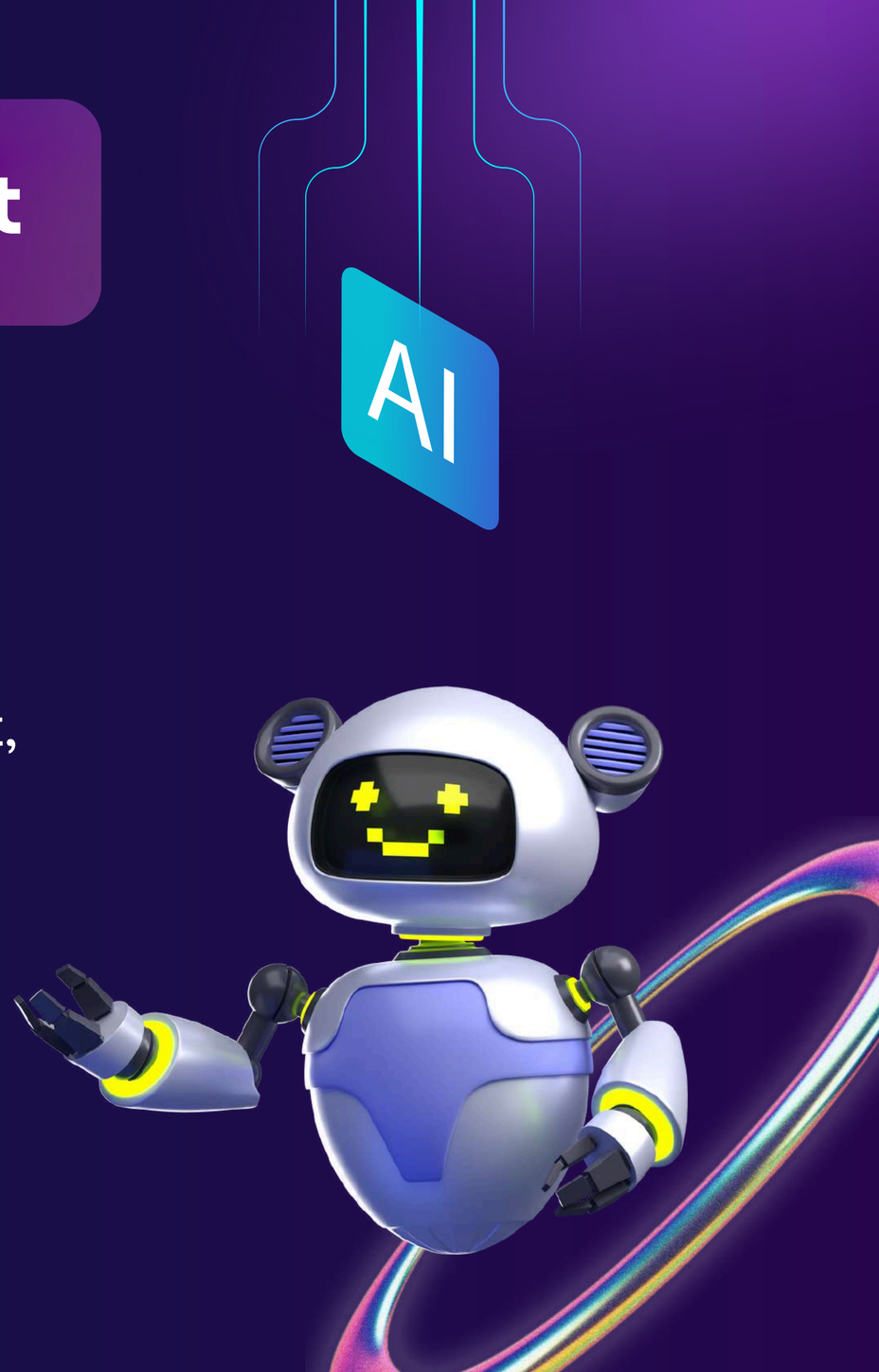
AI-Powered Social Media Caption Generator

PRESENTED BY
Lakshika Padmamali
Liaba Muneer
Aakash Kuragayala

Motivation and Problem Statement

- **Problem:** Manual generation of stylish, category-specific captions for social media is labor-intensive.
- **Challenges:** Maintaining Gen Z tone, trend alignment, and rapid adaptability across platforms.
- **Impact:** Inconsistent brand messaging and wasted creative effort.



Project Goals and Contributions

01

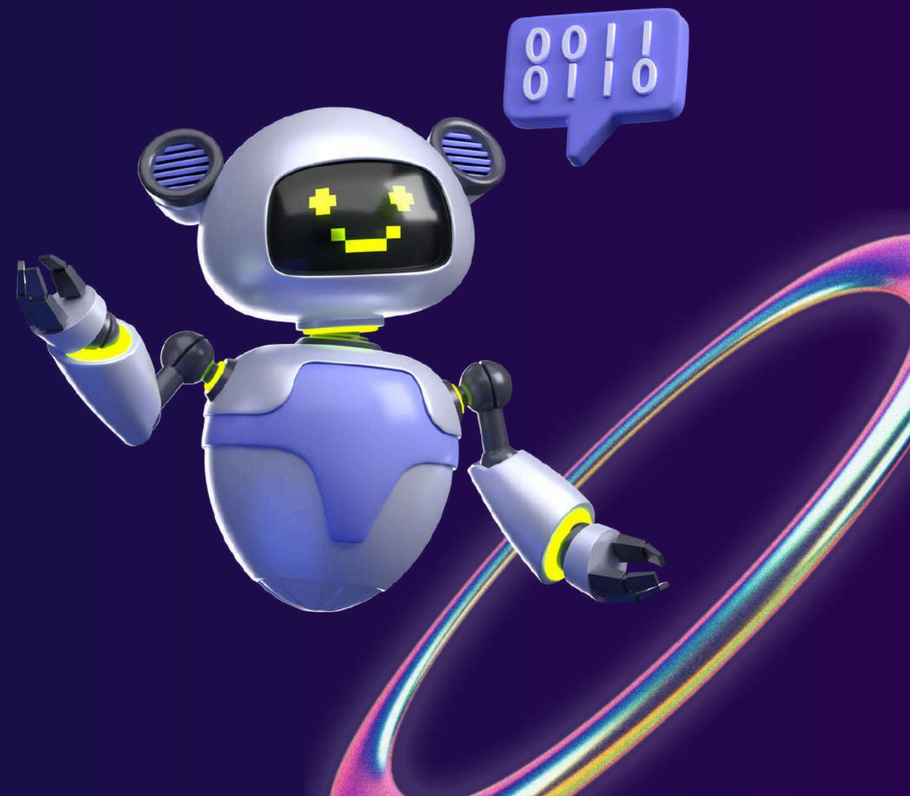
Goals:

- Automate high-quality caption creation
- Ensure contextual relevance and stylistic alignment○ Enable scalable, trend-aware, Gen Z-style captioning
- Deploy in real-time for non-technical users

02

Core Contributions:

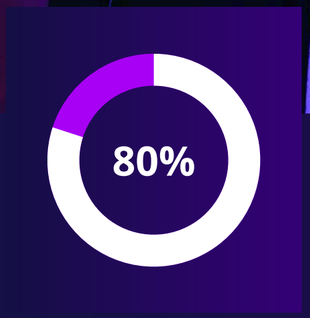
- Authentically scraped dataset
- RAG + CoT-enhanced DistilGPT2 model
- Interactive deployment interfaces (CLI & Web)

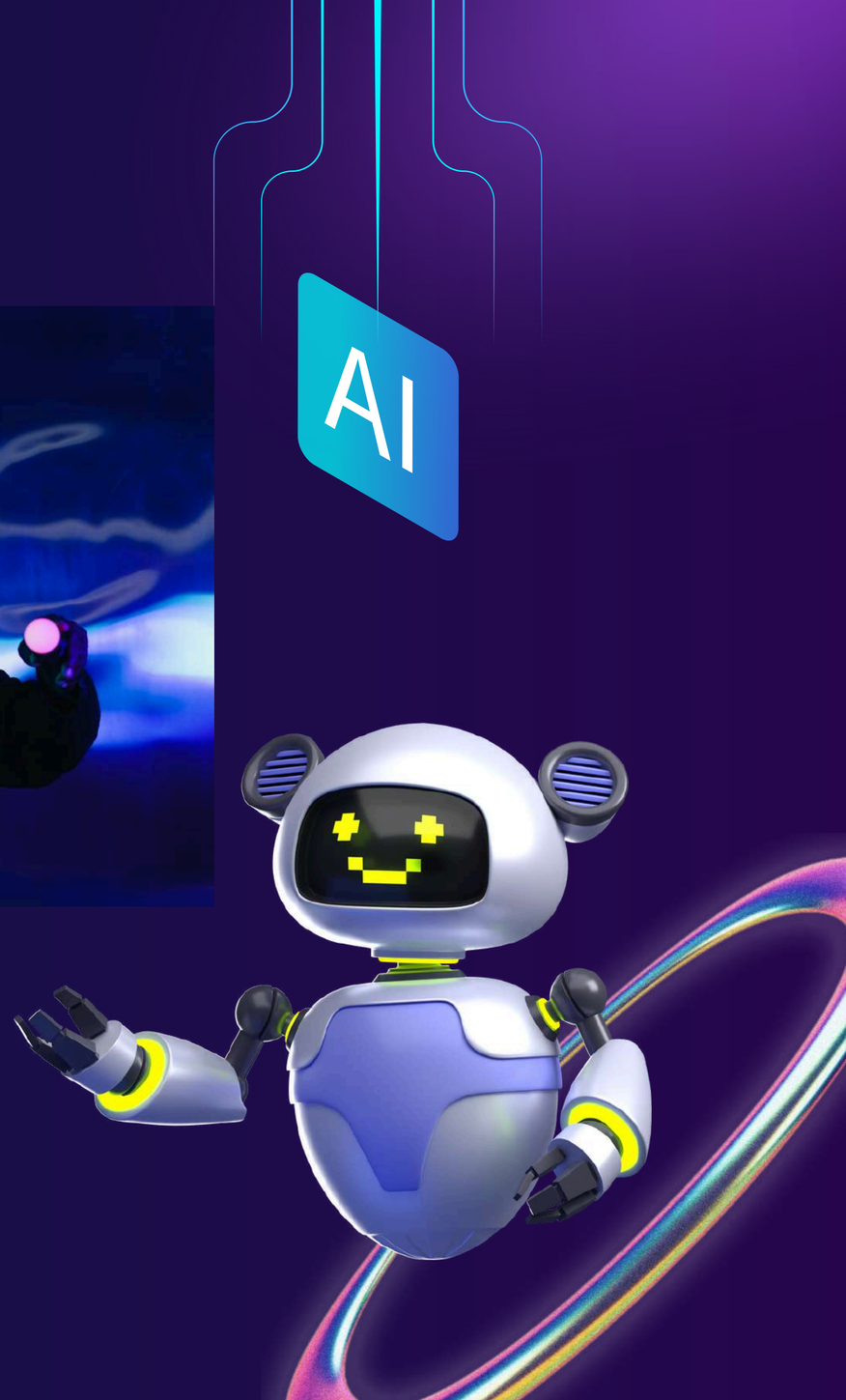


System Architecture Overview

Pipeline Modules:

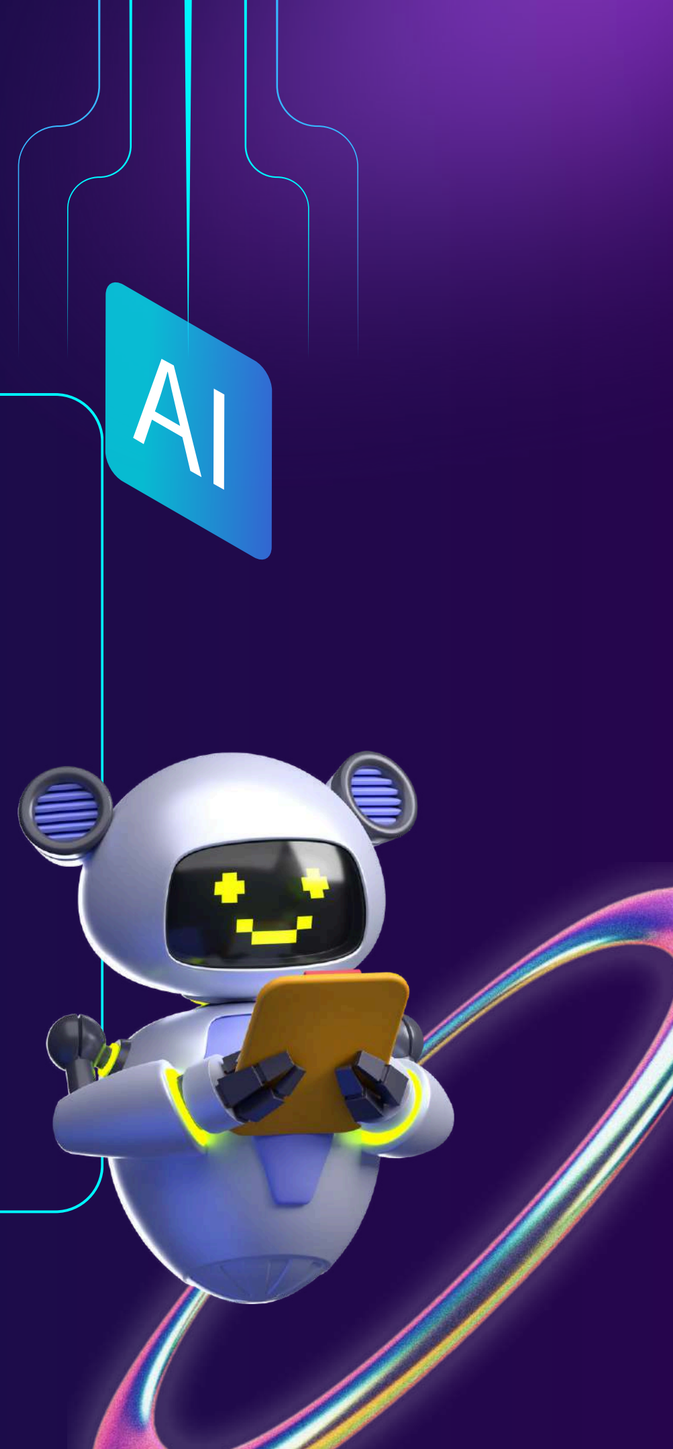
1. Data Acquisition via Web Scraping
2. Preprocessing and EDA
3. DistilGPT2 Fine-Tuning
4. Retrieval-Augmented Generation (FAISS + MiniLM)
5. Chain-of-Thought Prompting
6. User Interfaces (CLI & Web)





Data Acquisition and Preprocessing

- Source Sites: NDTV, Hindustan Times, Vogue India
- Scraping Tags: , ,
- Filter Criteria: ≤ 280 characters, presence of promotional keywords
- Cleaning: Token normalization, punctuation fix, whitespace trimming
- Synthetic Data Creation: Due to insufficient dataset quality, we utilized AI models (e.g., ChatGPT, Grok, Gemini, Copilot) to generate synthetic data or captions, ensuring robust and diverse content for analysis.

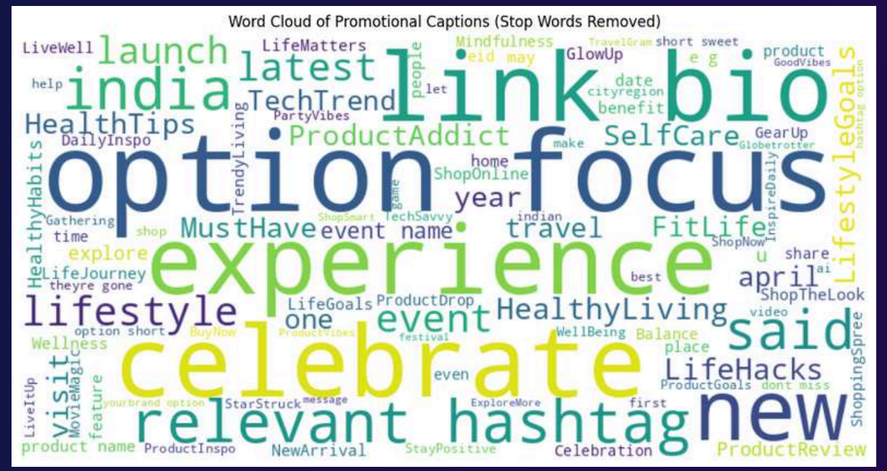


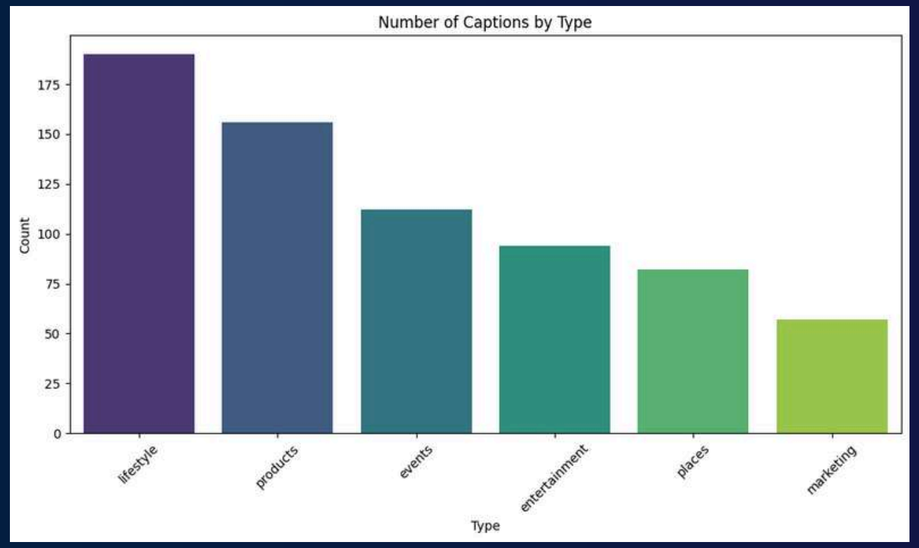
Exploratory Data Analysis (EDA)

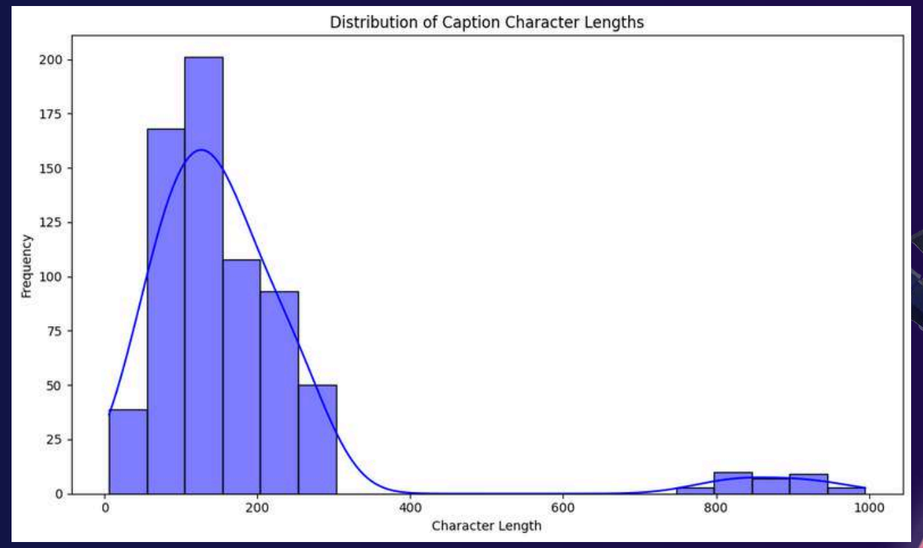
Dataset Stats:

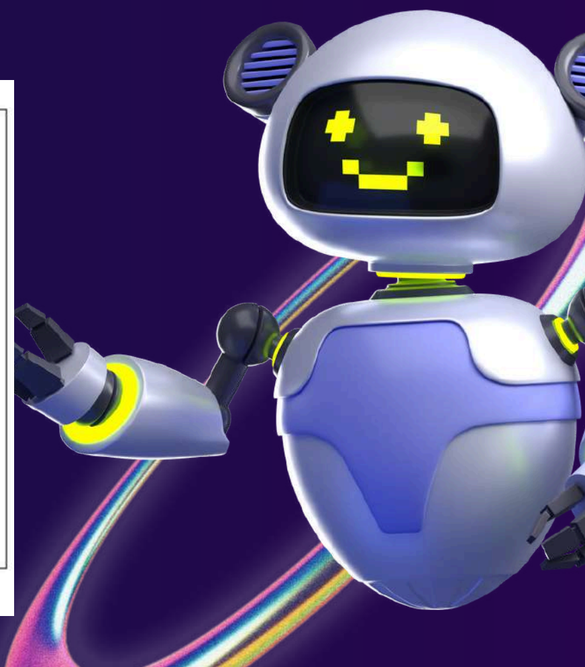
- 981 unique captions
- 6 domain categories
- Avg. 108.2 characters, 16.1 words

Visualization Highlights:









Language Model Fine-Tuning

- Model: DistilGPT2
- **Training:**

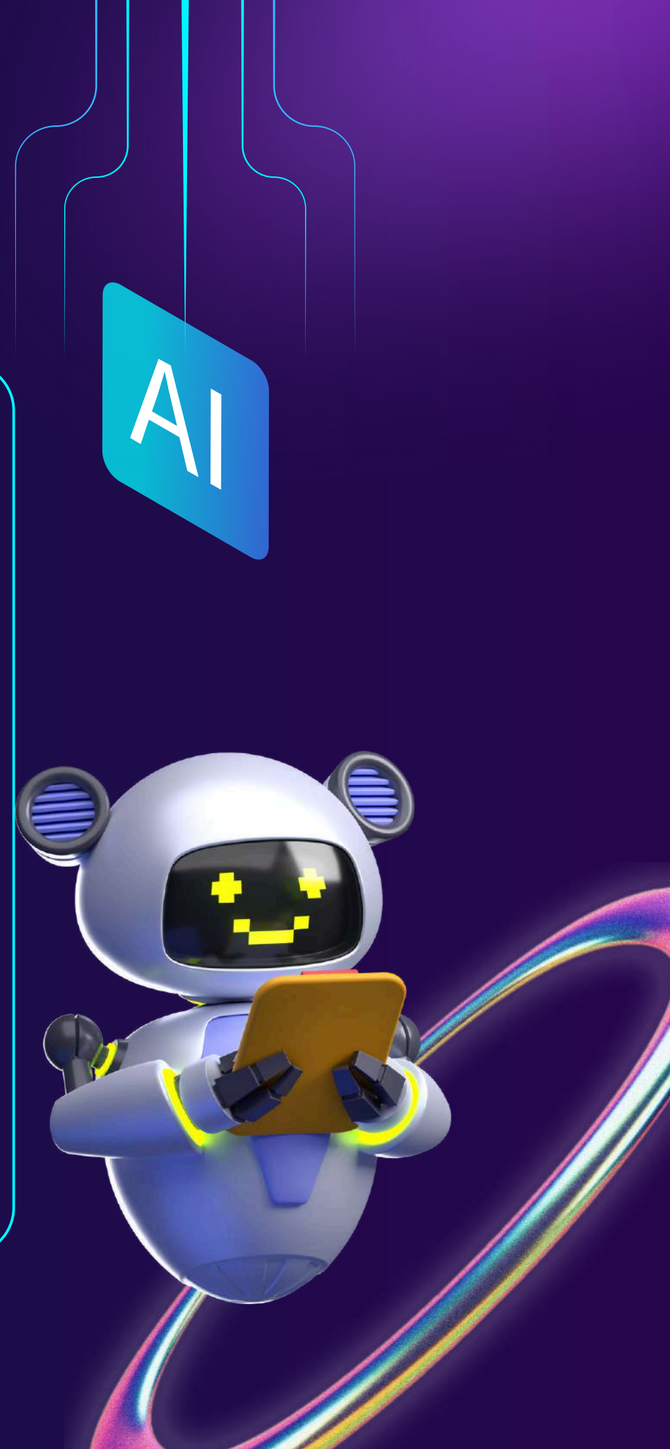
  5 epochs

  Loss: 2.70 → 1. 92
- **Fine-Tuning:**

  2 epochs

  Loss: 1. 91 → 1. 54

- Result: Fast convergence and style adaptation



Retrieval-Augmented Generation (RAG)

- Embedding Model: all-MiniLM-L6-v2
- Semantic Index: FAISS (FlatL2)
- Inference Use: Top-K captions retrieved to guide generation
- Speed: Retrieval latency under 100ms

Chain-of-Thought (CoT) Prompting

Steps in Prompting:

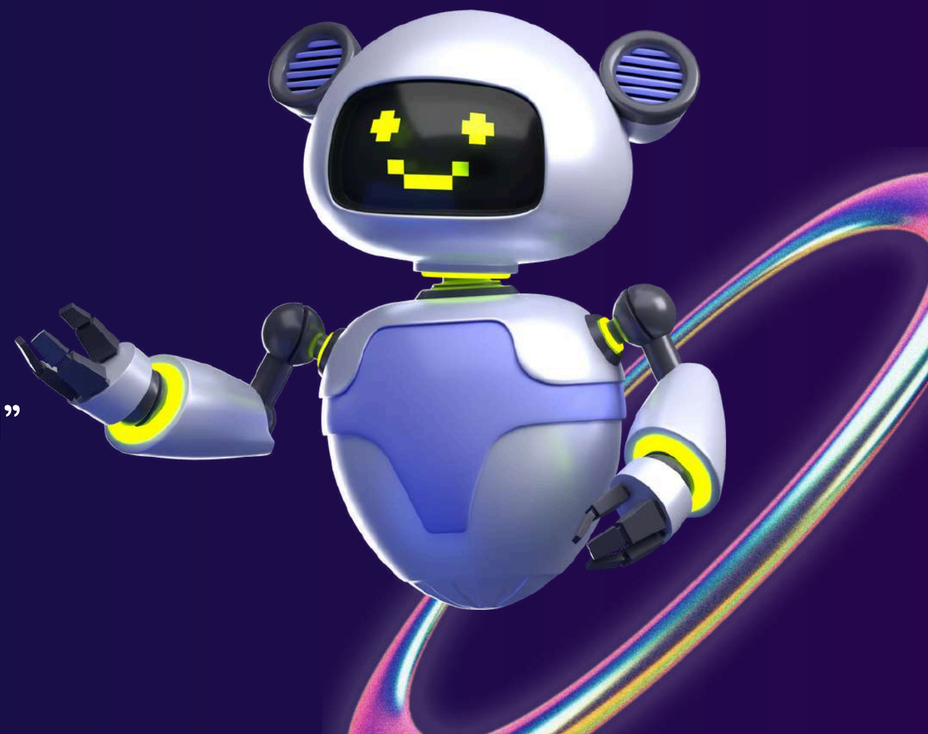
- User input analysis
- Contextual injection (from RAG)
- CoT-based caption output

Effect:

- Logical, structured, expressive captions

Example:

- Input: “Throwing a party this weekend”
- Caption: “Weekend turn-up loading... 🎉 #PartySzn#WeekendMood”



Output Examples and Variability

Prompt 1: “Launching a smart fitness band”

- Output: “Track every move in style! 🔥 #FitDrop #HealthGoals”

Prompt 2: “Introducing our new smartwatch”

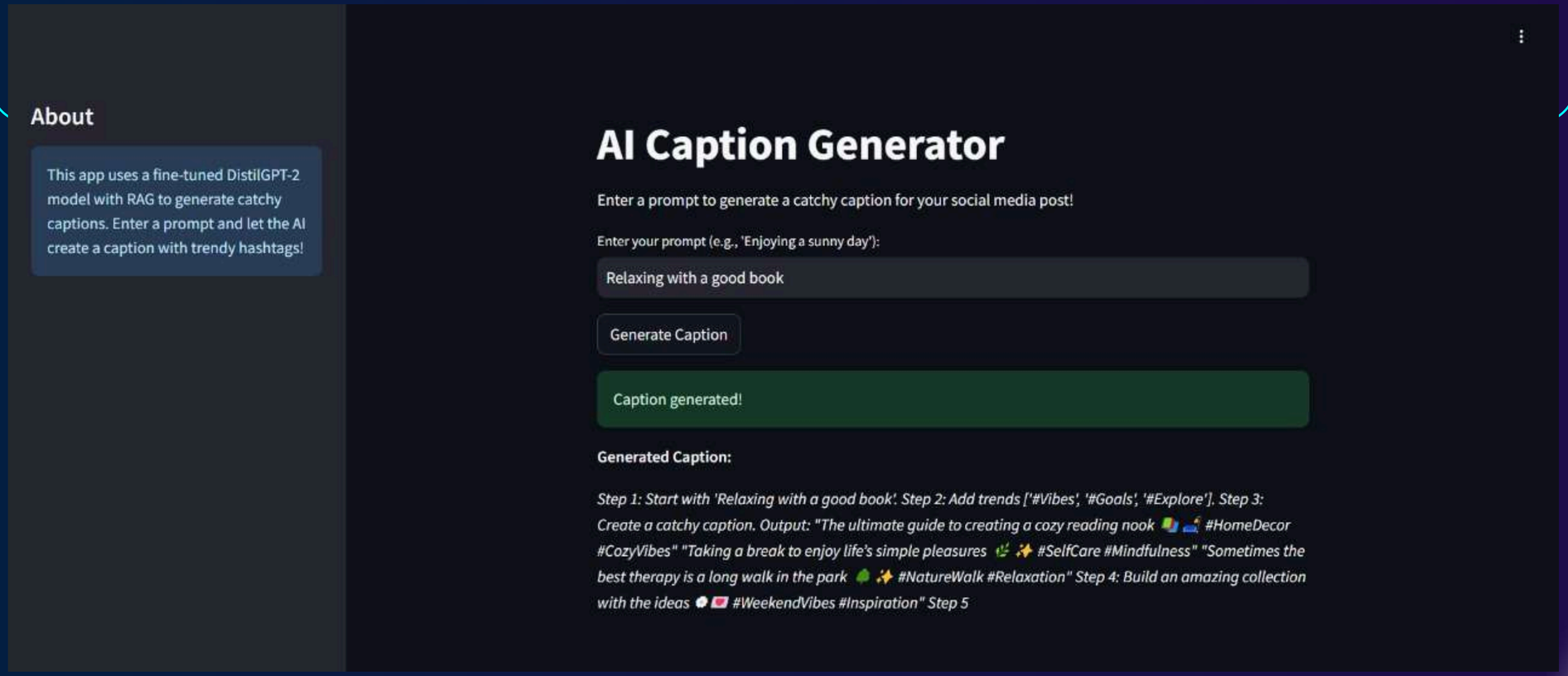
- Output: “Your wrist just got smarter ⌚ #NextGenGear #SmartTech”

Prompt 3: “Beach party this Friday”

- Output: “Sunset vibes & sand beneath our feet 🌊 #BeachBash #TurnUp

Deployment Interfaces

- CLI Chatbot: Scripting-friendly, text-only
- Streamlit Web App: Real-time captioning with input box and instant output
- Shared Backend: CoT + RAG + DistilGPT2 pipeline



Evaluation and Results

Training Convergence:

- Loss reduced smoothly over epochs

Retrieval Accuracy:

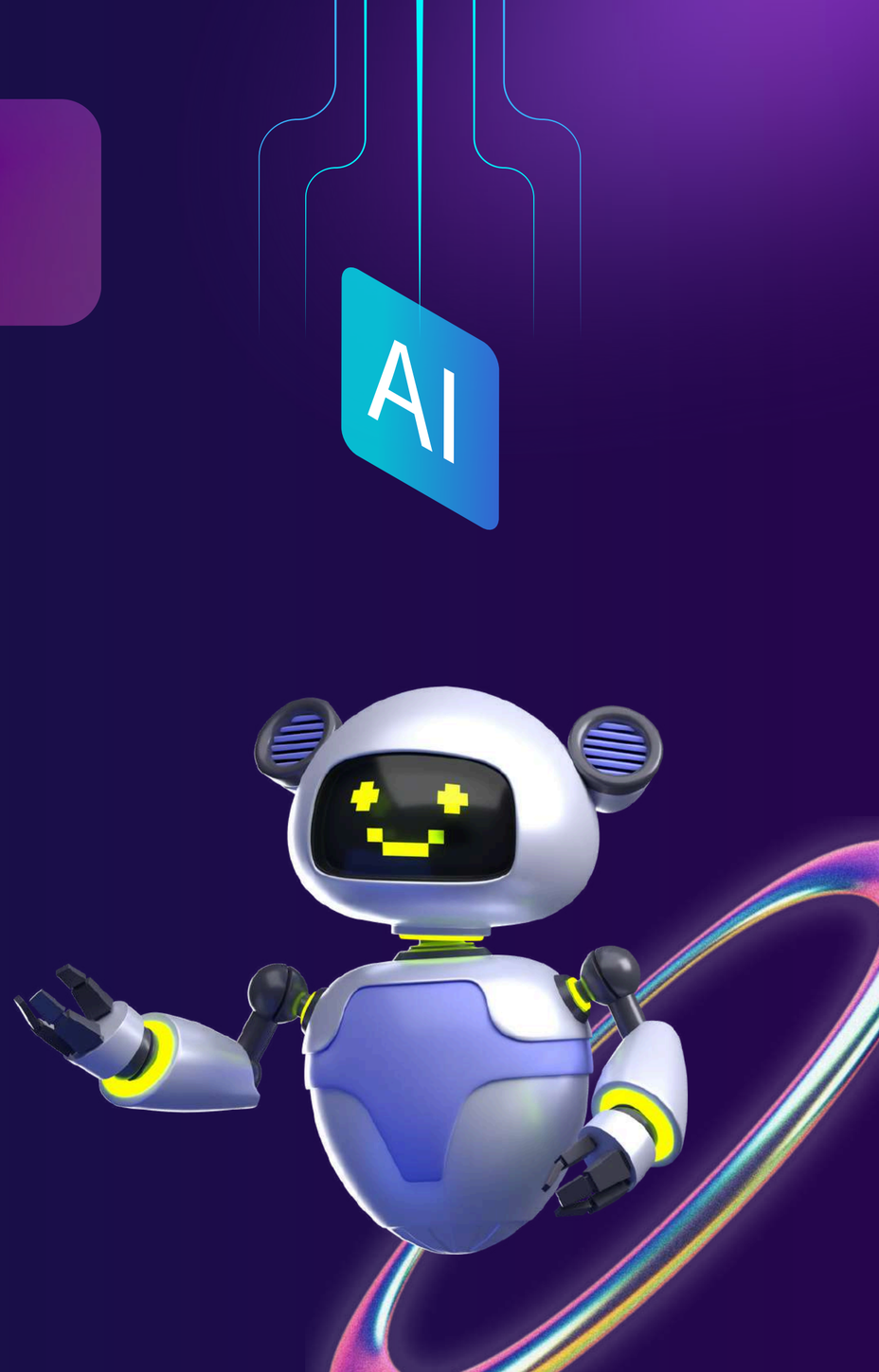
- Top-3 retrieved captions aligned contextually

Generation Quality:

- Human-like, stylistically accurate

Latency:

- Full pipeline < 1 second



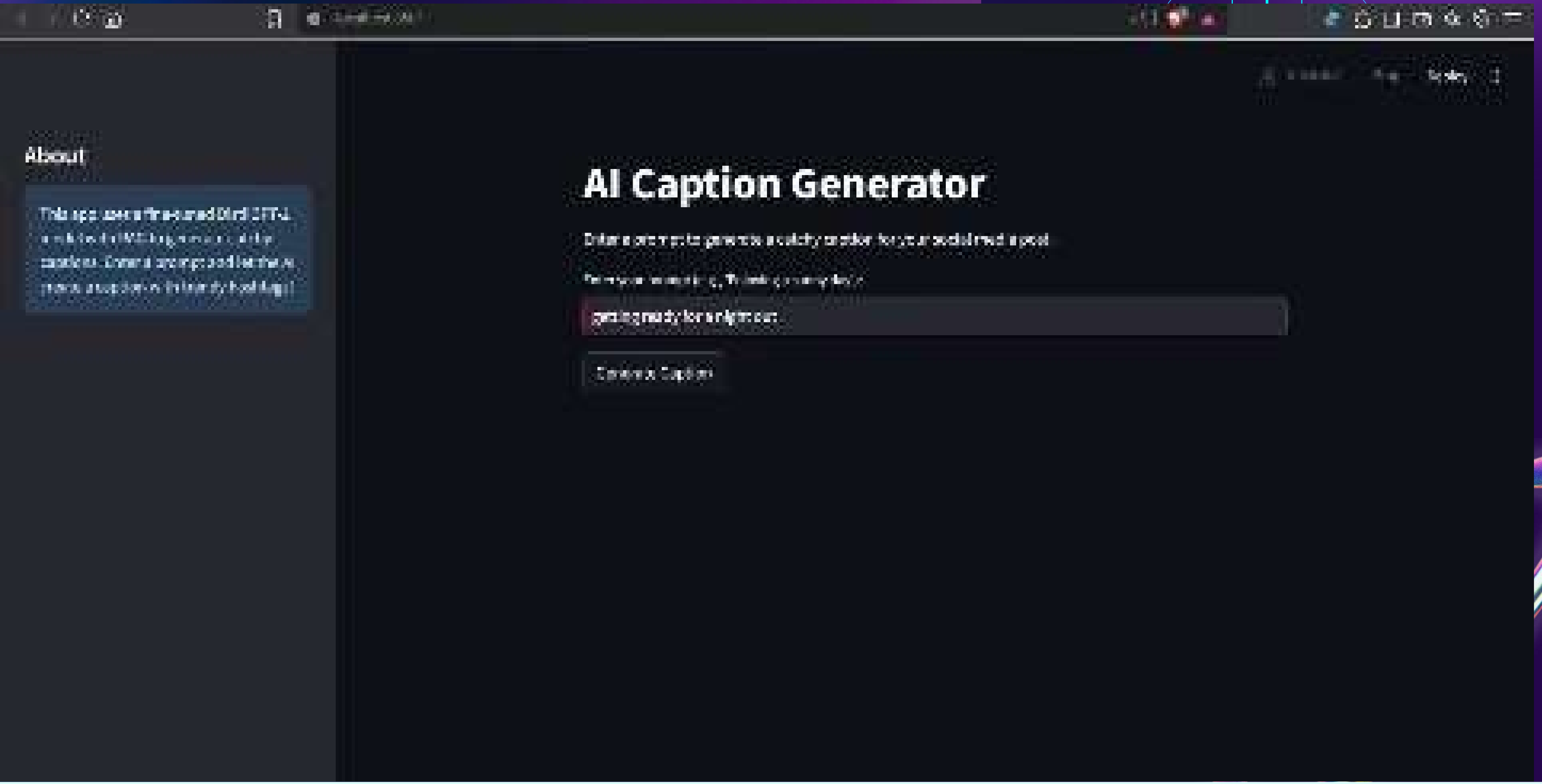
Demo Video

Title: Real-Time Caption Generation Demo

Content:

- User input prompt
- Retrieval + CoT in action
- Live output caption with hashtags

Demo Video



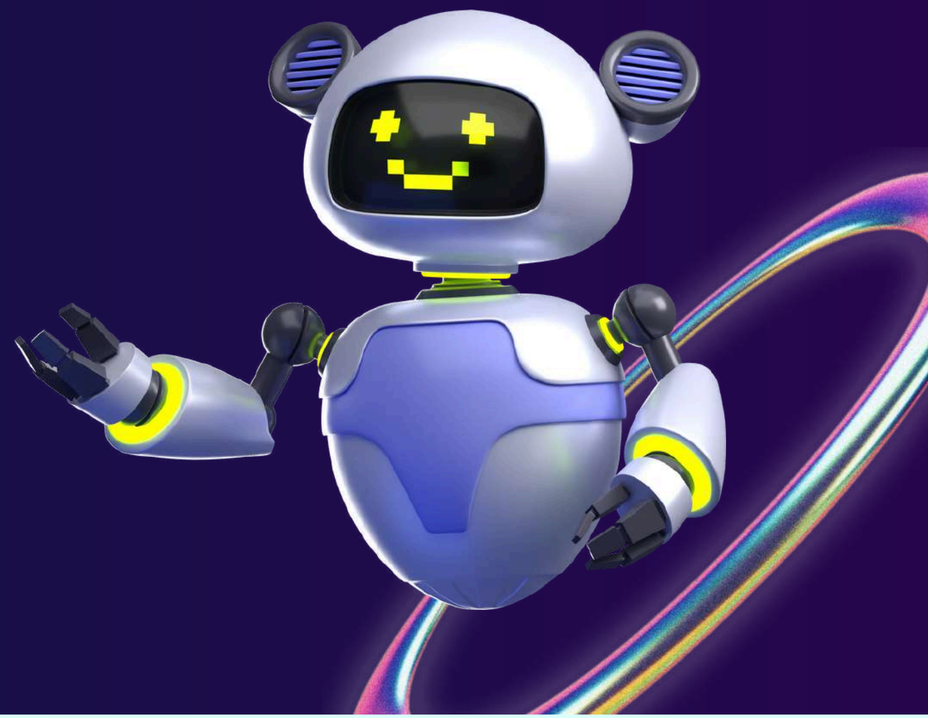
Limitations and Future Work

Current Gaps:

- No engagement predictor
- No trending hashtag auto-injection
- English-only generation

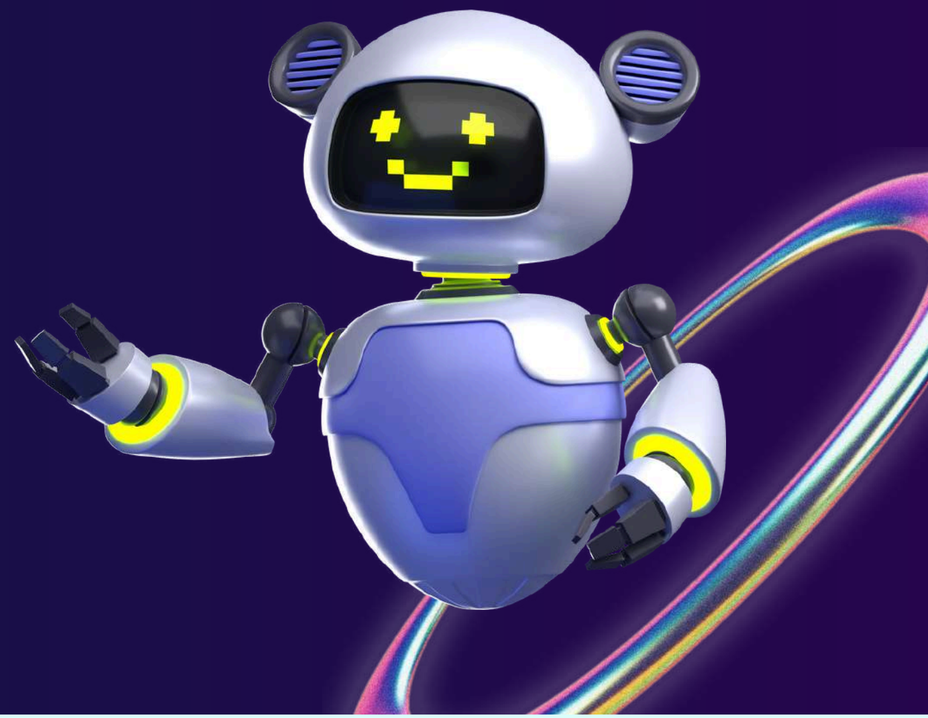
Future Enhancements:

- Multimodal input (BLIP/ViLT)
- Real-time trend APIs (Google/Twitter Trends)
- Multilingual model training (Hindi, Thai, Spanish)



Conclusion

- System proves automation of stylish, engaging captions is viable
- CoT + RAG improves quality beyond basic language models
- Modular, real-time system can serve brands, marketers, and creators
- Ready for multilingual and visual expansion



THANK YOU