HARIKRISHNA SAHU

RESEARCH SCIENTIST, AI FOR SCIENCE (PHD)

514 Altama Way, Acworth, 30102, United States | 470-309-7235 |
ORCID: 0000-0001-5458-9488 | Scopus: 7003538570 |
harikrishnasahu89@gmail.com | https://www.linkedin.com/in/harikrishna-chem/

PROFILE

Innovative computational scientist specializing in materials, molecules, and polymers design, including organic semiconductors and OLED-relevant compounds, using DFT, molecular dynamics, and AI/ML. Creator of polyT5, a polymer-specific large language model for property prediction and inverse design, and PSP, a toolkit for hierarchical polymer modeling. Proven track record delivering impactful solutions in academia and industry, including collaboration with Toyota Research Institute on polymer adaptations of CAMD. Driven to accelerate materials innovation by combining deep expertise with advanced computational methods.

TECHNICAL SKILLS

Machine Learning for Chemistry: Property prediction & generative design using tree-based models (Random Forest, XGBoost), fingerprint-based methods (Morgan, MACCS, ECFP), multi-layer perceptrons (MLPs), graph neural networks (GNNs).

Computational Chemistry: DFT (VASP, ORCA, Gaussian), molecular dynamics (LAMMPS).

Programming & Tools: Python (PyTorch, RDKit, Scikit-learn).

Domain Expertise: OLED/organic electronics, organic semiconductors, polymer informatics, dielectric, thermal & optoelectronic property prediction.

Advanced AI: Large language models (LLMs) for inverse design, NLP/LLM for literature data mining, experience with Hugging Face Transformers and PydanticAI; open to adopting LangChain, AutoGen, and related agentic frameworks.

PROFESSIONAL EXPERIENCE

Research Scientist II, Georgia Institute of Technology, Atlanta **May 2023 — Present**
Leading the development of computational and machine learning tools for accelerating materials discovery. Developed and deployed polyT5, a large language model tailored for polymer chemistry, alongside other AI-driven pipelines for property prediction, literature data extraction, and materials screening — identifying high-performance organic and inorganic compounds. Applied physics-informed multi-task ML to fuse experimental and simulation data for predicting solvent diffusivity in polymers. Collaborating with multidisciplinary teams and mentoring PhD students to bridge computational predictions with experimental validation.

Postdoc, Georgia Institute of Technology, Atlanta **Sep 2019 — May 2023**
Created the Polymer Structure Predictor (PSP) toolkit and integrated it into the CAMD platform, enabling autonomous, target-driven materials discovery workflows. Developed machine learning models for predicting electrical conductivity in *p*-doped polymers, screening over 838,000 candidates and identifying high-performance materials, alongside other ML/DFT-driven investigations.

Postdoc, Nanjing University, Nanjing **Mar 2017 — Jun 2019**
Designed organic materials for solar cell applications using machine learning, developing predictive models to screen over 10,000 candidates.

EDUCATION

Ph.D. in Chemistry **Jul 2011 — Oct 2016**
IIT Guwahati, Guwahati, India
Thesis Title: *In-silico investigation of optical and electronic properties of heterocyclic conjugated polymers*

MSc in Chemistry **Jul 2009 — May 2011**
NIT Rourkela, Rourkela, India

AWARDS

KRICT ChemDX Hackathon: Colaboratory, South Korea **Nov 2024**
Awarded first prize for the generative design of inorganic thermoelectric materials.

AMDD hackathon, Remote **May 2020**
Won first prize by demonstrating that numerous low-cost experiments can yield greater value than a few expensive ones.

SELECTED PUBCLICATIONS

1. An Encoder-Decoder Foundation Chemical Language Model for Generative Polymer Design, *arXiv*, 2510.18860, **2025.**
2. Machine Learning for Green Solvents: Assessment, Selection and Substitution, *Adv. Sci.*, e16851, **2025**.
3. From Corpus to Innovation: Advancing Organic Solar Cell Design with Large Language Models, *Submitted*, 2025.
4. polyBART: A Chemical Linguist for Polymer Property Prediction and Generative Design, In *Findings of the Association for Computational Linguistics: EMNLP*, 12104–12119, **2025**.
5. Solvent diffusivity in polymers: Simulation-experiment data fusion and multi-task machine learning, *Npj Comput. Mater.*, 11, 187, **2025**.
6. Stereoisomers of vicinal polydichloronorbornene for ultra-high-temperature capacitive energy storage, *Adv. Mater.*, 2417625, **2025**.
7. Elucidating Photochemical Conversion Mechanism of PDMS to Silica under Deep UV Light and Ozone, *J. Phys. Chem. Lett.*, 16, 747, **2025**.
8. A machine learning approach to predicting the spall strength of metals and alloys, *J. Appl. Phys.*, 137, 104905, **2025**.
9. High-temperature high-k polyolefin by rational molecular design, *Proc. Natl. Acad. Sci.*, 121, e2415388121, **2024**.
10. Low temperature 3D printing of transparent silica glass microstructures*, Sci. Adv.*, 9, eadi2958, **2023**.
11. Polymer Structure Predictor (PSP): A Python Toolkit for Predicting Atomic-Level Structural Models for a Range of Polymer Geometries, *J. Chem. Theory Comput.*, 18, 2737, **2022**.
12. polyG2G: A Novel Machine Learning Algorithm Applied to the Generative Design of Polymer Dielectrics, *Chem. Mater*., 33, 7008, **2021**.
13. An Informatics Approach for Designing Conducting Polymers, *ACS Appl. Mater. Interfaces*, 13, 53314, **2021**.
14. Designing Promising Molecules for Organic Solar Cells via Machine Learning Assisted Virtual Screening, J. Mater. Chem. A, 7, 17480, 2019.
15. Toward Predicting Efficiency of Organic Solar Cells via Machine Learning and Improved Descriptors, *Adv. Energy Mater.*, 8, 1801032, **2018**.
16. Computational Investigation of Charge Injection and Transport Properties of a Series of Thiophene-Pyrrole-based Oligo-Azomethines, *Phys. Chem. Chem. Phys.*, 16, 8563, **2014**.

LEADERSHIP & MANAGEMENT

1. Successfully led and coordinated research projects involving interdisciplinary teams, ensuring timely progress and collaboration.
2. Managed computational resources for the research group, including writing and securing large-scale computing proposals (e.g., 6 million ACCESS credits, 3 million XSEDE core-hours).
3. Maintained and updated critical research databases (Khazana) and implemented data backup protocols to safeguard research outputs.
4. Mentored graduate students on advanced computational modeling and machine learning techniques, supporting their professional growth and project success.

REFERENCES

Rampi Ramprasad at Georgia Institute of Technology rampi.ramprasad@mse.gatech.edu
Aditya N. Panda at Indian Institute of Technology Guwahati anp@iitg.ac.in
Huan Tran at Georgia Institute of Technology huan.tran@mse.gatech.edu