



AI-Driven Discovery of Crystal Structures for Optics and AI Assisted UED Alignment



Christopher Chow¹, Tianyi Zhao¹, Anita Nam¹, Viviana Vladutescu¹, Mikhail Polyanskiy², William Li², Li Geng¹, Vikas Teotia², Lufeng Leng¹, Marcus Babzien², Giovanni Ossola¹, Aldane Duncan¹, Axel Lojas¹, Henry Ye¹, Raul-Barajas Vargas¹, Efrain Luna¹, Avanti Williston¹, Sheikh Tusher¹

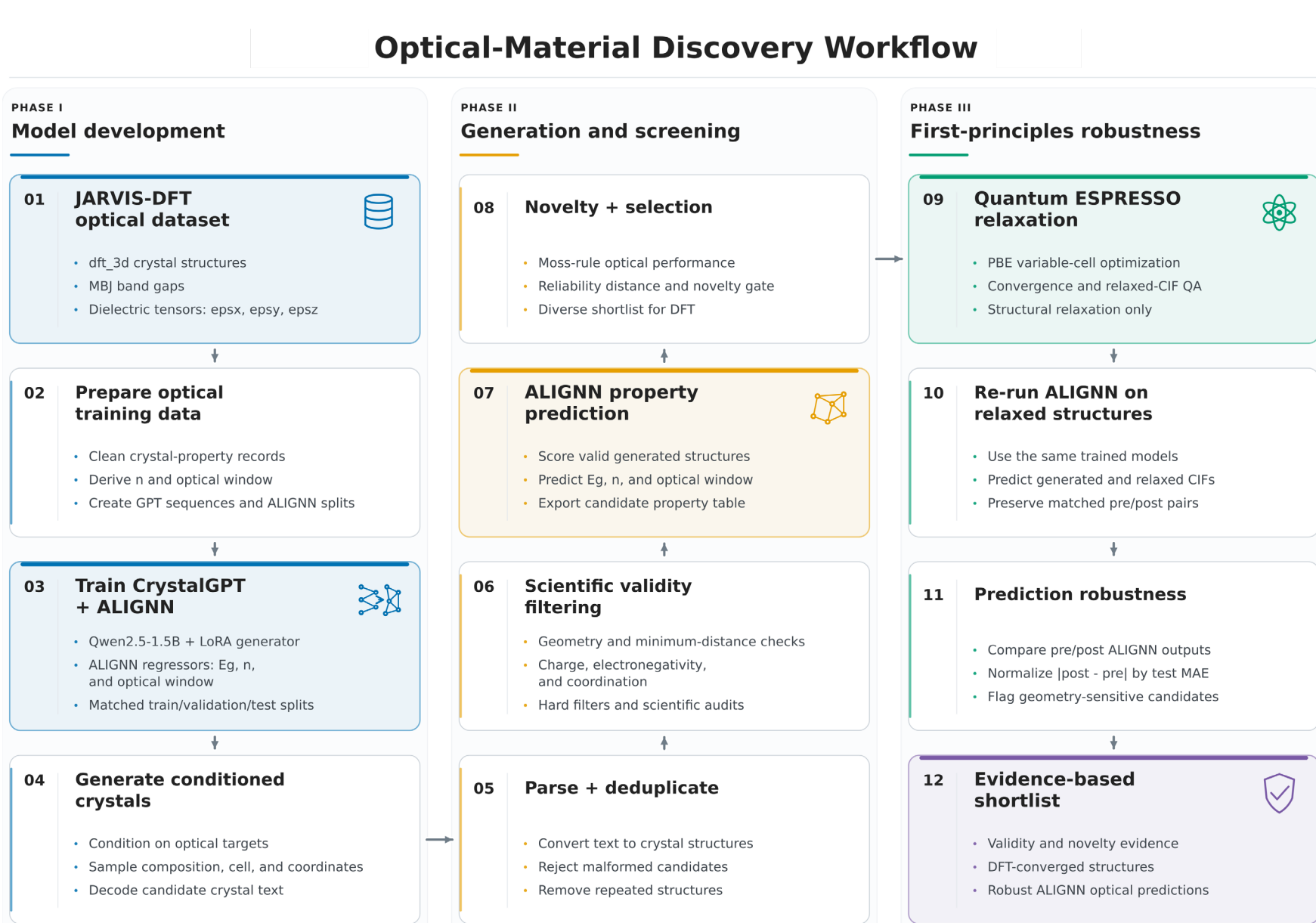
¹Department of ETET, New York City College of Technology, Brooklyn, NY 11201

²ASTD Department, Brookhaven National Laboratory, Upton, NY 11973

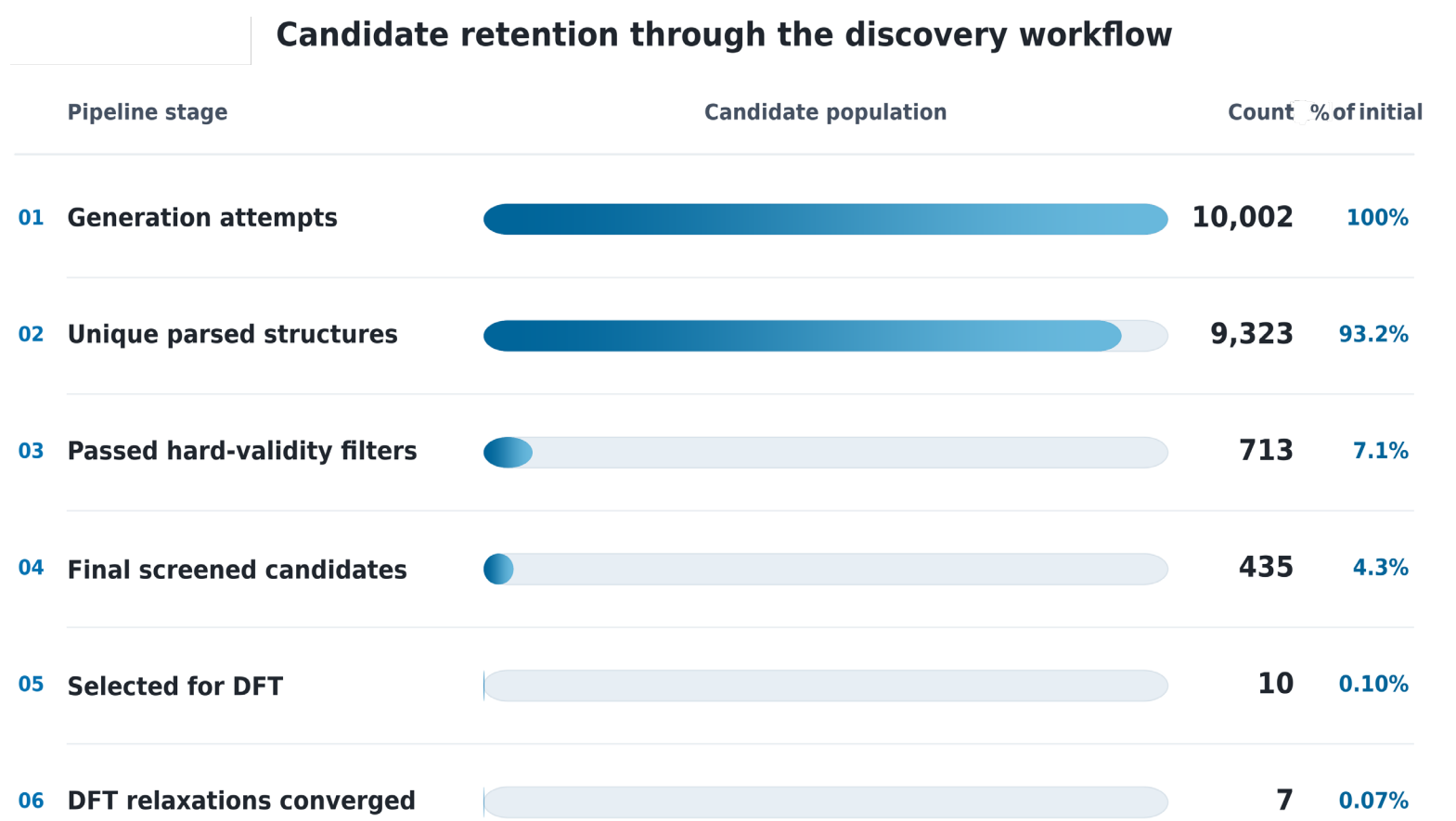
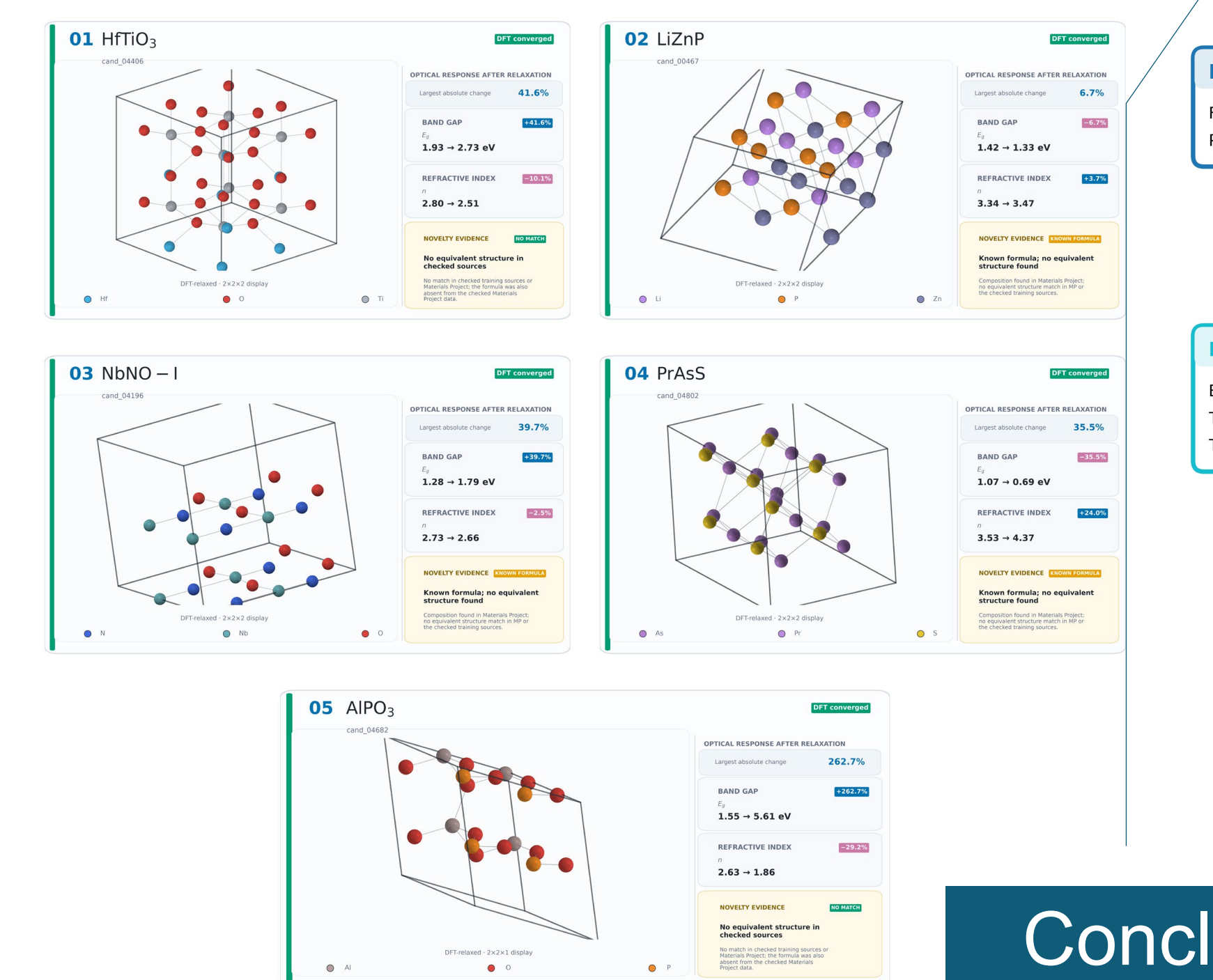
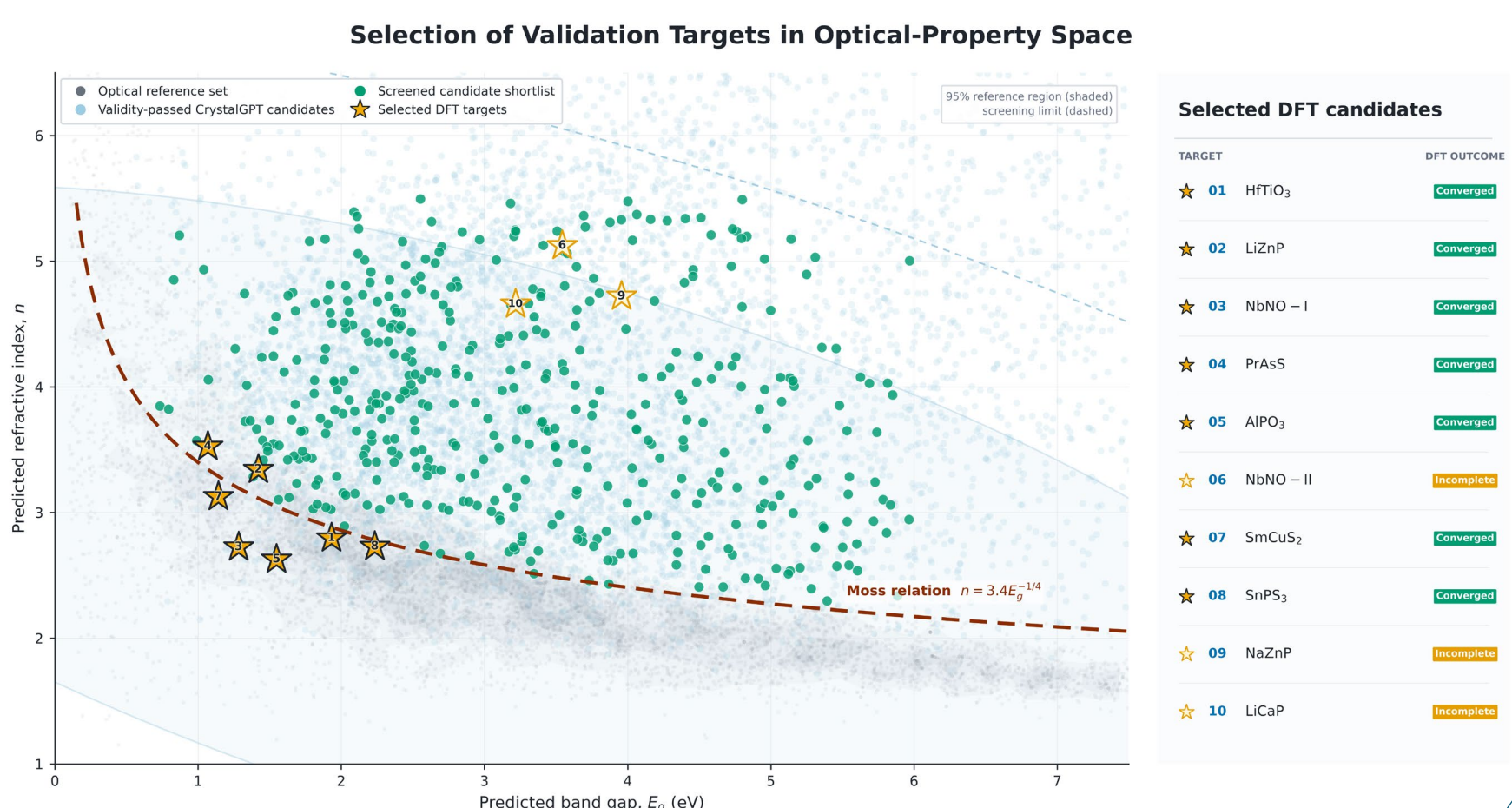
Abstract

Optical materials play pivotal roles in laser systems, ultra-fast electron diffraction (UED), and related instruments used in accelerator science. Identifying novel optical materials to improve accelerators remains difficult due to the vast number of material formulas and crystal structures that would need to be validated which are computationally expensive. This work develops an AI system for optical materials discovery that combines large language model (LLM) crystal generation, physical and chemical validity screening, graph neural network optical property prediction, and density functional theory (DFT) validation to identify promising candidate materials while evaluating the robustness of their predicted properties after structural relaxation. A dataset of 8,330 optical materials was prepared from JARVIS-DFT and used to fine-tune Qwen2.5-1.5B (CrystalGPT), a large language model, through low-rank adaptation (LoRA) to generate new materials and crystal structures. Three ALIGNN graph neural network models were also tested and finetuned on the optical dataset to predict bandgap (eV), refractive index, and transmission width (nm) achieving test MAE of 0.435, 0.181, and 1.406, respectively. The top 10 materials were chosen based on their novelty and chemical and physical plausibility. The crystal structures of the top materials used DFT structural relaxation with Quantum Espresso and the predicted bandgap and refractive index were obtained by using the fine-tuned ALIGNN model with the relaxed structures as input. The DFT structural relaxation computations resulted in 7 of the 10 candidate structures converging. A comparison of pre-relaxation and post-relaxation predictions was conducted and out of the top 10 candidates, 3 candidates were found to have low bandgap, refractive index, and crystal structure shifts illustrating that fine-tuned LLMs can generate chemically and physically valid crystal structures and materials. This project provided hands-on research experience in artificial intelligence, computational materials science, and scientific computing while strengthening the technical and analytical skills needed to contribute to multidisciplinary research in accelerator science.

Methods & Results

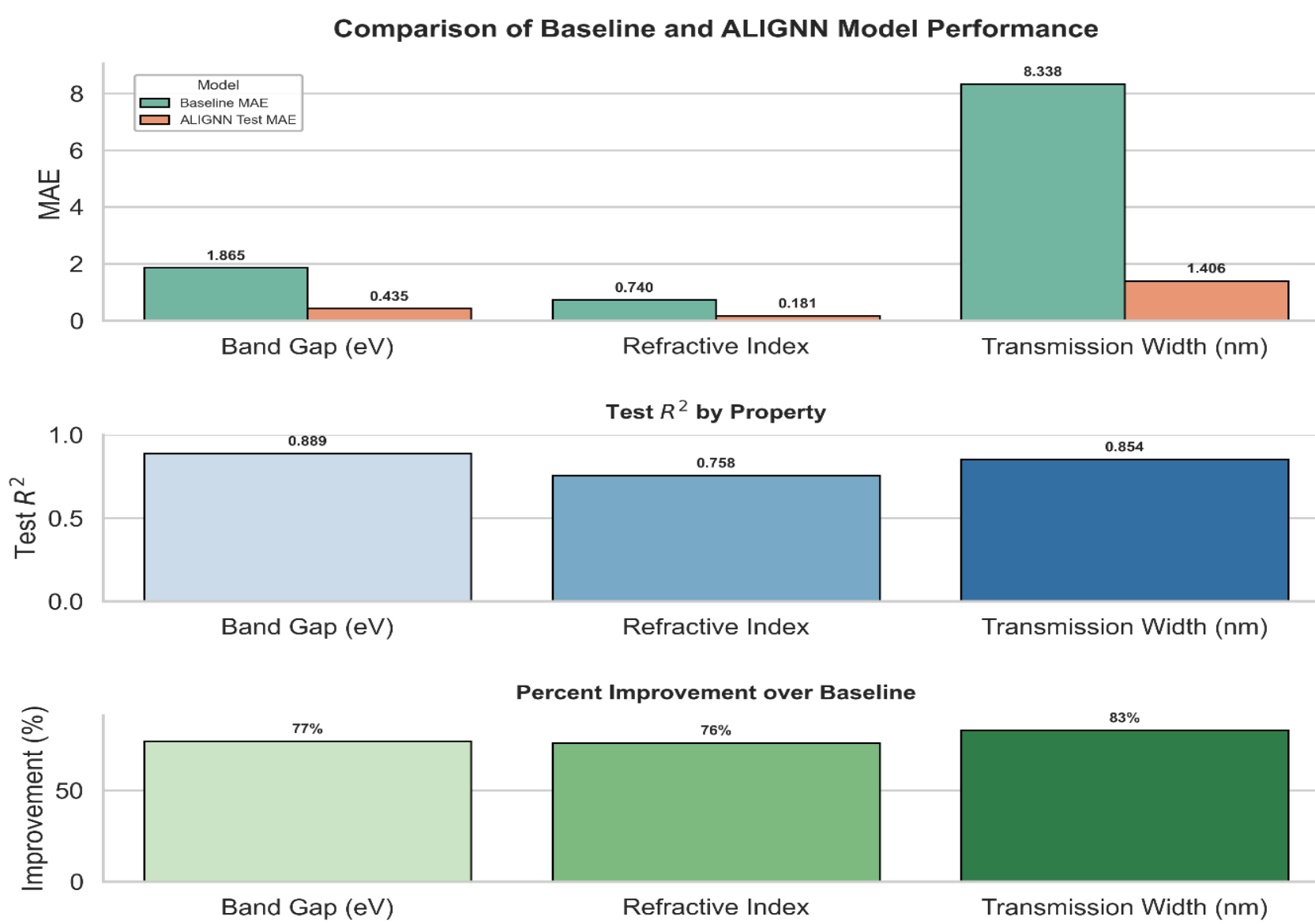
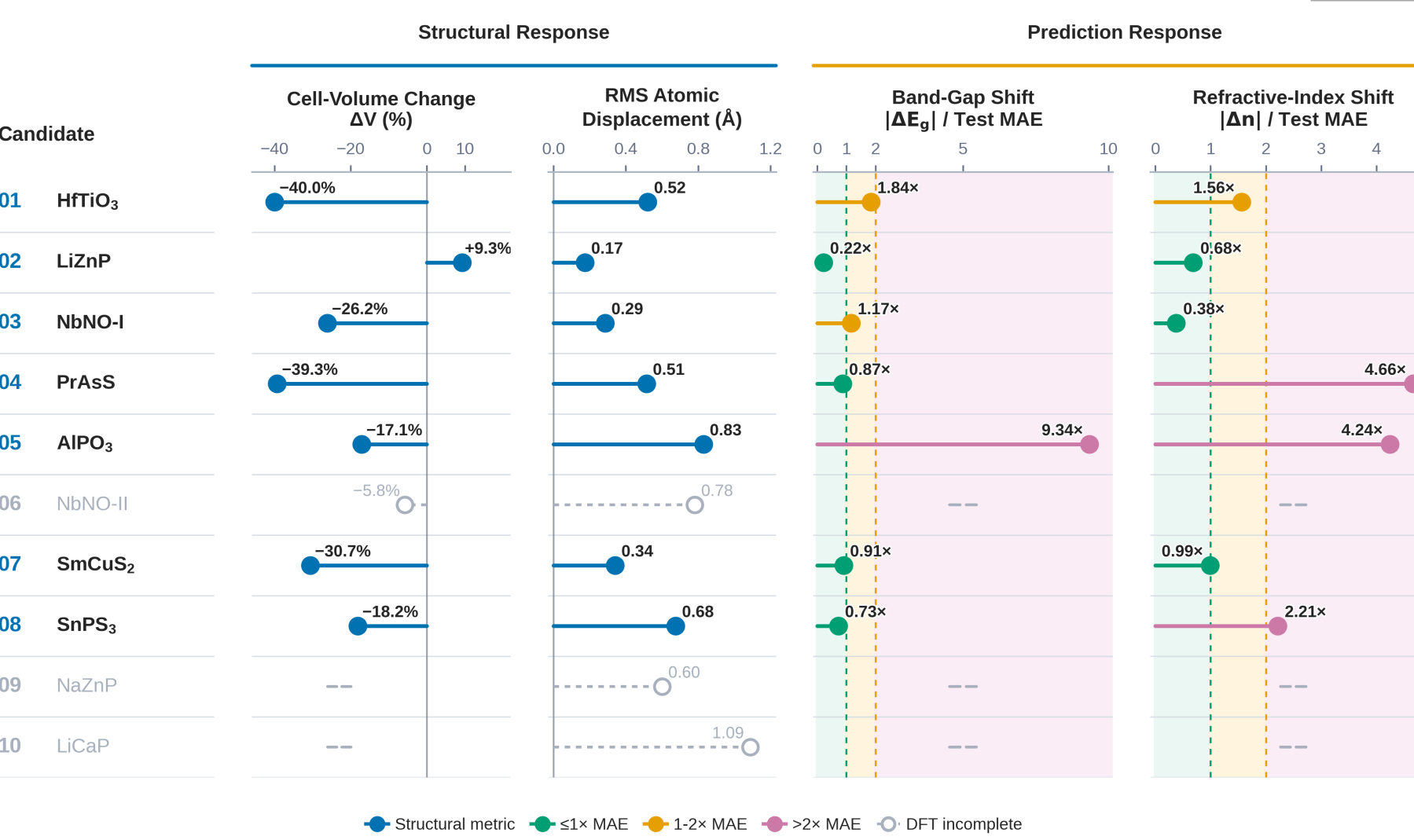


The workflow begins with preparation of the JARVIS-DFT optical dataset and training of the CrystalGPT generator together with three ALIGNN models. CrystalGPT then generates candidate crystal structures conditioned on target optical properties, which are parsed into crystal structures and evaluated using a series of structural and chemical validity checks. Candidates are next screened using the ALIGNN property predictions to identify the most promising materials for density functional theory (DFT) relaxation. The relaxed structures are subsequently re-evaluated with the same ALIGNN models to measure how sensitive the predicted optical properties are to structural relaxation.



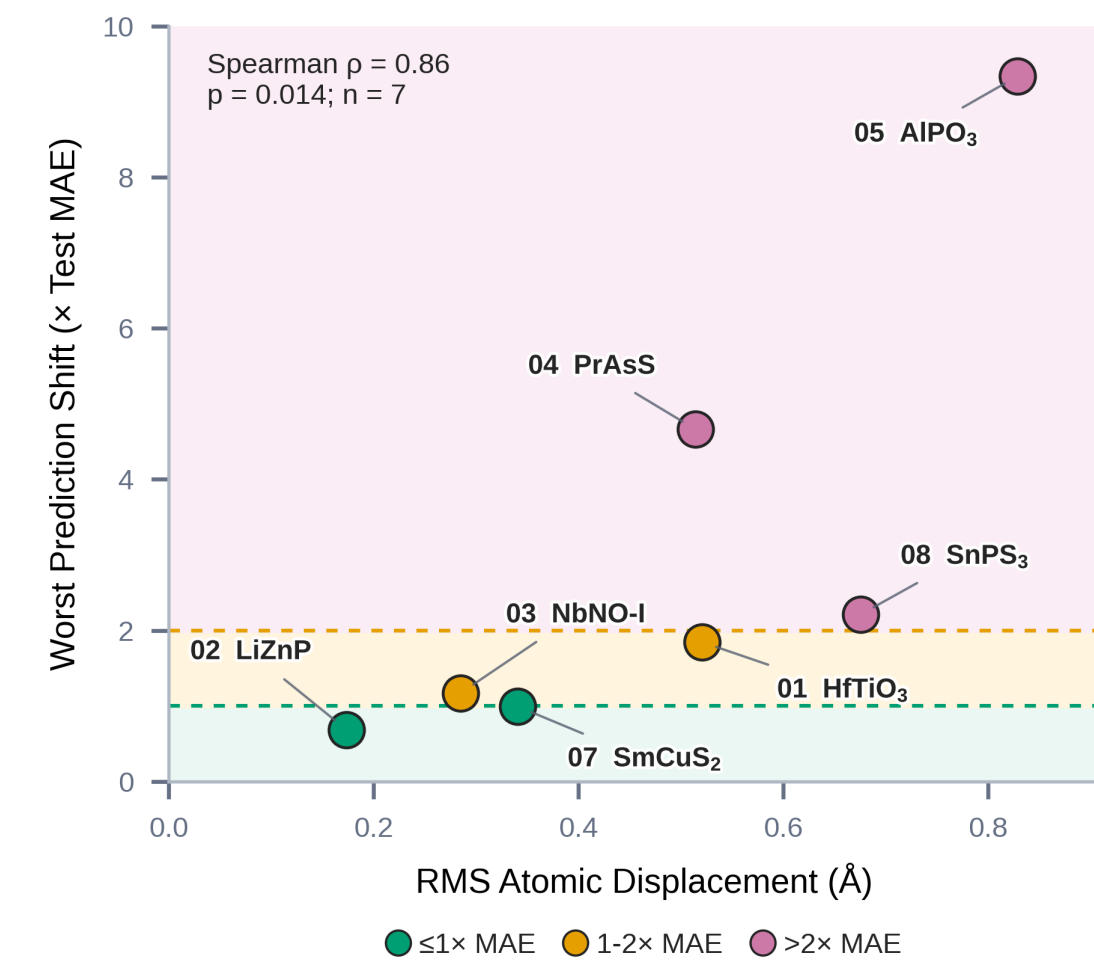
Of 10,002 generation attempts, 9,323 unique outputs were parsed into crystal structures, 713 passed every hard geometry and chemistry validity filter, and 435 passed the cumulative prediction, optical-property, Moss-relation, reliability, novelty, and realism filters. Ten candidates were selected for validation and seven Quantum ESPRESSO DFT relaxation calculations converged.

Candidate-Level Structural and Prediction Response After DFT Relaxation

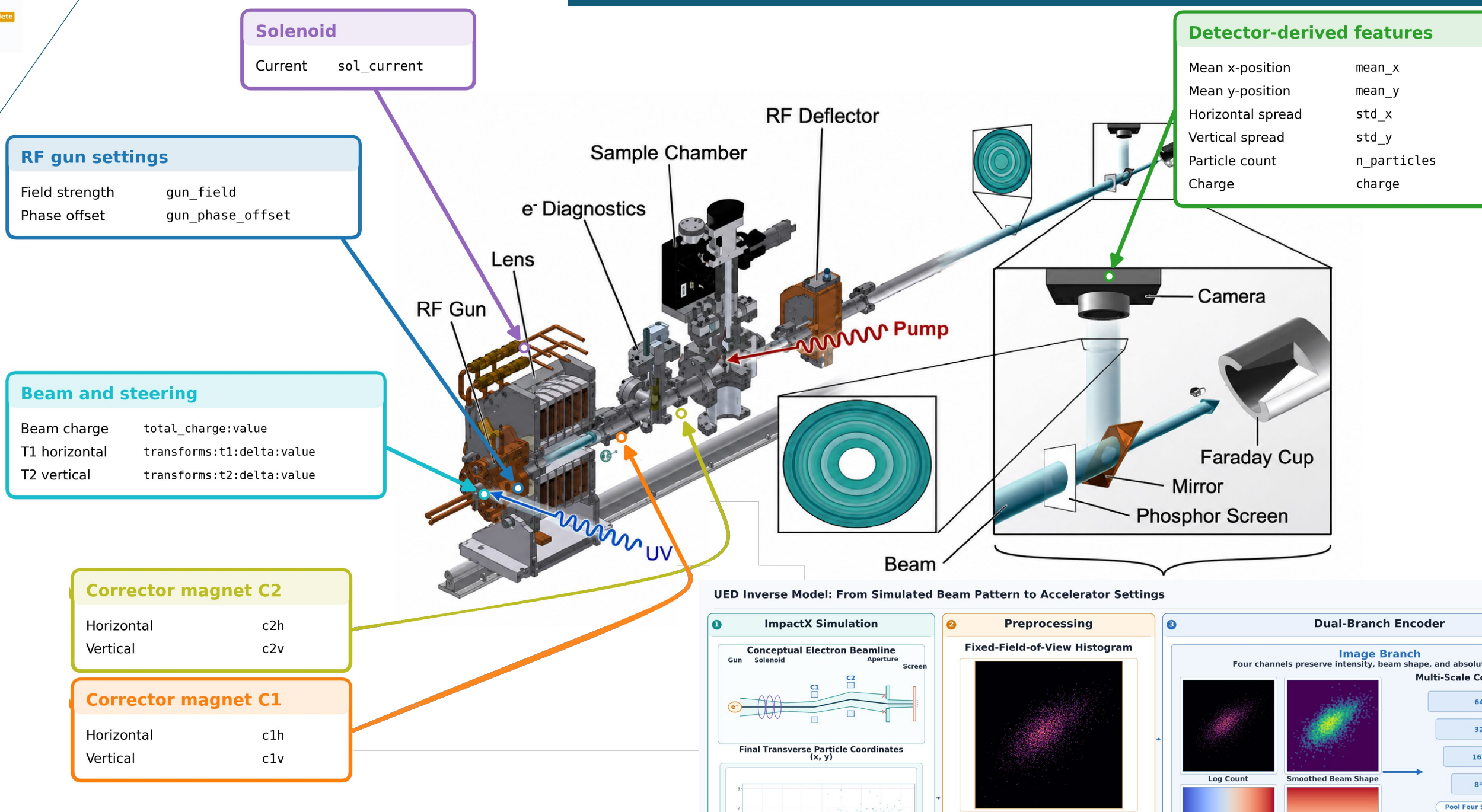


Property	Baseline MAE	ALIGNN Test MAE	Test R ²	Improvement
Band Gap (eV)	1.865	0.435	0.889	77%
Refractive Index	0.740	0.181	0.758	76%
Transmission Width (nm)	8.338	1.406	0.854	83%

Structure-Prediction Relationship After DFT Relaxation



AI assisted UED Alignment



UED CNN Model Results

Parameter / Description	R ²
Corrector Magnet 1 Horizontal (c1h)	0.4097
Corrector Magnet 1 Vertical (c1v)	0.3867
Corrector Magnet 2 Horizontal (c2h)	0.3597
Corrector Magnet 2 Vertical (c2v)	0.3664
Gun Electric Field (gun_field)	0.7848
Gun Phase Offset (gun_phase_offset)	0.7741
Solenoid Current (sol_current)	0.7678
Total Charge (total_charge:value)	0.8075
Transform 1 Offset (transforms:t1:delta:value)	0.8515
Transform 2 Offset (transforms:t2:delta:value)	0.8472
Macro R ²	0.6355

Conclusion & Future Work

- This AI-Deep Learning pipeline for optical materials discovery by combining large language models, graph neural networks, and DFT simulations provides a framework for effectively discovering and screening materials for further analysis and validation.
- Seven out of ten structures converged, suggesting that CrystalGPT can generate physical and chemically plausible crystal structures.
- Re-evaluating the relaxed structures with the ALIGNN models showed that three candidates, LiZnP, SmCuS₂, NbNO-I, retained similar predicted optical properties after relaxation, while others exhibited larger changes suggesting that LLMs can generate structures that are close to their relaxed state.
- The results also suggest that larger crystal structure movement can lead to a larger shift in predicted optical properties. Future work will expand DFT calculations to predict band gap and refractive index, thermodynamic stability calculations and analysis.
- The CNN-beam-feature model demonstrated strong overall performance achieving a macro (R²) of 0.636. It predicted charge, gun field, gun phase, solenoid current, and both transform offsets particularly well ((R²=0.768)–0.851). This model is a promising starting point for accelerating UED alignment.

References & Acknowledgements

- [1] Badrinarayanan, S., Magar, R., Antony, A., Meda, R. S., & Farimani, A. B. (2025). MOFGPT: Generative Design of Metal-Organic Frameworks using Language Models. Journal of Chemical Information and Modeling. DOI: 10.1021/acs.jcim.5c01625.
- [2] Raza, Søren, Kristian Sommer Thygesen, and Gururaj Naik. Breaking the Moss Rule. arXiv, 2026, https://arxiv.org/abs/2602.16247
- [3] K. Choudhary et al., "The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design," npj Computational Materials, vol. 6, art. 173, 2020, doi: 10.1038/s41524-020-00440-1
- [4] Hu, Edward J., et al. LoRA: Low-Rank Adaptation of Large Language Models. International Conference on Learning Representations (ICLR), 2022, https://arxiv.org/abs/2106.09685

These experiences provided a foundation for future accelerator science and engineering careers and were supported by The Science and Engineering Student Apprenticeship Program in Accelerator Research and Development (SESAPARD)-program of Reaching a New Energy Sciences Workforce Program of the U.S. Department of Energy, Office of Science, Grant #DE-SC0025742 and the Workforce Development and Science Education Program at Brookhaven National Laboratory. No export control work



U.S. DEPARTMENT
of ENERGY

www.bnl.gov



Brookhaven
National Laboratory