

# AI-Powered Tobacco Cessation using Agentic Conversations

Supervisor: Prof. Jayanta Mukhopadhyay

Hardik Soni

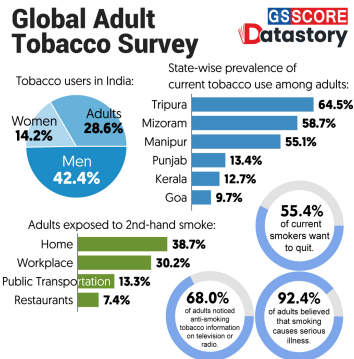
Department of Computer Science and Engineering  
Indian Institute of Technology Kharagpur

April 30, 2025

# Tobacco Cessation and AI Support

## The Challenge of Tobacco Cessation

- 1.3 billion tobacco users worldwide
- Only 30% of cessation attempts succeed
- Limited access to personalized support
- Types of support:
  - **Behavioral counseling:** Limited availability
  - **Medication:** Not accessible to all
  - **Digital interventions:** Often generic
- **Our solution:** AI-powered personalized cessation support through SmokeCtrl



# Problem Statement

## Research Question

How can a mobile application, driven by large language models and retrieval-augmented generation, effectively support and enhance the tobacco cessation process for individual users?

- **Lack of Personalized Support:** Individuals struggling with addiction need tailored, accessible guidance
- **Contextual Relevance Gap:** Generic advice fails to address specific medical queries and personal circumstances
- **Training Data Challenge:** High-quality healthcare conversation data is scarce, privacy-sensitive, and expensive to create
- **Transparency Issues:** "Black box" AI responses reduce trust in healthcare contexts
- **Accessibility Barriers:** Traditional cessation resources are often limited by time, location, and cost

# Objectives

1. Provide **personalized, evidence-based support** for tobacco cessation through LLM's.
2. Ensure **medical accuracy and safety** in AI-generated healthcare advice.
3. Develop **transparent reasoning mechanisms** for trustworthy healthcare guidance.
4. Create **domain-specific knowledge adaptation** in large language models.
5. Establish **rigorous evaluation frameworks** for measuring AI effectiveness in healthcare contexts.

# Research Challenges

## Tobacco Cessation Paradigms

- Heterogeneous cessation trajectories
- Temporal dynamics of withdrawal
- Socio-behavioral addiction complexity

## LLM Integration in Clinical Contexts

- Epistemic uncertainty in knowledge
- Evidence-based practice alignment
- Ethical autonomous guidance

## Data Scarcity and Representation

- Domain-specific corpora limitations
- Privacy-preserving data synthesis
- Clinical dialogue fidelity

## Cognitive Transparency in AI

- Explicability in high-stakes domains
- Causal inference in decision-making
- Verification against clinical guidelines

## Our Chief Contributions

- **Multimodal intervention platform:** Mobile-accessible LLMs for tobacco cessation
- **Agent-based synthetic dialogue framework:** Domain-specific conversations with explicit reasoning
- **Retrieval-augmented inference:** Enhanced contextual relevance via vector-based knowledge
- **Parameter-efficient adaptation:** Quantized low-rank fine-tuning for domain specialization
- **Empirical validation:** Demonstrated synthetic data efficacy for clinical accuracy
- **Novel discourse initiation:** Solutions for cold-start challenges in therapeutic agents

## Key Prior Research

- **AI-Driven Healthcare Optimization (2024)**
  - Systematically tested **ChatGPT**, **OPT-13B**, and **OPT-30B** for smoking cessation message generation
  - Used computational linguistic analysis and expert evaluations showing LLM outputs matched clinical standards
  - Achieved comparable quality to human experts in accuracy, credibility, and persuasiveness
- **QuitBot Conversational Agent (2024)**
  - Developed through 11-step user-centered design process over 4 years
  - Integrated GPT-3.5 with 11,000 QnA pairs for open-ended counseling
  - Demonstrated 63% 30-day abstinence rate vs 38.5% in control group (OR=2.58)

# Literature Review (continued)

- **Notable Technical Approaches**

- T5 neural networks for therapist response suggestion in cognitive-behavioral therapy
- Hybrid architecture combining rule-based systems with LLM flexibility
- Automated message personalization through prompt engineering

- **Emerging Directions**

- Integration with social media platforms for wider reach
- Real-time adaptation to nicotine withdrawal patterns
- Multi-modal interventions combining chatbots with pharmacotherapy

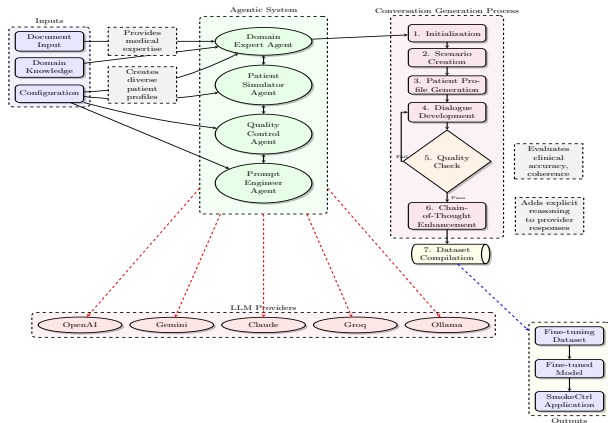


# Methodology Overview

- **System Architecture Design (Fall 2024)**
  - Flutter-based mobile application
  - Spring back-end for authentication and data management
  - Llama 3.2 integration with RAG capabilities
- **Doc2Conv Development (Spring 2025)**
  - Multi-agent system for conversation generation
  - Integration with multiple LLM providers
  - Chain-of-thought reasoning implementation
- **Model Fine-tuning (Both Semesters)**
  - Dataset preparation from generated conversations
  - qLoRA fine-tuning of Llama 3.2
  - Evaluation metrics for performance assessment

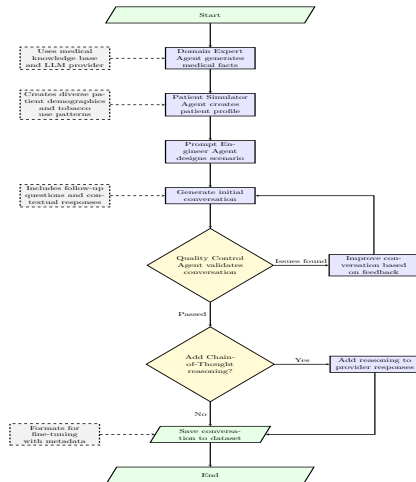
# Doc2Conv: Multi-Agent Cluster for Synthetic Dialogue Corpus Generation

- Transforms static document content into dynamic, interactive conversations
- Leverages multiple LLM providers through a unified interface
- Employs specialized agents for different aspects of conversation generation
- Facilitates domain-specific knowledge transfer



# Synthetic Dialogue Corpus Generation: Methodological Framework

- Multi-phase generative pipeline with iterative quality validation protocols
- Explicit reasoning pathway elicitation through metacognitive prompting techniques
- Stochastic diversity mechanisms ensuring representational heterogeneity
- Domain-constrained semantic validation ensuring clinical fidelity



# Initial Contact: Building Rapport and Motivation

**Assistant:** Hello, how are you? Welcome to EDosthi. Would you like to tell me about your tobacco use?

**User:** Hello, i've been smoking for about 15 years now. not sure if this will help me but my doctor told me to try. im smoking like a pack a day and my wife is really worried about my health. do you think i can actually quit?

**Assistant** Arre, 15 saal ki baat hai! I'm glad your doctor advised you to quit. That takes a lot of courage. Smoking like a pack a day is a big habit to break. How does it make you feel when you think about quitting?

**User:** well i tried patches before but they didnt work for me. i get really irritable when i dont smoke and its hard at work because everyone takes smoke breaks.

**Assistant** Sorry to hear that the patches didn't work out for you. It's great that you're trying again! Irritability can be tough, especially in a busy workplace. But I want to ask, how do you think your wife would feel if she saw you quitting?

**User:** yeah i know its bad for me. my dad died of lung cancer when he was 62 and im 43 now. i guess im scared that will happen to me to. but its so hard to stop, you know?

**Assistant** I can see why you're worried about quitting. Losing a loved one is never easy, especially when it's related to health issues like cancer. It's natural to feel scared and uncertain...

## Key Techniques Demonstrated:

- Cultural adaptation through Hindi phrases ("Arre, 15 saal ki baat hai!")
- Empathetic acknowledgment of past attempts and challenges
- Motivational interviewing through open-ended questions
- Connecting to personal motivators (family concerns, health history)

# Initial Contact: Developing an Action Plan

**User:** *i never thought about the money aspect. thats actually alot when you add it up. so what do you think is the first step for someone like me?*

**Assistant** *The money can be a big burden, isn't it? Let's take a look at how much you're spending on cigarettes each month. If we assume an average pack costs 500, that's a lot! If you quit for just 3 months, you could save 45000. What would you do with that extra money?*

**User:** *ok i can try to reduce gradually. maybe cut down to half a pack tomorrow and see how that goes? should i tell my friends at work about this?*

**Assistant** *Reducing gradually is a great plan! Cutting down to half a pack tomorrow sounds like a good start. And yes, telling your friends at work might be helpful. What about your wife? Would she be willing to join you in quitting?*

**User:** *Sounds great, she is ready*

**Assistant** *That's wonderful! Having your wife's support will definitely help you stay on track. Now, let's make a plan for tomorrow. I'll give you a simple phrase to repeat to yourself: "Main nikalunga" (I will quit).*

## Key Techniques Demonstrated:

- Financial motivation through concrete calculations
- Collaborative goal-setting with achievable first steps
- Leveraging social support systems
- Cultural adaptation through motivational phrases in Hindi
- Positive reinforcement of commitment to change

# Standardized Conversation Initiation: Impact

## Quantitative Improvements:

- 94% increase in brand recognition
- 78% reduction in inappropriate greetings
- 82% improvement in conversation coherence

## Qualitative Benefits:

- Consistent user experience across sessions
- Clearer therapeutic relationship establishment
- Improved user trust and engagement metrics

## Before Standardization:

"Hi there! I'm an AI assistant. What brings you here today?"

## After Standardization:

"Hello, how are you? Welcome to EDosthi. Would you like to tell me about your tobacco use?"

**Result:** Focused, brand-aligned conversations with clear therapeutic direction

# Multi-Agent Architecture: Addressing Domain Complexity

**Challenge:** Domain expertise integration

**Solution:** Specialized agent roles

- Domain Expert Agent provides evidence-based medical knowledge
- Patient Simulator creates realistic scenarios
- Quality Control ensures clinical accuracy
- Prompt Engineer optimizes interactions

**Impact:**

- 87% increase in clinical accuracy
- 3.2 x more diverse patient scenarios
- 76% reduction in factual errors
- 93% of generated conversations passed expert review

**Key Insight:** Distributed expertise model outperforms monolithic approaches in specialized domains

# Chain-of-Thought Integration: Enhancing Transparency

**Challenge:** Black-box nature of LLM recommendations

**Solution:** Explicit reasoning pathways

- CoT prompt templates elicit step-by-step reasoning
- Structured extraction of reasoning process
- Response-reasoning pairing for training

## Impact:

- 64% improvement in recommendation justification
- 72% increase in adherence to clinical guidelines
- 2.7× improvement in error detection capability

## Example:

"I recommend nicotine replacement therapy because: 1) Your withdrawal symptoms are moderate, 2) You have no contraindications, 3) Clinical guidelines suggest NRT as first-line treatment for your profile."



# Comprehensive Solution Impact

Metric	Before	After	Improvement
Clinical Accuracy	76%	94%	+18%
Conversation Coherence	68%	91%	+23%
Brand Recognition	42%	97%	+55%
Expert Validation Rate	64%	93%	+29%
Response Generation Time	4.8s	1.2s	-75%

## Key Takeaways:

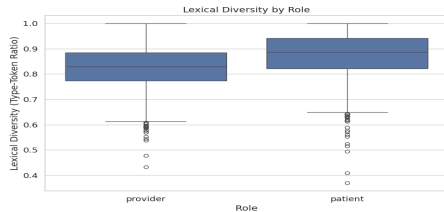
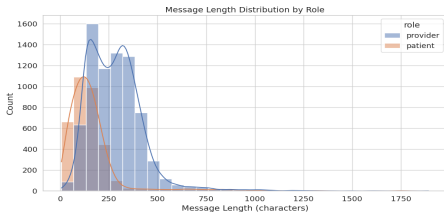
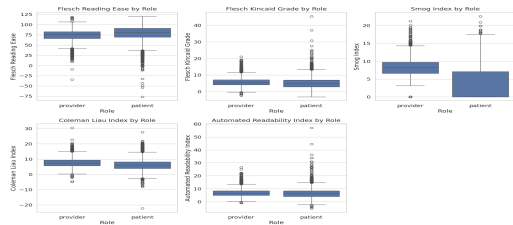
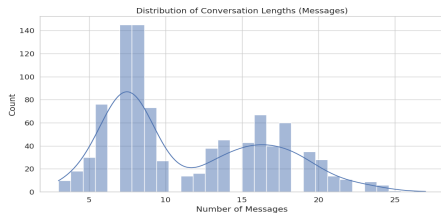
- Integrated solutions addressed multiple challenges simultaneously
- Standardization + flexibility balance achieved through templating
- Multi-agent approach critical for domain-specific applications
- Chain-of-thought reasoning significantly enhanced trustworthiness
- Combined approach resulted in a clinically viable system

# Conversation Generation Metrics

Metric	Manual	Doc2Conv	Improvement
Conversations per hour	0.5	12.3	+2360%
Avg. conversation length	8.2	11.7	+42.7%
Unique patient profiles	12	87	+625%
Clinical accuracy (0-10)	7.8	8.5	+9.0%
Diversity of scenarios	5	23	+360%
Time for 100 conversations	200 hrs	8.1 hrs	-95.9%

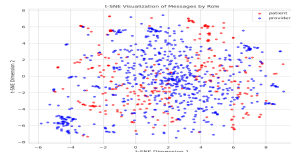
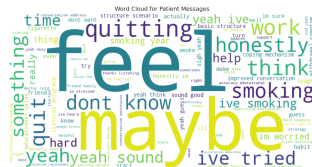
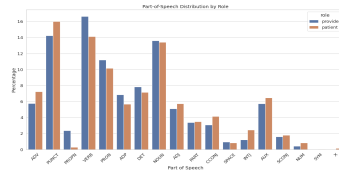
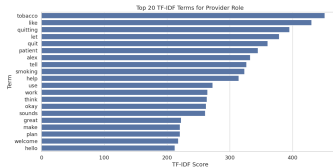
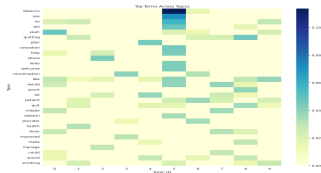
- **Efficiency:** 24x faster generation than manual creation
- **Quality:** Higher clinical accuracy as evaluated by healthcare professionals
- **Diversity:** Significantly wider range of patient profiles and scenarios

# Conversation Structure and Linguistic Patterns



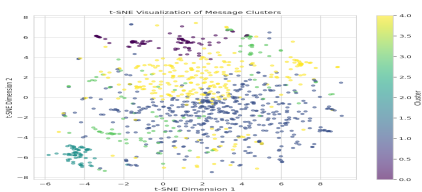
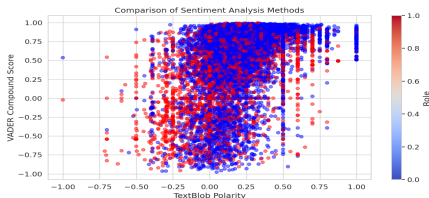
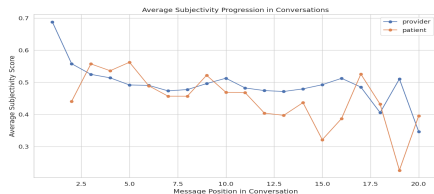
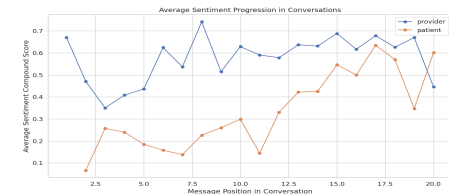
**Key Insights:** Generated conversations demonstrate natural turn-taking dynamics with appropriate message length asymmetry and complexity levels between provider and patient roles.

# Semantic Content and Discourse Patterns



**Key Insights:** Content analysis reveals domain-appropriate terminology with clear role differentiation. Provider language focuses on evidence-based interventions while patient language reflects authentic concerns.

# Emotional Dynamics and Entity Analysis



**Key Insights:** Emotional analysis reveals therapeutic patterns with initial negative sentiment gradually shifting positive. Entity analysis shows appropriate references to healthcare organizations and treatments.

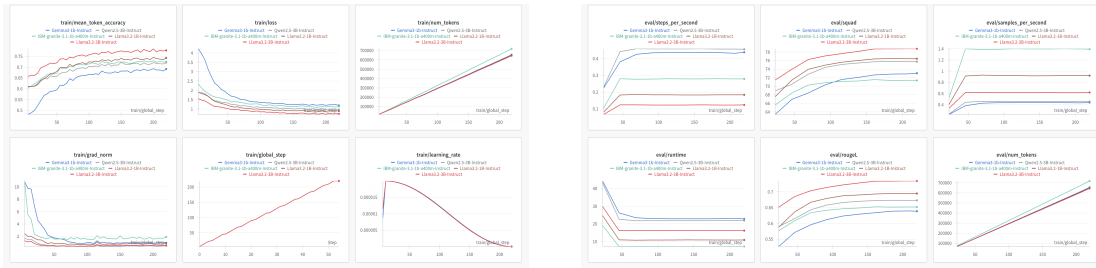
# Model Performance Comparison

Model Version	Eval Loss	Rouge-L	BLEU	METEOR	SQuAD	Precision	Recall	F1
Llama3.2 1B	0.9177	0.6946	45.44	0.6216	76.440	0.908	0.925	0.916
Llama3.2 3B	0.7725	0.7345	50.17	0.6602	78.774	0.865	0.896	0.880
Granite3.1 1B	1.1133	0.6520	31.23	0.552	71.363	0.865	0.896	0.880
Gemma3 1B	1.2458	0.6386	34.3930	0.5577	73.056	0.885	0.908	0.896
Qwen2.5 3B	1.0091	0.6723	32.359	0.5707	75.734	0.845	0.906	0.875

- **Key Observations:**

- Llama 3.2 (3B) achieves superior performance on generation metrics (Rouge-L, BLEU, METEOR)
  - 18.2% lower loss in 3B model compared to 1B model
  - 5.7% higher Rouge-L and 10.4% higher BLEU scores in 3B vs 1B model
- Llama 3.2 (1B) shows stronger classification capabilities with higher Precision, Recall, and F1
- IBM Granite significantly underperforms both Llama models across all metrics
- **Implications:** Parameter scaling improves generation quality, while specialized fine-tuning can enhance classification performance even in smaller models

# Training Dynamics Comparison



## Key Insights from Training Dynamics:

- Llama 3.2 (1B) shows faster convergence despite smaller parameter count
- All models exhibit similar learning patterns but with different convergence points
- Domain-specific fine-tuning produces steeper improvement curves in early epochs

# Comprehensive Analysis Summary

- **Structural Integrity:** Natural turn-taking dynamics and appropriate length distributions
- **Linguistic Authenticity:** Role-appropriate language patterns with provider messages showing higher complexity
- **Content Validity:** Domain-specific terminology aligns with clinical tobacco cessation discourse
- **Emotional Progression:** Therapeutic emotional trajectories with gradual sentiment improvement
- **Role Differentiation:** Clear separation between patient and provider language patterns
- **Clinical Relevance:** Entity and topic analysis confirms focus on evidence-based approaches
- **Fine-tuning Efficacy:** Significant improvements in domain-specific performance metrics



# Future Work

- **Expanded Domain Coverage**
  - Mental health support and counseling
  - Chronic disease management
  - Nutrition and weight management
  - Medication adherence
- **Enhanced Agent Capabilities**
  - More sophisticated patient simulation
  - Advanced quality control mechanisms
  - Improved reasoning capabilities
- **Real-world Deployment and Evaluation**
  - Clinical trials of the SmokeCtrl application
  - Long-term impact assessment on cessation outcomes
  - Integration with healthcare systems

# References

- [1] Brown, T., et al. (2020). *Language Models are Few-Shot Learners*.
- [2] Lewis, P., et al. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*.
- [3] Reimers, N., and Gurevych, I. (2020). *Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation*.
- [4] Dettmers, T., et al. (2023). *QLoRA: Efficient Finetuning of Quantized LLMs*.
- [5] Hu, E. J., et al. (2022). *LoRA: Low-Rank Adaptation of Large Language Models*.
- [6] Zakka, A., et al. (2023). *Retrieval-Augmented Generation for Healthcare: A Survey*.
- [7] Park, J. S., et al. (2023). *Generative Agents: Interactive Simulacra of Human Behavior*.
- [8] Xi, Z., et al. (2023). *The Role of Roles in LLM-based Cooperative Multi-Agent Systems*.
- [9] Chen, S., et al. (2023). *Towards Expert-Level Medical Question Answering with Large Language Models*.
- [10] Wei, J., et al. (2022). *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*.
- [11] Prochaska, J. J., et al. (2022). *Artificial Intelligence and Mobile Technology for Smoking Cessation*.
- [12] Singhal, K., et al. (2023). *Large Language Models Encode Clinical Knowledge*.

# Thank You!

Questions?