Title: Kruti

URL: https://en.wikipedia.org/wiki/Kruti

PageID: 80222264

Categories: Category:2025 in artificial intelligence, Category:2025 software, Category:AI software, Category:Chatbots, Category:Generative pre-trained transformers, Category:Indian inventions, Category:Large language models, Category:Products introduced in 2025, Category:Proprietary software, Category:Virtual assistants

---

Kruti is a multilingual AI agent and chatbot developed by the Indian company Ola Krutrim. It is designed to perform real-world tasks for users, such as booking taxis and ordering food, by integrating directly with various online services. It is notable for its ability to understand and respond in multiple Indian languages . [1] [2]

Developed by a team founded by Bhavish Aggarwal , Kruti functions as an "agentic" AI, meaning it can reason, plan, and execute multi-step tasks to fulfill a user's request. The backend technology combines several open-source large language models with Ola's proprietary Krutrim V2 model. The system was developed to work primarily on smartphones, addressing the Indian market's specific needs, including language diversity and potential bandwidth constraints. [3] [4]

Kruti was officially released in June 2025, replacing an earlier chatbot from the company that was also named Krutrim. Initially supporting 13 languages, the company plans to expand its capabilities to 22 Indian languages. [5] [6]

Background

Kruti is an improved version of Ola's Krutrim chatbot, which was first launched in 2023 and was intended to be replaced by Kruti. [7] [8] It was officially released on 12 June 2025 as an upgrade to passive chatbots, with support for text and voice in 13 Indian languages . As an agentic AI, it can execute tasks with customization and reasoning , providing adaptive answers based on user preferences and past interactions. [2] [6] [4]

Kruti is optimized for smartphone usage and designed to accommodate bandwidth constraints and usage patterns in India. To ensure scalability and cost-effective performance, it combines various open-source large language models with Ola's own Krutrim V2, which has 12 billion parameters. [1] [5] Its speech recognition is built to identify regional Indian languages, dialects, and accents. [9] Due to its integration with numerous apps and services, Kruti is context-aware and can proactively complete tasks. [10] [11] [12] [13] [14]

Initially connected only with Ola ecosystem services, Krutrim intends to expand and incorporate various Indian services into Kruti, with the goal of adding services from Blinkit , Swiggy , and Uber with respective voice command support. [15] [5] On 20 June 2025, Krutrim acquired the AI platform BharatSah‘AI’yak to increase its involvement in government, education, and agriculture projects. This acquisition will allow Kruti to assist in broadening the scope of BharatSah'AI'yak's work on India-centric, vernacular retrieval-augmented generation AI bots. [16]

Development

Kruti is designed to perform tasks with minimal user help, eliminating the need to switch between different applications. It can accept documents, images, and text as input. [17] The underlying agentic framework allows individual agents to work together sequentially or concurrently to efficiently complete a user's request. When a query is submitted, Kruti breaks it into smaller sub-tasks that are carried out via agent-to-agent protocols, reportedly with over 90% accuracy for faster performance. [1] [18] [19] This component of Ola Krutrim's AI stack, which includes AI training and inference systems and foundation models , allows Kruti to be modified for diverse industries like healthcare, education, and finance. [3] [20]

Kruti connects directly with company databases or APIs using a model context protocol to improve its training. It displays answers as summaries, tables, or narratives based on user patterns. [21] The system is designed to gradually learn user behavior and can handle online payments using credit / debit cards or UPI . [22] Ola Cabs and Open Network for Digital Commerce have begun implementing Kruti within their networks as it undergoes reliability testing for broader cross-platform support. [1]

See all

India portal

Haptik

OpenAI Operator

ChatGPT Deep Research

Manus

Devin AI

AutoGPT

References

v

t

e

Autoencoder

Deep learning

Fine-tuning

Foundation model

Generative adversarial network

Generative pre-trained transformer

Large language model

Model Context Protocol

Neural network

Prompt engineering

Reinforcement learning from human feedback

Retrieval-augmented generation

Self-supervised learning

Stochastic parrot

Synthetic data

Top-p sampling

Transformer

Variational autoencoder

Vibe coding

Vision transformer

Waluigi effect

Word embedding

Character.ai
ChatGPT
DeepSeek
Ernie
Gemini
Grok
Copilot
Claude
Gemini
Gemma
GPT 1 2 3 J 4 4o 4.5 4.1 OSS 5
1
2
3
J
4
4o
4.5
4.1
OSS
5
Llama
o1
o3
o4-mini
Qwen
Base44
Claude Code
Cursor
Devstral
GitHub Copilot
Kimi-Dev
Qwen3-Coder
Replit
Xcode
Aurora
Firefly
Flux

GPT Image 1
Ideogram
Imagen
Midjourney
Qwen-Image
Recraft
Seedream
Stable Diffusion
Dream Machine
Hailuo AI
Kling
Midjourney Video
Runway Gen
Seedance
Sora
Veo
Wan
15.ai
Eleven
MiniMax Speech 2.5
WaveNet
Eleven Music
Endel
Lyria
Riffusion
Suno AI
Udio
Agentforce
AutoGLM
AutoGPT
ChatGPT Agent
Devin AI
Manus
OpenAI Codex
Operator
Replit Agent
01.AI
Aleph Alpha

Anthropic
Baichuan
Canva
Cognition AI
Cohere
Contextual AI
DeepSeek
ElevenLabs
Google DeepMind
HeyGen
Hugging Face
Inflection AI
Krikey AI
Kuaishou
Luma Labs
Meta AI
MiniMax
Mistral AI
Moonshot AI
OpenAI
Perplexity AI
Runway
Safe Superintelligence
Salesforce
Scale AI
SoundHound
Stability AI
Synthesia
Thinking Machines Lab
Upstage
xAI
Z.ai
Category
v
t
e
History timeline
timeline

Companies

Projects

Parameter Hyperparameter

Hyperparameter

Loss functions

Regression Bias–variance tradeoff Double descent Overfitting

Bias–variance tradeoff

Double descent

Overfitting

Clustering

Gradient descent SGD Quasi-Newton method Conjugate gradient method

SGD

Quasi-Newton method

Conjugate gradient method

Backpropagation

Attention

Convolution

Normalization Batchnorm

Batchnorm

Activation Softmax Sigmoid Rectifier

Softmax

Sigmoid

Rectifier

Gating

Weight initialization

Regularization

Datasets Augmentation

Augmentation

Prompt engineering

Reinforcement learning Q-learning SARSA Imitation Policy gradient

Q-learning

SARSA

Imitation

Policy gradient

Diffusion

Latent diffusion model

Autoregression

Adversary

RAG

Uncanny valley

RLHF

Self-supervised learning

Reflection

Recursive self-improvement

Hallucination

Word embedding

Vibe coding

Machine learning In-context learning

In-context learning

Artificial neural network Deep learning

Deep learning

Language model Large language model NMT

Large language model

NMT

Reasoning language model

Model Context Protocol

Intelligent agent

Artificial human companion

Humanity's Last Exam

Artificial general intelligence (AGI)

AlexNet

WaveNet

Human image synthesis

HWR

OCR

Computer vision

Speech synthesis 15.ai ElevenLabs

15.ai

ElevenLabs

Speech recognition Whisper

Whisper

Facial recognition

AlphaFold

Text-to-image models Aurora DALL-E Firefly Flux Ideogram Imagen Midjourney Recraft Stable Diffusion

Aurora

DALL-E

Firefly

Flux

Ideogram

Imagen

Midjourney

Recraft

Stable Diffusion

Text-to-video models Dream Machine Runway Gen Hailuo AI Kling Sora Veo

Dream Machine

Runway Gen

Hailuo AI

Kling

Sora

Veo

Music generation Riffusion Suno AI Udio

Riffusion

Suno AI

Udio

Word2vec

Seq2seq

GloVe

BERT

T5

Llama

Chinchilla AI

PaLM

GPT 1 2 3 J ChatGPT 4 4o o1 o3 4.5 4.1 o4-mini 5

1

2

3

J

ChatGPT

4

4o

o1

o3

4.5

4.1

o4-mini

5

Claude

Gemini Gemini (language model) Gemma

Gemini (language model)

Gemma

Grok

LaMDA

BLOOM

DBRX

Project Debater

IBM Watson

IBM Watsonx

Granite

PanGu-Σ

DeepSeek

Qwen

AlphaGo

AlphaZero

OpenAI Five

Self-driving car

MuZero

Action selection AutoGPT

AutoGPT

Robot control

Alan Turing

Warren Sturgis McCulloch

Walter Pitts

John von Neumann

Claude Shannon

Shun'ichi Amari

Kunihiko Fukushima

Takeo Kanade

Marvin Minsky

John McCarthy

Nathaniel Rochester

Allen Newell

Cliff Shaw
Herbert A. Simon
Oliver Selfridge
Frank Rosenblatt
Bernard Widrow
Joseph Weizenbaum
Seymour Papert
Seppo Linnainmaa
Paul Werbos
Geoffrey Hinton
John Hopfield
Jürgen Schmidhuber
Yann LeCun
Yoshua Bengio
Lotfi A. Zadeh
Stephen Grossberg
Alex Graves
James Goodnight
Andrew Ng
Fei-Fei Li
Alex Krizhevsky
Ilya Sutskever
Oriol Vinyals
Quoc V. Le
Ian Goodfellow
Demis Hassabis
David Silver
Andrej Karpathy
Ashish Vaswani
Noam Shazeer
Aidan Gomez
John Schulman
Mustafa Suleyman
Jan Leike
Daniel Kokotajlo
François Chollet
Neural Turing machine
Differentiable neural computer

Transformer Vision transformer (ViT)

Vision transformer (ViT)

Recurrent neural network (RNN)

Long short-term memory (LSTM)

Gated recurrent unit (GRU)

Echo state network

Multilayer perceptron (MLP)

Convolutional neural network (CNN)

Residual neural network (RNN)

Highway network

Mamba

Autoencoder

Variational autoencoder (VAE)

Generative adversarial network (GAN)

Graph neural network (GNN)

Category