SANUJ KUMAR

Ph.D. Candidate, Computer Science | Natural Language Processing, Text Mining & Trustworthy AI
New Mexico State University, Las Cruces, New Mexico
sanujkr@nmsu.edu | LinkedIn | GitHub | Google Scholar | Hugging Face | +1 575-993-9910

Research Profile

My primary research program is in natural language processing and **focused analysis**: designing models that surface subtle, user-relevant evidence and preserve information that dominant corpus patterns can obscure. I develop methods for **unsupervised multi-document summarization, targeted topic modeling, evidence-preserving representation/generation, contrastive learning, and diffusion-based text modeling**. A complementary applied research line studies **trustworthy and reliable AI** through human-in-the-loop auditing, calibrated deferral, model-disagreement analysis, retrieval-augmented reasoning, and multimodal/LLM reliability. **PHMCS is a major application of this broader trustworthy-AI agenda to longitudinal public-health communication.**

Research Interests

Focused Analysis & Text Mining • Unsupervised Multi-Document Summarization • Targeted Topic Modeling • Evidence-Preserving Representation Learning • Natural Language Processing • Contrastive Learning • Diffusion Models for Text • Trustworthy / Reliable AI • Human-in-the-Loop & Agentic AI • Natural Language Inference • Retrieval-Augmented Reasoning • LLM/SLM Reliability & Failure Analysis • Multimodal AI Evaluation • Efficient LLM Training

Education

Ph.D. Candidate, Computer Science — New Mexico State University, Las Cruces, NM *Aug 2021 – Present*

- Research concentration: Natural Language Processing / Text Mining; expected completion in 2026.
- Cumulative GPA: **3.97/4.00**.
- Dissertation research centers on *focused analysis*: steering models toward user-specific aspects of interest to reveal important signals that dominant themes can obscure.

M.S., Computer Science — New Mexico State University, Las Cruces, NM *Aug 2019 – Aug 2021*

- Cumulative GPA: **3.97/4.00**.

Academic & Research Experience

Research Assistant — Natural Language Processing / Text Mining, New Mexico State University *Jan 2020 – Present*

- Develop my primary doctoral research agenda in **focused analysis**, creating NLP methods that steer models toward user-specified aspects of interest and preserve subtle or minority evidence that can be lost when representations collapse toward dominant themes.
- Design focused and personalized generative models for **unsupervised abstractive multi-document summarization and targeted topic modeling** using Transformers, variational autoencoders, contrastive objectives, hierarchical/evidence-aware representations, and diffusion-based text modeling.
- As a secondary applied research line, lead NLP/AI development for the NSF-funded **Public Health Message Consistency System (PHMCS; Award #2346727)**, auditing longitudinal public-health communication across **104,670 agency messages** and **7.3 million candidate message pairs**; the project identified **6,212 confirmed perceived contradictions**.
- Developed PHMCS as a **trustworthy human-in-the-loop AI workflow** combining semantic retrieval, DeBERTa-v3 NLI, symbolic checks, selective LLM reasoning, rationale/disagreement signals, and human adjudication; the validated protocol improved contradiction-classification F1 from approximately **0.85 to 0.94**.
- Improve Transformer/LLM training efficiency using low-bit quantization, LoRA/instruction tuning, sparse computation, gradient checkpointing, computation-graph reuse, and memory-access optimization; collaborate across computer science, sociology, public health, and psychology on reliable model evaluation and human-AI workflows.

AI Research Fellow — Multimodal Reliability & Failure Analysis, Handshake AI *Jul 2025 – Present*

- Build evaluation suites and regression checks for multimodal reasoning and retrieval systems; perform systematic failure-mode analysis and optimize multi-stage prompting to improve reliability and user-facing response quality.

Instructor & AI Integration Reviewer — zyBooks *Jul 2025 – Present*

- Review and improve AI-enhanced computer-science learning content, including question variants, hints, grading logic, technical correctness, and instructional quality.

Lead AI/ML Trainer — Break Through Tech AI, Cornell University / Cornell Tech *May 2025 – Jul 2025*

- Delivered applied AI/ML instruction, workshops, and hands-on programming activities supporting **1,000+ learners**; coordinated an industry guest session with Keurig Dr Pepper and supported hackathon participation.

Graduate Assistant — Integration Developer, New Mexico State University *Feb 2020 – Jun 2024*

- Contributed to the NSF-funded **iCREDITS smart-grid project (Award #1914635)**, applying Spark-iForest for distributed multivariate anomaly detection and real-time decision support across multi-feeder grid systems; reduced detection latency by approximately **30%**.

Android Systems Engineer (Freelance / Open-Source) — XDA Developers Community *Jul 2017 – Nov 2019*

- Built and maintained AOSP-based custom ROMs for ARM devices; worked across Android framework, HAL, Linux kernel, device bring-up, debugging, Camera2 integration, power/wakelock behavior, Kconfig, and MIUI porting.

Teaching Experience

Co-Teacher — Psychology Research Courses & Emotion/Social Decision-Making Lab — New Mexico State University *Fall 2025 – Present*

- Served as **Co-Teacher for PSYC 4991 Research** in **Fall 2025 and Spring 2026** and as Co-Teacher for the **Ketelaar Emotion & Social Decision-Making Lab**, documenting interdisciplinary teaching and research-training experience in psychology.

Course Instructor — Java Programming (CS 452), New Mexico State University *Jun 2024 – Aug 2024*

- Developed course materials and guided students through hands-on Java programming using zyBooks-based activities and assessments.

Teaching Assistant — New Mexico State University *Feb 2020 – Jun 2024*

- Supported Object-Oriented Programming, Data Mining, Computer Organization & Assembly, Java, Python, and Database Management Systems; led labs, created recorded tutorials, and provided individualized student support.

Mentoring & Student Development

Project Assistant & Mentor — Research Experiences for Undergraduates (REU), NMSU *May 2021 & May 2022*

- Mentored a student through research-paper writing and publication; guided 10+ students on graph data, automation, surveys, and research workflows; delivered a workshop on remote research-server access.

Student Mentor — Computing Alliance of Hispanic-Serving Institutions (CAHSI), NMSU *Oct 2022 – Dec 2022*

- Mentored **20+ undergraduate students** in the CAHSI student group through structured discussions of Google-style programming/problem-solving questions, providing step-by-step guidance on algorithms, reasoning, debugging, and solution development.
- Helped students build confidence in technical problem solving by making solution strategies explicit and encouraging peer discussion rather than only supplying final answers.

Academic Event Coordination, Outreach & Community Engagement

Technical/Event Coordinator — Great Minds in STEM / CAHSI Data Analytics Challenge *Oct 2022*

- Helped create, package, test, and organize the **CodaLab competition bundle/environment** used for the Data Analytics Challenge and supported competition logistics for **20+ teams**.
- Assisted with technical setup and competition readiness, helping translate challenge materials into a reproducible submission/evaluation workflow for participating student teams.

Graduate Research & Arts Symposium (GRAS) — Graduate Student Council *2024–2025*

- As Vice President of Activities, helped coordinate university-level research-event logistics, programming, and participant support for approximately **90–100 graduate presenters/participants** from diverse disciplines.

Graduate Student Orientation — New Mexico State University *Fall 2024*

- Served as a graduate-student presenter/peer resource, sharing practical guidance for incoming graduate students and supporting orientation programming.

Research Computing Outreach — NMSU REU *May 2021*

- Designed and delivered an REU workshop on remote access and effective use of Computer Science research servers, helping undergraduate researchers become productive on shared computing infrastructure.

Peer-Reviewed Publications

---

1. Taylor, M. A., Stoltz, D. S., Harper, H., **Kumar, S.**, Nandhikonda, S. R., & Burks, L. (2026). *A Simulation-Based Slope Metric for Anchor List Reliability in Word Embedding Spaces.* **Sociological Methodology**.
2. Khubchandani, J., Batra, K., Wiblishauser, M., . . . , **Kumar, S.**, . . . , & Balls-Berry, J. E. (2026). *The Greatest Global Public*

Health Threats in the Post-COVID-19 Pandemic Era. **Journal of Medicine, Surgery and Public Health**.

3. **Kumar, S.**, & Le, T. M. V. (2024). *FoTo: Targeted Visual Topic Modeling for Focused Analysis of Short Texts*. Proceedings of LREC-COLING 2024, 7406–7416.
4. Ramirez, R. B., **Kumar, S.**, Le, T. M. V., & Cao, H. (2022). *Focused Stochastic Neighbor Embedding for Better Preserving Points of Interest*. IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT).
5. **Kumar, S.**, & Le, T. M. V. (2021). *A Word Embedding Topic Model for Robust Inference of Topics and Visualization*. Proceedings of the First International Conference on AI-ML Systems (AIMLSystems).

Preprints & Manuscripts in Progress

1. Stoltz, D. S., Taylor, M. A., & **Kumar, S.** (2026). *Selecting Language Models for Social Science: Start Small, Start Open, and Validate*. arXiv:2601.10926 [cs.CL, cs.AI].
2. **Kumar, S.** & Le, T. M. V. *Personalized Query-Driven Unsupervised Abstractive Summarization for Multi-Document Texts*. Dissertation research / manuscript in progress.
3. **Kumar, S.**, Nandhikonda, S. R., Harper, H., Khubchandani, J., Burks, L., Omidian, S., & Taylor, M. A. *Auditing Evolving Public Health Communication with Human-in-the-Loop AI: A Computational Framework for Detecting Perceived Inconsistency*. Manuscript in preparation.
4. Taylor, M. A., Harper, H., Khubchandani, J., **Kumar, S.**, Nandhikonda, S. R., Burks, L., & Omidian, S. "(Don't) Be Afraid": *Public Health Agencies, Contradiction Networks, and the Unintended Consequences of Crisis Communication on Social Media*. Manuscript in progress.

Conference Presentations, Posters & Research Dissemination

- **Kumar, S.**, et al. (2026). *Human-in-the-Loop Agentic Auditing of COVID-19 Public-Health Messaging: Detecting Contradictions, Updates, and Differences*. Mayo Clinic AI Research Summit 2026, Rochester, MN, Jun. 4–5, 2026. **Poster #89. Best Presentation & Poster Award.**
- Taylor, M. A., et al. (2026). "Be Afraid. Don't Be Afraid": *Public Health Agencies, Contradiction Networks, and the Unintended Consequences of Crisis Communication on Social Media*. NSF SaTC PI Meeting, College Park, MD, Aug. 6–7, 2026. Coauthor/presenter.
- Taylor, M. A., et al. (2026). Same work presented at the **American Sociological Association Annual Meeting**, New York, NY, Aug. 7–11, 2026. Coauthor/presenter.
- **Kumar, S.**, Ketelaar, T., & Nevarez, J. (2026). *Who Joins a Twitter Mob?: Inter-Elite Competition, Moral Outrage, and the Dynamics of Digital Collective Action*. Rocky Mountain Psychological Association (RMPA), Boise, ID, Mar. 26–28, 2026. Co-presenter.
- **Kumar, S.** & Le, T. M. V. (2025). *FoTo: Targeted Visual Topic Modeling for Focused Analysis of Short Texts*. Center for Decoding the Universe Quarterly Forum, Stanford Data Science, Stanford University, Stanford, CA, Oct. 29, 2025. Poster presentation.
- Dhanapal, K., & **Kumar, S.** (2025). *Analyzing Bikeshare User Preferences Across Urban Zones: A Case Study of San Francisco's Zoning Influence on Ridership*. Northeast Decision Sciences Institute Annual Conference, Hershey, PA, Mar. 27–29, 2025.
- Taylor, M. A., Stoltz, D. S., Harper, H., **Kumar, S.**, Nandhikonda, S. R., Khubchandani, J., & Burks, L. (2025). *Semantic Moorings / A Simulation-Based Slope Metric for Semantic Relation Reliability in Word Embedding Spaces*. International Roundtable on Computational Social Science, Norrköping, Sweden. Coauthor.
- **Kumar, S.** & Le, T. M. V. (2024). *FoTo: Targeted Visual Topic Modeling for Focused Analysis of Short Texts*. LREC-COLING 2024, Torino, Italy, May 20–25, 2024.
- **Kumar, S.** (2024). Graduate Student Orientation, New Mexico State University. Presenter, Fall 2024.
- **Kumar, S.** & Le, T. M. V. (2021). *A Word Embedding Topic Model for Robust Inference of Topics and Visualization*. AIMLSystems 2021.
- **Kumar, S.** (2021). *REU Workshop: Usage of Computer Science Research Servers*. New Mexico State University.
- **Kumar, S.** (2020). *New Generation's Approaches, Technologies & Career Opportunities*. Webinar/presentation.

Selected Conference Participation & Research Travel

- **Mayo Clinic AI Research Summit 2026** — Rochester, Minnesota, Jun. 4–5, 2026. First-author/co-presenter of PHMCS human-in-the-loop AI work; Poster #89; **Best Presentation & Poster Award**.
- **NSF Security, Privacy, and Trust in Cyberspace (SaTC) PI Meeting** — College Park, Maryland, Aug. 6–7, 2026. Represented the NSF-funded public-health messaging collaboration through research dissemination, project discussion, and cross-disciplinary networking.
- **American Sociological Association 121st Annual Meeting** — New York City, Aug. 7–11, 2026. Coauthor/presenter on public-health contradiction-network research.
- **Rocky Mountain Psychological Association 96th Annual Convention** — Boise, Idaho, Mar. 26–28, 2026. Co-presenter

on digital collective-action / social-media research.
- **Stanford Data Science, Center for Decoding the Universe Quarterly Forum** — Stanford, California, Oct. 29, 2025. Poster presenter; participated in cross-disciplinary discussions on foundation models and agentic AI for scientific discovery.
- **Northeast Decision Sciences Institute Annual Conference** — Hershey, Pennsylvania, Mar. 27–29, 2025. Coauthor on urban bikeshare / zoning analytics research.
- **LREC-COLING 2024** — Torino, Italy, May 20–25, 2024. Author of peer-reviewed FoTo paper in computational linguistics / NLP.

Awards, Fellowships & Scholarships

- **Best Presentation & Poster Award** — Mayo Clinic AI Research Summit 2026, Rochester, MN (first author/co-presenter), Jun 2026.
- **TAYLOR-NSF-PHMC GRAD Fellowship / graduate research support** — multi-year support through the NSF Public Health Messaging project (2021–2027; NSF Award #2346727, project total $620,095).
- **GRAD SUCCESS Scholarship** — $5,342, New Mexico State University, 2024–2025 academic year.
- **GRAD SUCCESS Scholarship** — $1,660, New Mexico State University, 2021–2022 academic year.
- **NMSU C/S EPSCoR Scholarship** — $3,800 for data-intensive computing research, Feb 2021.

Selected Research & Technical Projects

Focused / Query-Driven Unsupervised Multi-Document Summarization — Doctoral Research *2021 – Present*
- **Primary doctoral research** developing representation-learning and generation methods that preserve document- and review-specific evidence while producing focused abstractive summaries; explores Transformer, VAE, contrastive, hierarchical attention, soft/discrete selection, and diffusion-based designs.

FoTo — Targeted Visual Topic Modeling for Focused Analysis of Short Texts *2021 – 2024*
- Foundational focused-analysis work developing targeted topic-modeling and visualization methods that reveal structures aligned with a user-specified focus rather than only dominant corpus themes; published at **LREC-COLING 2024**.

Public Health Message Consistency System (PHMCS) — NSF Award #2346727 *2021 – Present*
- **Secondary applied research line and trustworthy-AI case study**: human-in-the-loop framework for detecting and adjudicating contradictions, legitimate updates, and contextual differences in longitudinal federal/state public-health messaging; integrates retrieval, NLI, symbolic reasoning, selective LLM inference, model disagreement, rationales, calibrated escalation, and human review.

IntentGuard — Continuous Intention/Behavior Verification for Autonomous Agents *2026*
- Prototype safety layer for autonomous agents that continuously checks task drift, policy violations, and anomalous behavior and can intervene or generate structured reports; designed as a containerized Python evaluation component.

Care4U — Agentic Assistant Prototype *2026*
- Agentic assistant prototype exploring multi-step planning and retrieval-grounded reasoning.

University Website Content Analysis Project (UWCAP) — Interdisciplinary NMSU Collaboration *2025 – Present*
- Develops reproducible workflows for archived university website collection and content analysis, including image/text extraction, Wayback Machine quality logging, structured longitudinal datasets, and research-assistant coordination across law and medical school web archives.

iCREDITS Smart-Grid Integration — NSF Award #1914635 *2020 – 2024*
- Applied distributed Spark-iForest anomaly detection and visualization for smart-grid decision support; improved reliability and reduced detection latency by approximately 30%.

Leadership & University Service

Secondary Advisor, Indian Student Association — New Mexico State University *2025 – Present*
- Advise student leadership, support new-student onboarding, and help strengthen cultural programming and community representation; continue serving as an experienced point of contact for incoming international students.

Vice President of Activities, Graduate Student Council — New Mexico State University *Aug 2024 – Jan 2025*
- Coordinated university-level graduate programming and research-event logistics, including the Graduate Research & Arts Symposium with approximately 90–100 participants.

President, Indian Student Association — New Mexico State University *Sep 2021 – Sep 2022*
- Led a 50+ member organization; organized and supported cultural/community programs including potlucks, Holi, Cultural Bazaar, and the flagship *Roshni* event, serving approximately **350–400+ attendees** across three consecutive years; supported

onboarding for **30+ incoming students** and participated in community-service initiatives including the Big Event and Keep State Great.

President, Computer Science Graduate Student Organization — New Mexico State University *Sep 2021 – Sep 2022*

- Organized orientations, faculty/student interview sessions, coffee hours, and activities supporting graduate-student community building.

Certifications & Professional Development

- LangGraph — Develop LLM-Powered AI Agents with LangGraph, Oct 2024.
- LangChain — Develop LLM-Powered Applications with LangChain, Apr 2024.
- edX Verified Certificate — Large Language Models: Application through Production, Apr 2024.
- Advanced Data Science Specialization, Jul 2020.

Technical Skills

- **Programming & Data:** Python, Java, C++, SQL, JavaScript, PHP, React, Node.js, Django, HTML/CSS, Apache Spark / PySpark / MLlib, HiveQL, MongoDB, relational databases.
- **NLP / Machine Learning:** PyTorch, TensorFlow, Hugging Face Transformers, DeBERTa, BERT/GPT-family models, NLI, topic modeling, abstractive summarization, contrastive learning, VAEs, diffusion models, classical ML, clustering, anomaly detection.
- **LLMs / Agents / Retrieval:** DeepSeek-R1, LLaMA/Ollama, GPT-OSS, LangChain, LangGraph, RAG, n8n, multi-stage prompting, committee/multi-agent workflows, few-shot prompting, Chain-of-Thought / Tree-of-Thought methods.
- **Efficient AI / Systems:** LoRA, instruction tuning, Unsloth, quantization, gradient checkpointing, sparse computation, computation-graph reuse, GPU-memory/performance optimization, Docker, REST APIs.
- **Cloud / Engineering Tools:** AWS, Azure AI, Git, JIRA, Tableau, Pinecone, Cascade CMS.

Last updated: September 2026