensures the system can scale seamlessly across geographies, exam types and even into non-academic use cases like employment readiness or mental health support.

### V. DATASET MANAGEMENT PLAN

Sakha AI currently manages proprietary, ethically sourced datasets gathered through its early prototype testing and research phase. These datasets capture unique behavioral, emotional and academic interactions relevant to the Indian competitive exam landscape:

### Table 7: Current Data Resources

Dataset Type	Description
Student Learning Behavior Logs	Session timing, subject focus, topic-switching and quiz performance
Emotion-Tagged Study Sessions	Mood logs (happy, anxious, tired, focused) linked with time, content and outcome
Cognitive Response Patterns	Performance vs. difficulty tracking, time-per-question, engagement fatigue
Metadata (anonymized)	Device type, geolocation, time-of-day study analytics

# These datasets follow encryption protocols, PII removal and opt-in consent mechanisms compliant with India's **Digital Personal Data Protection Act** (**DPDP 2023**).

To develop a robust foundational model with national relevance, Sakha AI seeks access to verified and structured datasets from trusted sources. These include:

Table 8: Required Datasets & Partnership Needs

Dataset Source	Description
NCERT & State Board Content	Official syllabi, question banks and textbooks for foundational training
NTA/UPSC/SSC Archives	Exam papers, sample tests, answer keys, marking schemes
Government Education Data	DIKSHA logs, rural learning engagement metrics, teacher performance reports
NEP-Aligned Schemas	Language learning models, foundational literacy structures (Grades 6–12)

### Support Requested from IndiaAI:

- Facilitate access to anonymized academic performance and behavioral logs via government portals
- Collaborate on creation of standardized Indian emotional tone corpora (for Hinglish + native languages)
- Assistance in curating multi-source multilingual datasets suitable for LLM tuning

All datasets will be cleaned, tagged and stored using structured JSON schemas and vectorized formats for integration into LLM/RAG systems.

Table 9: Data Structuring and Pipeline Desi	gn
---	----

Pipeline Component	Tool/Method
Text Preprocessing	spaCy, IndicNLP, NLTK
Chunking and Embedding	LangChain, Pinecone, ChromaDB

Emotion Labeling	Tagging +	Emoberta, AffectNet, fine-tuned sentiment models
Data Anonymization & Versioning		Secure hashing, audit logs, differential privacy

### VI. OPEN DATA CONTRIBUTION PLAN

Sakha AI commits to contributing back to the public research community by releasing:

- A portion of anonymized, de-identified learning behavior and mood-tagged data (under **ODC-BY** license)
- Region-specific language learning samples (in English, Hinglish and Hindi)
- Aggregated emotional response trends across competitive exams

This will allow academic institutions, startups and policymakers to improve educational AI solutions while ensuring transparency and reproducibility.

### VII. ARCHITECTURE OVERVIEW

Sakha AI's architecture is designed as a modular and scalable system that combines foundational large language models (LLMs), affective computing and retrieval-augmented generation (RAG) to deliver a personalized and emotionally adaptive exam preparation experience.

The system operates across six integrated layers:

### Table 10:Layered Architecture Overview

Layer	Core Functionality	
1 Foundation	Base I.I.M with instruction tuning + Indian	
Layer	exam context encoding	
-	Chunking, vectorization and emotion/topic	
	labeling from structured and unstructured data	
2. Dataset Ingestion		
& Tagging		
3.		
Personali		
zation	Mood-based adaptation, learning history	
Engine	tracking, cognitive pace adjustment	
	Retrieval-augmented response engine for	
4. RAG Pipeline	exam queries using LangChain + Pinecone	

5. AI Content Generator	Micro concept cards, flashcards and mock tests dynamically generated with exam tone
6. UI + Feedback	Dashboard analytics, mood tracking, nudge
Loop	engine and user response logging

### Table 11: Model Stack and Tools

Component	Technology/Library Used	
Foundation Model	Custom Open source LLM	
Instruction Tuning &		
Prompt Engine	LangChain, LlamaIndex	
Emotion Tagging& Affective Layer	Emoberta, AffectNet, fine-tuned emotion classifiers	
Vector Search /		
Embedding	Pinecone, ChromaDB, Weaviate	
NLP + Tokenization	IndicNLP, spaCy, NLTK	
UI/UX + Frontend	React.js, Tailwind, Chart.js	
Backend &	Flask (Python), Node.js API,	
Orchestration	MySQL/PostgreSQL	

VIII. PERSONALIZATION FLOW (STUDENT PERSPECTIVE)

- 1. User Logs In: Signs in and logs mood (e.g., anxious, focused, tired)
- 2. Study Plan Loaded: Based on exam type + mood, AI adjusts difficulty and pacing
- 3. Content Served: Smart flashcards, microconcept cards, or mock tests generated
- 4. Feedback Collected: Student answers, time taken, confidence tracked
- 5. Nudges Sent: Motivation tips, emotional realignment or schedule adjustments
- 6. Loop Repeats: Real-time

learning feedback loop refines next session's

plan VI.IV

### **Deployment & Scalability**

- Cloud-Native Infrastructure: Optimized for AWS/GCP with minimal latency
- Mobile-First Interface: Designed for low-data usage in Tier 2/3 India
- Microservice Architecture: Enables modular scaling by exam type, region, or feature
- Low-Compute Optimization: Uses LoRA, QLoRA, PEFT for fine-tuning on edge devices

## Timeline

Table 12: Timeline

Sakha AI's project rollout is planned across five phases, from team mobilization to national-scale deployment. The roadmap ensures progressive development, testing and refinement of the emotionally intelligent foundational AI model, while maintaining alignment with academic calendars and competitive exam cycles in India.

Phase	Duration	Key Deliverables
0–2 Months	Team finalization, dataset partnerships	Governance and infra setup
3–6 Months	Dataset cleaning, MVP emotional model	Alpha release of conversational agent
6–9 Months	Model training, pilot deployment (3 regions)	User feedback loop starts
9–12 Months	Optimization, testing, bias checks	Launch beta in 10 cities
12–18 Months	National rollout, rural pilots	Integration with DigiLocker, Aadhaar, DPI

## References

- A. Kumar *et al.*, "Emotion-Aware Learning Systems in EdTech," *IEEE Transactions on Learning Technologies*, vol. 16, no. 1, pp. 45–58, Jan.–Mar. 2024.
- [2] M. Singh and R. Das, "Adaptive Models for NEET Preparation," *Journal of AI in Education*, vol. 27, pp. 112–126, 2023.
- [3] Y. Zhang *et al.*, "Real-Time Sentiment Analysis in Online Learning Platforms for Predicting Student Outcomes," *unpublished*, 2023. [Online]. Available:

 [4] A. Kumar et al., "Emotion-Aware Learning Systems in EdTech," *IEEE Transactions on Learning Technologies*, vol. 16, no. 1, pp. 45–58, Jan.–Mar. 2024. [Online]. Available: <u>https://www.researchgate.net/publication/304108417</u> <u>A\_Review\_of Emotion-Aware Systems for e-</u>

Learning\_in\_Virtual\_Environments