Avan Kumar, PhD

Contact Information

1019 East University Dr.
Tempe, AZ 85281 USA

Email: akuma469@asu.edu
Connect on: LinkedIn

Research Interests

Generative AI (including Large Language Models and NLP), Machine Learning (with Deep Learning and Explainable Tools), Multi-Objective Optimization, and Sustainable Process & Product Design (including Life Cycle Assessment and Sustainable Systems Development), AI-Driven Circular System Discovery, and Life Cycle Assessment

Education

Arizona State University, Tempe, USA — Dec '23 - Present
Postdoctoral Researcher, School of Sustainability (AI for Sustainable Systems)

Indian Institute of Technology, Delhi (IIT Delhi), Hauz Khas, Delhi — July '19 - May '24
Ph.D., Chemical Engineering (Deep Learning Modelling and Text-Mining for Sustainable Systems)

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

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

Work Experience

[3] A Comprehensive Life Cycle Assessment of Sorbent-Based Fluidized Direct Air Capture (DAC) Technology. ***A project was jointly carried out with DACLab Ltd.*** — Aug '25 - Present
Developing an iterative life cycle assessment (LCA) framework for the proposed DAC technology that incorporates distance- and energy-sensitivity analyses, evaluates the potential use of waste heat, and identifies effective and sustainable pathways from the carbon capture stage through utilization stages.

[2] AI/ML-based Modelling, Optimization and Variability Identification of Trastuzumab Upstream Cell Culture Process. ***A project was jointly carried out with Biocon Ltd.*** — Aug '23 - May '24

- Developed a predictive model to assess the impact of potential variations in process parameters on the product and critical quality attributes.
- Optimized the process to intensify product output beyond the average of the current process while also impacting product quality attributes.
- Analyzed the impact of all monitored process parameters on the product and its associated critical quality attributes.

[1] Predicting the Performance of a Driver in a Racing Simulator for Racing Unleashed AG.
A project was carried out jointly with Racing Unleashed. — Jul '22- Aug '23

- Developed a framework for identifying the top drivers across a completed race using statistical approaches.
- A convolutional neural network (CNN)-based approach was developed to separate the top drivers from the rest based on various attributes. The approach achieved 98% accuracy in its predictions.
- An autoencoder was used to calculate the reconstruction error across various deciles and to predict a driver's performance during a race.
- Developed an ensemble method to predict the outcome for a driver during a race. The framework achieved an efficacy of 95%.

Research Experience

Post-Doctorate Researcher, *Advisor: Prof. Bhavik R. Bakshi* — Dec '23 - Present
At the School of Sustainability, Arizona State University, Tempe, AZ, USA
Ongoing AI-enabled Projects for Achieving Net-Zero in the Chemical and Materials Industry:

- Developing Life Cycle Inventory and Environmental Impact Databases Using the Customized Sustain-LLaMA Framework
- An Automated LLM-Driven Framework for Discovering Sustainable Circular Pathways for Chemicals and Polymers
- Generative AI Models and Frameworks for Sustainable-by-Design Materials
- Circular and Sustainable Pathways for the Life Cycle of Lithium Batteries

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 convolutional feature extractor with a machine learning-based model to provide preliminary photoactive catalysts that enhance photocatalytic reactions for solar energy utilization, facilitating chemical production to support a sustainable environment.
- **Data-driven Explainable Machine Learning and Optimization Framework:** Gaussian Process Regressor (GPR) with an explainable tool (Shapley) 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 End-of-Life (EoL) Technologies:** An NLP-based framework that streamlines the identification of effective PET plastic EoL technologies, promoting plastic circular economy practices.

M.Tech. Degree, *Advisor: Prof. Ravi P. Jaiswal* Jul '17 - May '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:** By incorporating natural luminescent dyes into the ethylene vinyl acetate (EVA) layer of solar modules, efficiency was enhanced by $\approx 9\%(\pm 4)$.

Journal Publications

[11] Kundu, Kaushik, **Kumar, Avan**, Hariprasad Kodamana, and Kamal K. Pant. "Data-driven optimization of catalyst composition and reaction temperature for hydrogen-rich syngas yield from biomass gasification: Experimental validation and characterization." Energy Conversion and Management 354 (2026): 121243.

[10] **Kumar, Avan**, Sunghoon Kim, and Bhavik R. Bakshi. "Role of artificial intelligence in the chemical industry transition to a sustainable, circular, and net-zero future." Current Opinion in Chemical Engineering 2026 51, 42, 101234.

[9] **Kumar, Avan**, Farshid Nazemi, Manojkumar Ramteke, Hariprasad Kodamana, and Bhavik R. Bakshi. "A Large Language Model-based Framework to Retrieve Life Cycle Inventory and Environmental Impact Data from Scientific Literature" Environmental Science & Technology 2025 59, 42, 22533–22543.

[8] **Kumar, Avan**, Harshitha Chandra Jami, Bhavik R. Bakshi, Manojkumar Ramteke, and Hariprasad Kodamana. "An evolutionary study on technologies for polyethylene terephthalate waste recycling using natural language processing." Computers & Chemical Engineering (2025): 109011.

[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* 63, no. 41 (2024): 17585-17598.

[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.

[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.

[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.

[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.

[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.

Under Review/ Preparation Papers

[4] "A Multi-Objective Optimization Framework for Screening Sustainable Downstream Supply Chains through Industrial Symbiosis of Agricultural Residues." *(Under Revision in the industrial & engineering chemistry research)*

[3] "A Knowledge Graph-Embedded Large Language Model to Strategize Battery Recycling Technology Selection" *(Under Revision in Computer & Chemical Engineering)*

[2] "Development of a Circular Reaction Network Using Scientific Literature and a Customized Large Language Model for Methanol Chemical" *(Under Review in Digital Discovery)*

[1] "METHANE-LLAMA: An Agentic Domain-Adapted Large Language Model Framework for Automated Methane Abatement Decision Support" *(Under Preparation)*

Book Chapters

[2] **Kumar, Avan**, and Hariprasad Kodamana. "Process Modeling and Optimal Evaluation Analysis for Direct CO2 Conversion to Methanol." In Comprehensive Methanol Science, pp. 190-210, Elsevier, 2025. 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**, Sunghoon Kim, George Stephanopoulos, and Bhavik R. Bakshi, "A Customized Large Language Model for Creating Circular Reaction Networks for Discovering Sustainable Chemical Processes," 2025 AIChE Annual Meeting at Boston, MA on November 2^{nd}–6^{th} 2025.

[5] **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.

[4] **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.

[3] **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.

[2] **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.

[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 December 11^{th}–14^{th} 2022.

Invited Talks

[2] A collaborative talk with Prof. Bhavik R. Bakshi for Amazon's Sustainability Speaker Series on "Sustainable Transformation of the Chemicals and Materials Industry to Net-Zero Emissions" on January 29^{th}, 2025.

[1] 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 November 21^{st}, 2024.

Journal Reviewer

[1] Environmental Science & Technology, [2] Digital Discovery, [3] Applied Engineering, and [4] Sustainable Energy and Fuels

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

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

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

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

Class Representative (CR)
At Chemical Engineering IIT BHU — Jul '17 May '19

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

Research Mentorship

Bilal Rasheed and RASHED AHMED (IIT Delhi Abu Dhabi) students, 2024–present

Shobhitha Sharma and Robin Rathore Ph.D. student, 2024–present

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