

Suman Adhya

Research Associate-I @ [IACS](#)

Jadavpur, Kolkata-700032, West Bengal, INDIA

 +91 9007657578 |  adhyasuman30@gmail.com |  AdhyaSuman |  [sumanadhyaa](#)

SKILLS & EXPERTISE

GenAI Stack	LLMs, RAG, Prompt Engineering
Languages	Python (Proficient), C, R
Frameworks	PyTorch, Hugging Face, LangChain, PyTorch-Geometric, TensorFlow, Scikit-learn, Pandas, Sentence-Transformers, Gensim
Tools & Viz	Streamlit, Plotly, Matplotlib
Expertise	NLP, Topic Modeling, Unsupervised Representation Learning, Contrastive Learning, VAEs, GNNs, Latent Space Geometry, Knowledge Distillation

PROFESSIONAL EXPERIENCE

Research Associate-I | *Indian Association for the Cultivation of Science* Jan 16, 2026 – Present

- **LLM Safety & Alignment:** Collaborating on research to develop safety guardrails and alignment methods for LLMs.

Ph.D. Scholar | *Indian Association for the Cultivation of Science* Oct 1, 2020 – Jan 15, 2026

- **Publication Venues:** AAAI (Demo), ACL, NAACL, IEEE-TAI, EACL, ECIR, LREC (workshop).
- Works included methodological innovations to end-to-end system building (detailed in Key Projects).

EDUCATION

Oct, 2020 – Jan, 2026 **Ph.D., Computer Science, IACS**, India Defended: 15/01/2026
Thesis Title: *Studies on Neural Topic Models for Document Corpora*

Aug, 2018 – Jul, 2020 **M.Sc., Mathematics & Computing, IACS**, India

2015 – 2018 **B.Sc. (Honours), Mathematics**, University of Calcutta, India

KEY PROJECTS

Generative AI Systems & Large-Scale Analysis

- **DTECT: Dynamic Topic Explorer & Context Tracker** [🔗 (AAAI '26 Demo)]
 - Built a unified system for tracking thematic evolution, proposing a custom temporal scoring metric to extract salient, trend-setting terms across large text corpora.
 - Engineered an interactive exploration interface powered by a LangChain-driven RAG pipeline, utilizing FAISS and MMR to ground an LLM chat assistant for automatic topic labeling and dynamic querying.
- **Scalable Analysis of Indian Parliamentary Question Hours Data** [🔗 (LREC '22)]
 - Processed 298k+ Lok Sabha Q&A pairs (1999–2019) using dynamic topic models to evaluate the temporal evolution of legislative discourse.
 - Quantified policy drift by mapping extracted topic shifts to socio-economic milestones, including the 2007 credit crisis and the post-2016 digital payments transition.

Unsupervised Representation Learning

- **S2WTM: Posterior-Collapse-Free Spherical Latent Space Modeling** [DOI] (ACL '25)
Leveraged Spherical Sliced-Wasserstein distance to model hyperspherical latent spaces, effectively mitigating posterior collapse. Outperformed Euclidean baselines with a 54.6% NPMI improvement and superior coherence validated by LLMs.
- **GINopic: Graph-Informed Text Modeling** [DOI] (NAACL '24)
Fused TF-IDF representations with graph-structured knowledge using Graph Isomorphism Networks (GIN). This dual-stream architecture captures global semantic correlations, achieving state-of-the-art coherence and classification accuracy across 5 benchmark datasets.
- **Contrastive Learning for Unsupervised Neural Topic Models** [DOI] (IEEE TAI '25)
Designed a model-agnostic Decoder Negative Sampling (VAE-DNS) framework compatible with 7 neural architectures. Implemented Triplet Loss to enforce latent space disentanglement, achieving superior cluster separation on Wiki40B (verified via UMAP).

Model Compression & Efficiency

- **Wasserstein Knowledge Distillation for Topic Models** [DOI] (ECIR '23)
Engineered a compression framework to distill a hybrid Teacher (BoW + SBERT) into a streamlined SBERT-only Student. Utilized Squared 2-Wasserstein distance to align latent geometries, achieving a model size reduction of 55.4% on 20NG while enabling the compressed student to surpass the teacher's topic coherence.
- **Regularization Dynamics in Generative Models** [DOI] (EACL '23)
Empirically demonstrated that high dropout rate disrupts representation learning in VAEs, establishing low-dropout protocols ($p < 0.2$) as essential for preserving generation quality in VAEs.

SELECTED PUBLICATIONS

For a complete list, see my [Google Scholar profile](#).

- **Suman Adhya** and Debarshi Kumar Sanyal. **DTECT: Dynamic Topic Explorer & Context Tracker.** *AAAI*, 2026 (Demo). [DOI] · [Code]
- **Suman Adhya** and Debarshi Kumar Sanyal. **S2WTM: Spherical Sliced-Wasserstein Autoencoder for Topic Modeling.** *ACL*, 2025. [DOI]
- **Suman Adhya**, Avishek Lahiri, Debarshi Kumar Sanyal, and Partha Pratim Das. **Evaluating Negative Sampling Approaches for Neural Topic Models.** *IEEE TAI*, 2025. [DOI]
- **Suman Adhya** and Debarshi Kumar Sanyal. **GINopic: Topic Modeling with Graph Isomorphism Network.** *NAACL*, 2024. [DOI]
- **Suman Adhya**, Avishek Lahiri, and Debarshi Kumar Sanyal. **Do Neural Topic Models Really Need Dropout? Analysis of Dropout in Topic Modeling.** *EACL*, 2023. [DOI]
- **Suman Adhya** and Debarshi Kumar Sanyal. **Improving Neural Topic Models with Wasserstein Knowledge Distillation.** *ECIR*, 2023. [DOI]

AWARDS & LEADERSHIP

- **Competitive Research & Travel Grants:** Awarded multiple grants from ACM India, ANRF, and Microsoft Research to present at top-tier conferences (AAAI '26, ACL '25, EACL '23).
- **INSPIRE-SHE Award (DST):** Nationally recognized for ranking in the top 1% in Class 12th Board Exams.
- **Institutional Leadership (IACS):** Served as Student Representative for the Anti-Ragging Cell, and as Research and Technical Assistant and System Manager for the Research Scholars Association.