CrystalGen: Benchmarking GPT, Diffusion, and Flow Models for Atomic Structure Discovery

C. Rhys Campbell, Kamal Choudhary, Aldo Romero

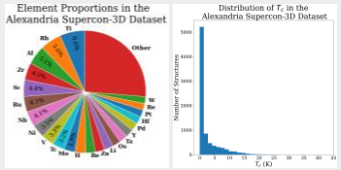
Material Measurement Laboratory, National Institute for Standard and Technology, Gaithersburg, MD 20899

Abstract

We benchmark GPT, diffusion and Riemannian-flow models for crystal-structure generation on JARVIS Supercon 3D and Alexandria Supercon 3D, using consistent metrics (KL divergence of lattice-parameter distributions, MAE of individual parameters, plus space-group, formula and Bravais-lattice recovery). AtomGPT achieves the lowest mean KL divergence (0.0258 nats) and mean MAE (5.13 Å), outperforming CDVAE and FlowMM despite longer training and inference times. This standardized benchmark supports faster innovation in AI-driven materials discovery.

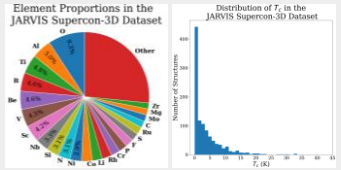
Alexandria Supercon 3D Dataset

Num Structures: 8253



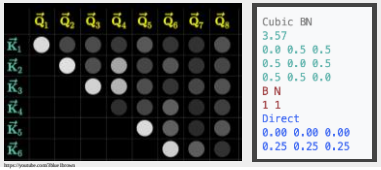
JARVIS Supercon 3D Dataset

Num Structures: 1000



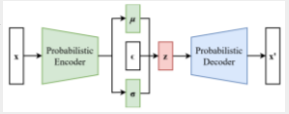
AtomGPT

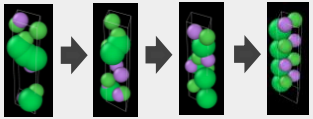
Chemical Formula
Scalar Property
→ Finetuned Large Language Model →
Lattice Parameters



CDVAE

Crystal Diffusion Variational Autoencoder (CDVAE)





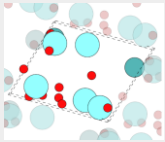
FlowMM

Crystal structure generation with Riemannian Flow Matching.

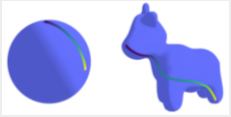
Integrate the following ODE over a curved manifold to evolve an initial crystal system into a more physical one.

$$\frac{d}{dt}\psi_t(x) = u_t(\psi_t(x))$$

$$\psi_0(x) = x \qquad t \in [0,1]$$



Geodesics on Curved Manifolds:



Results

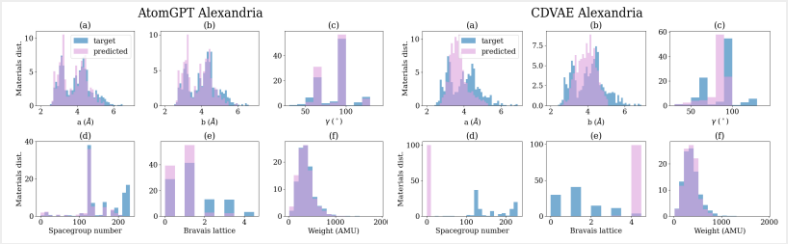


Figure 1: Reconstruction performance of AtomGPT and CDVAE on the Alexandria test set for three representative lattice parameters, space-group number, Bravais lattice, and atomic weight (FlowMM not shown due to space constraints). AtomGPT is the only model to correctly predict both the space group and Bravais lattice, and its lattice-parameter estimates most closely match the target values; FlowMM's predictions fall between those of AtomGPT and CDVAE in accuracy and achieve 100% atomic-weight correctness due to hardcoded formulas.

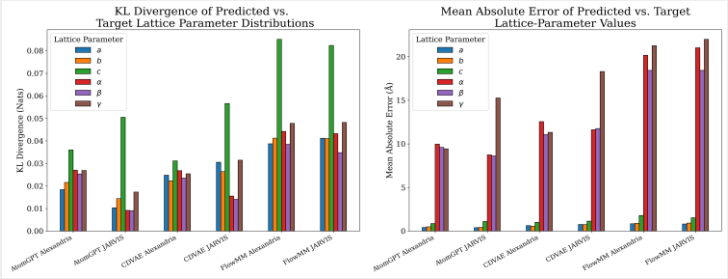


Figure 2: Comparison of KL divergence and MAE across all six lattice parameters for AtomGPT, CDVAE, and FlowMM on both JARVIS and Alexandria datasets. AtomGPT achieves the lowest errors, marginally outperforming CDVAE and markedly outperforming FlowMM; CDVAE also outperforms FlowMM by a significant margin.

Predicted vs. Target	JARVIS Supercon 3D			Alexandria Supercon 3D		
	AtomGPT	CDVAE	FlowMM	AtomGPT	CDVAE	FlowMM
Mean KLD (Nats)	0.0258	0.0294	0.0485	0.0252	0.0257	0.0493
Mean MAE (Å)	5.76	7.39	10.8	5.13	6.19	10.6

Table 1: Model reconstruction performance on the test set. The mean Kullback–Leibler divergence (KLD) across all six lattice parameters quantifies how much the predicted and target distributions differ, while the mean mean absolute error (MAE) across those parameters indicates the average individual prediction error.

Conclusions

AtomGPT consistently outperforms both CDVAE and FlowMM on reconstruction accuracy, achieving the lowest mean KL divergence and mean MAE on both JARVIS and Alexandria benchmarks. CDVAE's predictions, while only marginally less accurate, tend to regress toward the mean of each lattice parameter rather than faithfully reproducing the target distribution. FlowMM strikes a middle ground; its lattice-parameter curves are smoother than CDVAE's but still fall short of AtomGPT's fidelity. These gains come at a cost: AtomGPT requires roughly five to ten times more compute for training and inference. Going forward, optimizing AtomGPT's efficiency could yield a more practical balance of accuracy and speed.

Future Work

1) Instead of using default hyperparameters, conduct thorough hyperparameter sweeps for all three models.

2) Compute AtomGPT benchmarks using many different pretrained LLMs and determine the relationship between number of parameters and prediction accuracy.

3) Benchmark Microsoft's MatterGen crystal structure generation model.

References

AtomGPT: https://github.com/usnistgov/atomgpt
CDVAE: https://github.com/txie-93/cdvae
FlowMM: https://github.com/facebookresearch/flowmm
JARVIS Tools: https://jarvis.nist.gov/
Alexandria Database: https://alexandria.icams.rub.de/