Avan Kumar

CONTACT INFORMATION
1019 East University Dr.
Tempe, AZ 85281 USA
Email: akuma469@asu.edu
Connect on: LinkedIn

RESEARCH INTERESTS
Sustainable Process Development, Process and Product Design, Literature-Mining, Circular Economy, Life Cycle Assessment, Generative AI, Large Language Models, Natural Language Processing, Explainable AI, Machine Learning, Deep Learning, and Multi-Objective Optimization.

EDUCATION

Indian Institute of Technology, Delhi (IIT Delhi), Hauz Khas, Delhi — Aug '19 - May '24
Ph.D., Chemical Engineering

Indian Institute of Technology, BHU (IIT BHU), Varanasi, Uttar Pradesh — Aug '17 - May '19
M.Tech., Chemical Engineering

Delhi Technological University (DTU), Rohini, Delhi — Aug '13 - May '17
B.Tech., Polymer Science and Chemical Technology

RESEARCH EXPERIENCE

Post-Doctorate Researcher, *Advisor: Prof. Bhavik R. Bakshi* — Dec '23 - Present
At the School of Sustainability, Arizona State University, Tempe, AZ, USA
Research Area:

- Developing a Customized Large Language Model for Creating a Structured Database from Unstructured Text Resources to Develop Life Cycle Inventories, named "Sustain-GPT" for Chemical and Material Industry (CMI).
- Extraction of Reaction Information from Scientific Literature to Build a Circular Reaction Network Using a Tailored Large Language Model.

PhD Degree, *Advisor: Prof. Hariprasad Kodamana* — Jul '19 - Dec '23
At the Chemical Engineering Department, IIT Delhi, Hauz Khas, Delhi, India
Thesis Title- *"Application of Deep Learning and Natural Language Processing for Sustainable Process Development"*

- **Photo-catalyst Classification and Band Gap Prediction Using Deep Learning Models:** Developed an integrated machine learning framework to enhance photo-catalytic reactions for solar energy utilization, facilitating methanol production to support sustainable fuel alternatives.
- **Data-driven Explainable Machine Learning and Optimization Framework:** Gaussian Process Regression (GPR) with Shapley tool and Multi-objective Bayesian Optimization (MOBO) were implemented to model and identify the optimal input features for maximizing methanol yield from carbon monoxide, enabling efficient and sustainable chemical processes.
- **Customized (Large) Language Models for Knowledge Extraction Across Various Fields:** This includes (1) the "Extend-SciBERT" and (2) "H_2-BERT" models for hydrogen production, (3) "Recycle-BERT" for plastic waste recycling, and (4) "CCU-LlaMA" for carbon capture and utilization (CCU).
- **A Study on the Evolutionary Nature of PET Plastic Waste Recycling Technologies:** An NLP-based framework that streamlines the identification of effective PET plastic recycling methods, promoting circular economy practices.
- **Implementation of Machine Learning and Optimization Tools on the Biomass Gasification Process:** A predictive GPR model was trained on experimental data to estimate syngas yield. The weighted multi-objective Bayesian optimization was implemented to achieve optimal solutions and was further validated through experiments.

M.Tech. Degree, *Advisor: Prof. Ravi P. Jaiswal* — Jul '17 - Dec '19
At the Chemical Engineering Department, IIT BHU, Varanasi, Uttar Pradesh, India
Thesis Title- *"Luminescent Downshifting Natural Dyes to Enhance Light Harvesting Efficiency of mc-Silicon Solar Module"*

- **Modification of Multi-Crystalline Silicon Solar Module:** Incorporating natural luminescent dyes into solar modules enhances efficiency while providing an eco-friendly and sustainable solar energy system.

Journal Publications

[10] "Development of a Circular Reaction Network Using Scientific Literature and a Customized Large Language Model for Methanol Chemical" *(Under Preparation)*

[9] "Developing a Customized Large Language Model to extract the Life Cycle Inventory for Methanol and Plastic Packaging Waste End-of-life." *(Under Preparation)*

[8] **Kumar, Avan**, Bhavik R. Bakshi, Manojkumar Ramteke, and Hariprasad Kodamana. "An Evolutionary Study on Technologies for Polyethylene Terephthalate Waste Recycling Using Natural Language Processing." *(Under Review)*

[7] Kundu, Kaushik, **Avan Kumar**, Hariprasad Kodamana, and Kamal K. Pant. "Obtaining high H2-rich syngas yield and carbon conversion efficiency from biomass gasification: From characterization to process optimization using machine learning with experimental validation." *Fuel* 378 (2024): 132931.

[6] Jami, Harshitha Chandra, Pushp Raj Singh, **Avan Kumar**, Bhavik R. Bakshi, Manojkumar Ramteke, and Hariprasad Kodamana. "CCU-Llama: A Knowledge Extraction LLM for Carbon Capture and Utilization by Mining Scientific Literature Data." *Industrial & Engineering Chemistry Research* (2024).

[5] **Kumar, Avan**, Sreedevi Upadhyayula, and Hariprasad Kodamana. "A Convolutional Neural Network-based gradient boosting framework for prediction of the band gap of photo-active catalysts." *Digital Chemical Engineering* 8 (2023): 100109. doi:10.1016/j.dche.2023.100109

[4] **Kumar, Avan**, Bhavik R. Bakshi, Manojkumar Ramteke, and Hariprasad Kodamana. "Recycle-BERT: extracting knowledge about plastic waste recycling by natural language processing." *ACS Sustainable Chemistry & Engineering* 11, no. 32 (2023): 12123-12134. doi:10.1021/acssuschemeng.3c03162

[3] **Kumar, Avan**, Kamal K. Pant, Sreedevi Upadhyayula, and Hariprasad Kodamana. "Multiobjective Bayesian optimization framework for the synthesis of methanol from syngas using interpretable Gaussian process models." *ACS Omega* 8, no. 1 (2022): 410-421. doi:10.1021/acsomega.2c04919

[2] **Kumar, Avan**, Swathi Ganesh, Divyanshi Gupta, and Hariprasad Kodamana. "A text mining framework for screening catalysts and critical process parameters from scientific literature-A study on Hydrogen production from alcohol." *Chemical Engineering Research and Design* 184 (2022): 90-102. doi:10.1016/j.cherd.2022.05.018

[1] Singh, Juhi, **Avan Kumar**, Anubha Jaiswal, Swati Suman, and Ravi P. Jaiswal. "Luminescent down-shifting natural dyes to enhance photovoltaic efficiency of multicrystalline silicon solar module." *Solar Energy* 206 (2020): 353-364. doi: 10.1016/j.solener.2020.05.067

Book Chapters

[2] **Kumar, Avan**, and Hariprasad Kodamana. "Process Modeling and Optimal Evaluation Analysis for Direct CO2 Conversion to Methanol." (2024). doi:10.1016/B978-0-443-15740-0.00113-0

[1] **Kumar, Avan**, and Hariprasad Kodamana. "An NLP-based framework for extracting the catalysts involved in Hydrogen production from scientific literature." In *Computer Aided Chemical Engineering*, vol. 52, pp. 1457-1462. Elsevier, 2023. doi: 10.1016/B978-0-443-15274-0.50232-8

International & National Conferences

[6] **Kumar, Avan**, Amrita Sen, and Bhavik R. Bakshi, "Sustain-Gpt: A Large Language Model for Creating a Structured Database from Unstructured Text Resources to Develop Life Cycle Inventories," 2024 AIChE Annual Meeting at San Diego on October 27^{th}–31^{st} 2024.

[5] **Kumar, Avan**, Frashid Nazemi, and Bhavik R. Bakshi, "Developing a Life Cycle Inventory Database for Plastic Packaging End of Life Technologies Using Customized Large Language Models.", ACLCA 2024 at Snowbird, Utah, on September 24^{th}–26^{th}, 2024.

[4] **Kumar, Avan** and Hariprasad Kodamana, "A Neural Network Framework for material selection based on band gap using Chemical Embedding", 14th European Congress of Chemical Engineering and 7th European Congress of Applied Biotechnology at the CityCube in Berlin/Germany on September 17^{th}–21^{st} 2023.

[3] **Kumar, Avan**, and Hariprasad Kodamana, "An NLP-based framework for extracting the catalysts involved in hydrogen production from scientific literature", 33rd European Symposium on Computer-Aided Process Engineering (ESCAPE-33) at Athens, Greece, June 18^{th}–21^{nd} 2023.

[2] Kundu, Kaushik, **Avan Kumar**, K.K. Pant, Hariprasad Kodamana, and R.R. Sonde, "Prediction of Syngas Yield for Biomass Gasification using Multivariate LSTM", IIChE - CHEMCON at Kolkata on December 27^{th} - 30^{th}, 2023.

[1] **Kumar, Avan**, Sreedevi Upadhyayula, and Hariprasad Kodamana, "A deep learning framework for prediction of the band gap photoactive catalyst", 10th Asian Symposium on Process Systems Engineering: Systems Engineering for the Digitalization Era at IIT Madras on 11^{th}–14^{th} December 2022.

Invited Talks

Invitation for *International Industrial Ecology Day 2024* by Prof. Qingshi Tu for the title "Sustain-Gpt: A Large Language Model for Creating a Structured Database from Unstructured Text Resources to Develop Life Cycle Inventories" on Nov 21^{st}, 2024

Industrial Projects

[2] AI/ML-based Modelling, Optimization and Variability Identification of Trastuzumab Upstream Cell Culture Process Jul '23- Jul '24

[1] Predicting the Performance of a Driver in a Racing Simulator for Racing Unleashed AG Jul '22- Aug '23

Journal Reviewer

[1] Environmental Science & Technology

[2] Digital Discovery

Grant Proposals

Environmental Protection Agency - role played in proposal

Amazon - wrote this proposal

Teaching Experiences

Chemical Engineering at IIT Delhi Aug '20 - May '23
Teaching Assistant: Process Data Analytics, Statistical Methods & Chemical Engineering, and Process Control lab

Chemical Engineering at IIT BHU Aug '18 - May '19
Teaching Assistant: Process Dynamics & Control, Fluid Dynamic lab, Chemical Reaction Engineering (CRE) lab, Heat Transfer lab

Certificate Course in Data Science and Machine Learning, CEP IIT Delhi Mar '21 - Nov '23
Teaching Assistant: Generate Code and give Tutorial on Machine/Deep Learning Models

Awards & Recognition

Researcher Scholar Travel Award (RSTA) at IIT Delhi May '23

Research Excellence Travel Award (RETA) at IIT Delhi Aug '23

Nucleus OCS PhD coordinator
At Chemical Engineering, IIT Delhi Aug '22 - May '23

Head of Polymer Clay Event
In Tatva, Technical Fest at DTU Feb '16

Class Representative (CR)
At Chemical Engineering IIT BHU Aug '17 Jun '19

Research Mentorship

Kaushik Kundu, Ph.D. student, 2022-present

Shubham Garampalli and Yamini Vijay Khajekar, Dual Degree students, 2023-2024

Harshitha Chandra Jami and Pushp Raj Singh, Dual Degree students, 2022-2023

Khushee Namdeo and Vishruta Kathuria, B.Tech. students, 2021-2022

Swathi Ganesh (IIT Madras) and Divyanshi Gupta (IIT Delhi), B.Tech. students, 2020-2021

Industrial Internships

[2] Delhi Jal Board, Delhi, India Jun '16- Jul '16

[1] Tirupati Structural Limited, Sahibabad, UP, India Dec '15- Jan '16

Professional Skills

Operating System (OS): Ubuntu, Window
Machine Language: Python, SQL, LaTex
Python Tools: NLTK, Genism, SpaCy, NumPy, Pandas, BS4, Matplotlib, Seaborn, etc.
Technical Frameworks: Pytorch, TensorFlow, Scikit-Learn, LaTex
Technical Expertise: Large Language Models, Natural Language Processing Tools, Machine Learning, Deep Learning, Explainable AI, Generative AI

Professional Certificates	[1] Programming for Everybody (Getting Started with Python) [2] Understanding and Visualizing Data with Python [3] Natural Language Processing with Classification and Vector Spaces [4] Transfer Learning for NLP with TensorFlow Hub [5] Fine Tune BERT for Text Classification with TensorFlow [6] What is Data Science? [7] Deep Learning with PyTorch: Image Segmentation
Community Service	**American Center for Life Cycle Assessment (ACLCA), 2023** - role played as a Volunteer