A critical assessment of generative AI for inorganic materials discovery

Nathan Szymanski & Chris Bartel

Design of Materials on Computers
DMC LAB
Bartel Research Group

bartel.cems.umn.edu
Chemical Engineering & Materials Science
University of Minnesota

IPRIME

IPRIME Annual Meeting | May 28, 2025 | 4D Program Review

Generative AI for text and images has become mainstream

CEMS
Chemical Engineering
& Materials Science

ChatGPT

Claude

Gemini

deepseek

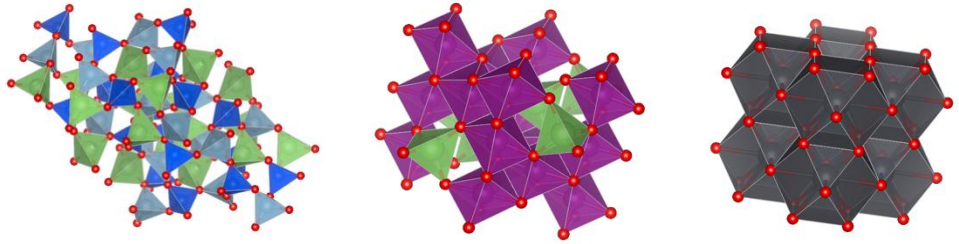
Adobe Firefly

Midjourney

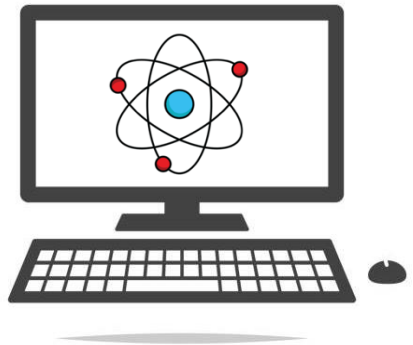
imagen

DALL·E

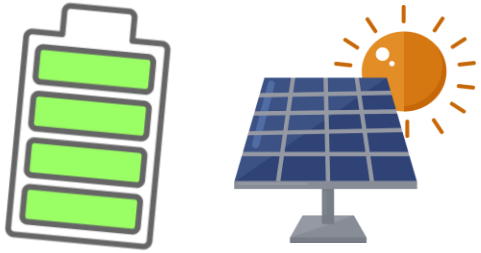
Can AI generate *materials* with desired functionality?



Structure and composition



Quantum chemistry calculations

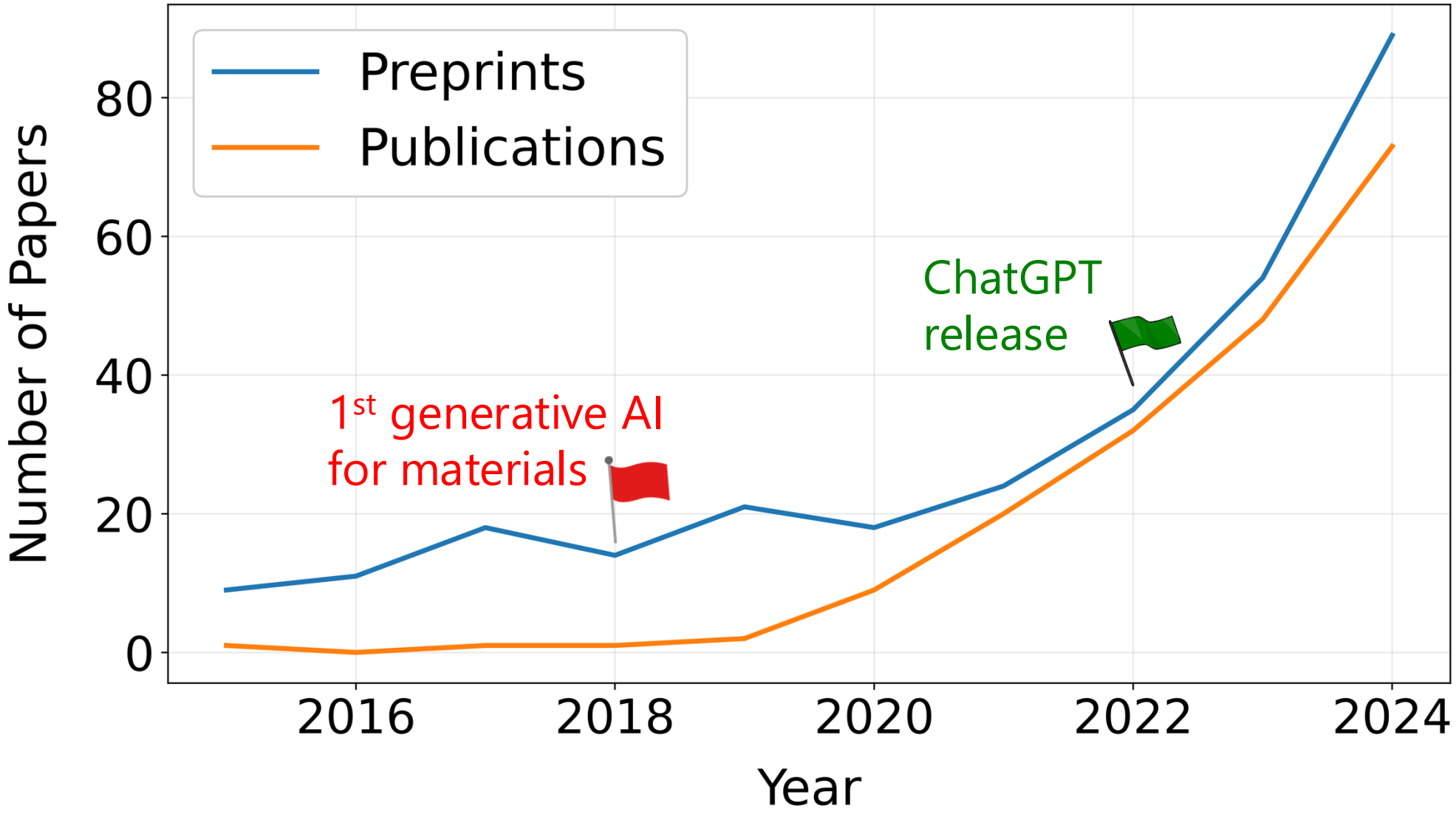


Predicted functionality

Generative AI may help us solve the *inverse problem*

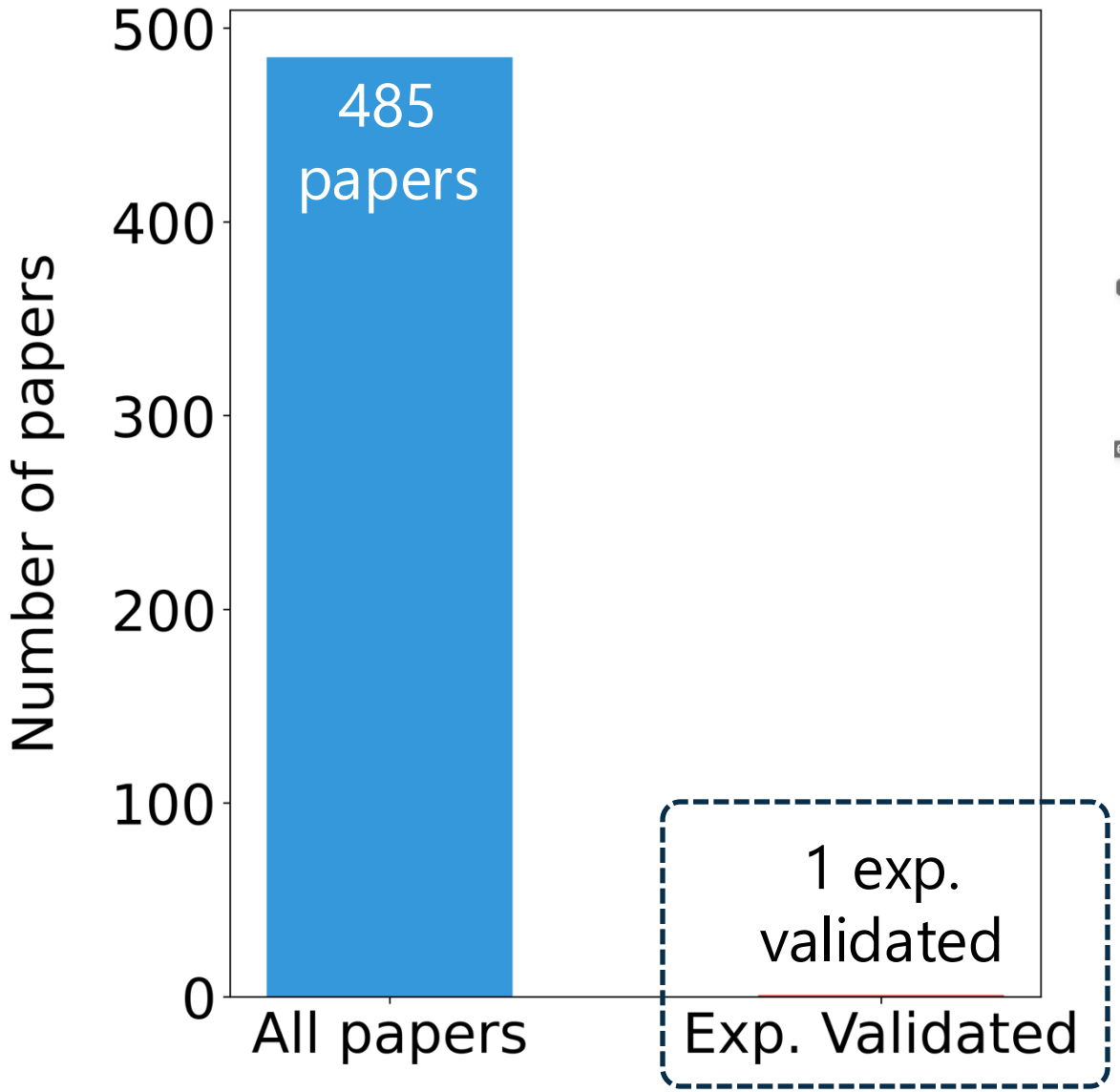
Given a desired functionality: propose hypothetical materials

Many generative models are being developed for materials



These are papers with ***generative*** and ***materials*** in their titles

A gap remains between *proposed* and *synthesized* materials



Article

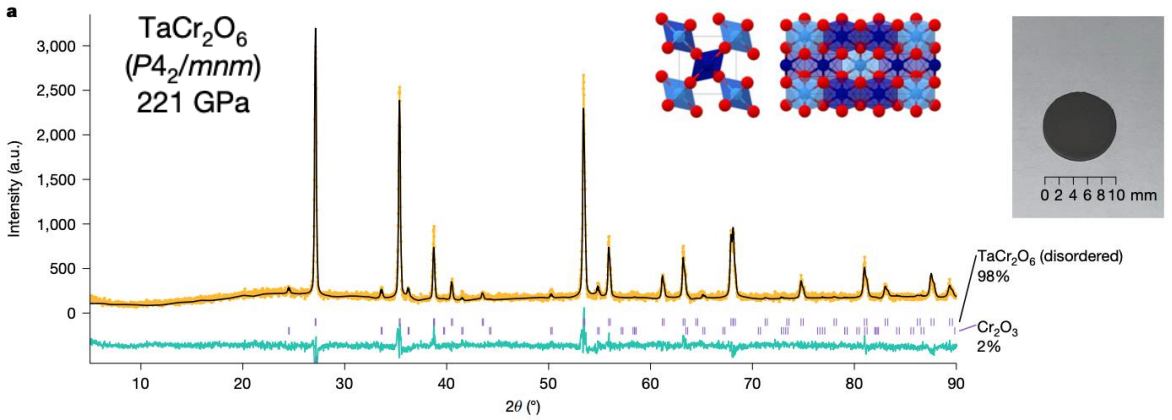
nature

A generative model for inorganic materials design

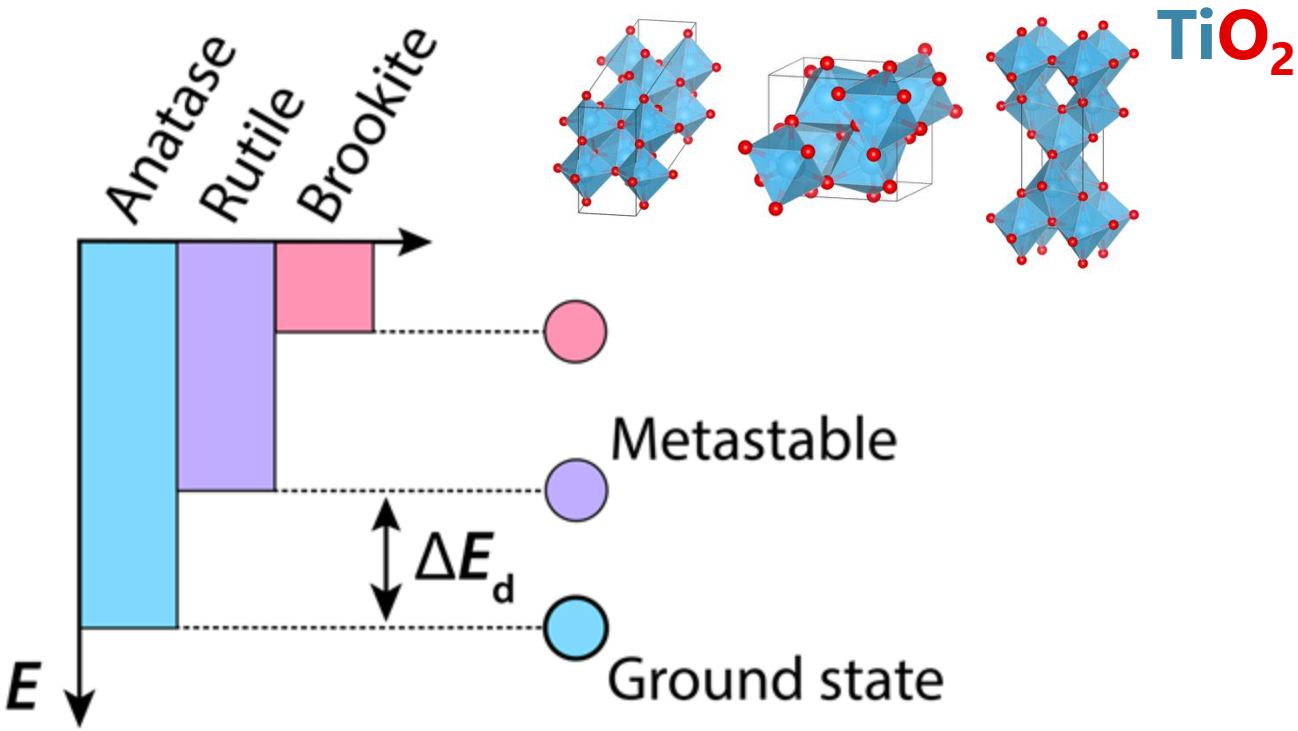
https://doi.org/10.1038/s41586-025-08628-5
Received: 17 January 2024
Accepted: 10 January 2025
Published online: 16 January 2025
Open access
Check for updates

Claudio Zeni[1,8], Robert Pinsler[1,8], Daniel Zügner[2,8], Andrew Fowler[1,8], Matthew Horton[3,8], Xiang Fu[1], Zilong Wang[4], Aliaksandra Shysheya[1], Jonathan Crabbé[1], Shoko Ueda[1], Roberto Sordillo[1], Lixin Sun[1], Jake Smith[3], Bichlien Nguyen[3], Hannes Schulz[2], Sarah Lewis[1], Chin-Wei Huang[5], Ziheng Lu[6], Yichi Zhou[7], Han Yang[6], Hongxia Hao[6], Jielan Li[6], Chunlei Yang[4], Wenjie Li[4], Ryota Tomioka[1,8] ✉ & Tian Xie[1,8] ✉

The design of functional materials with desired properties is essential in driving technological advances in areas such as energy storage, catalysis and carbon capture. Generative models accelerate materials design by directly generating new materials given desired property constraints, but current methods have a low success rate in proposing stable crystals or can satisfy only a limited set of property constraints.



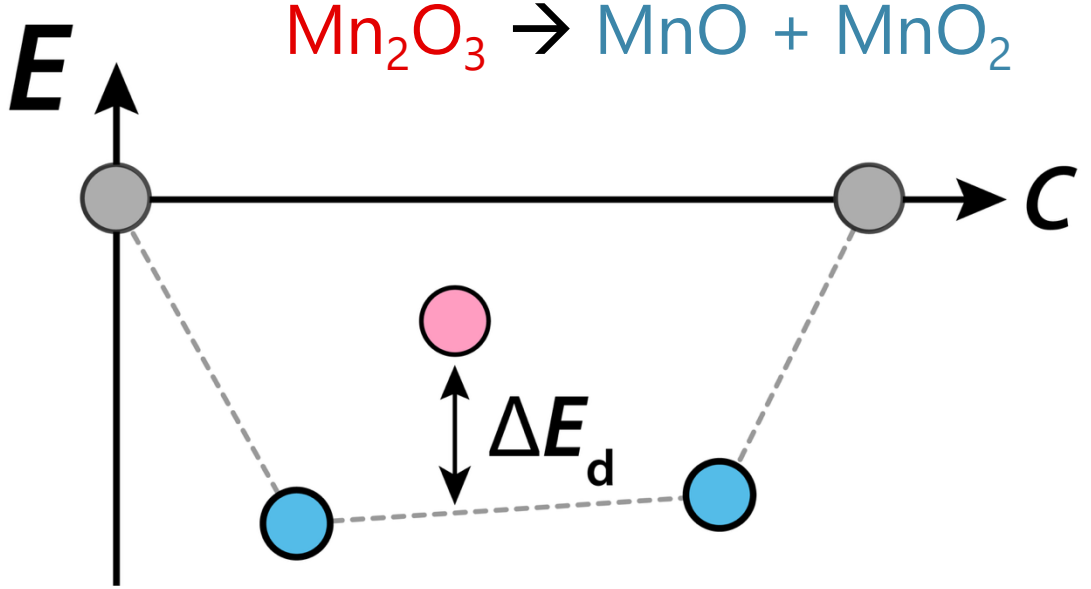
How can we quantify the likelihood of synthesizability?



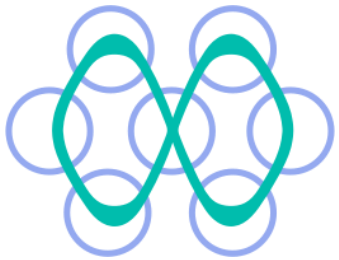
TiO_2

Example:

$Mn_2O_3 \rightarrow MnO + MnO_2$

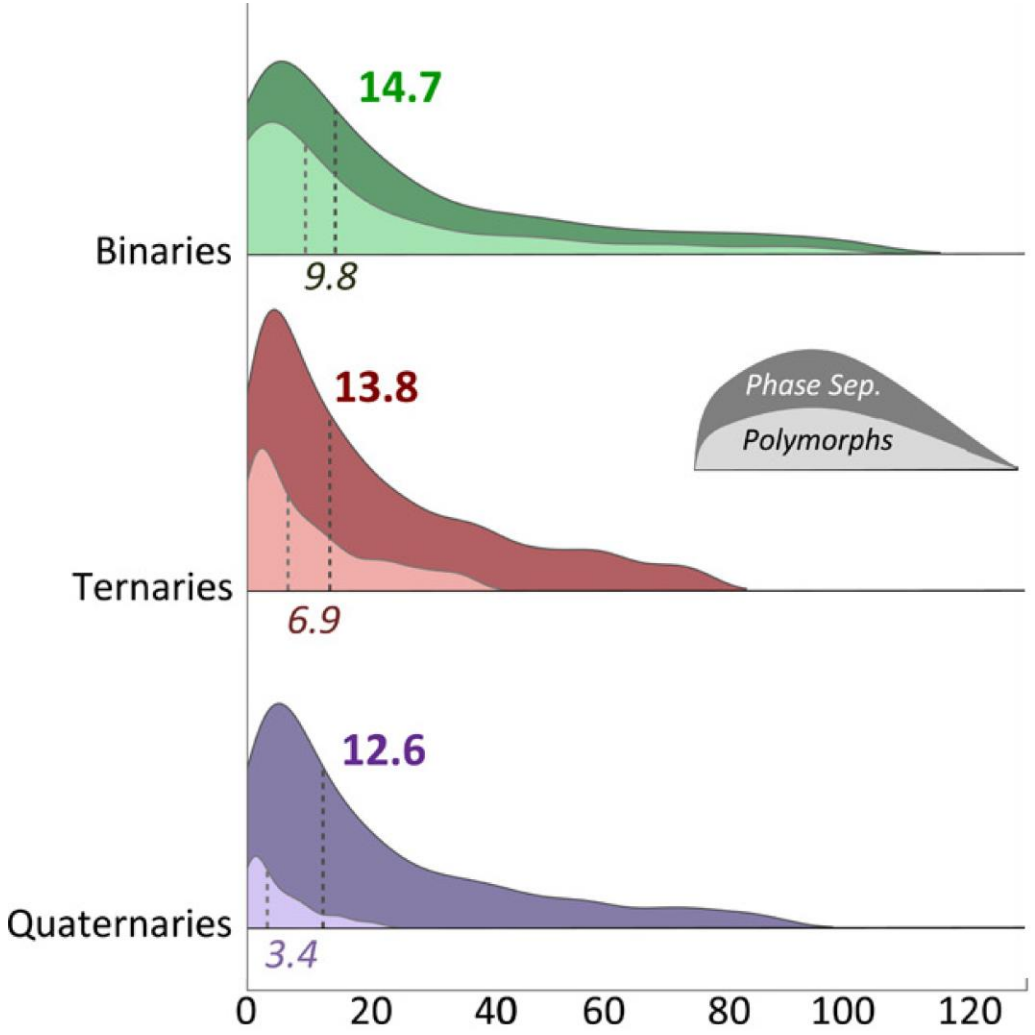


ΔE_d is the **decomposition energy**, which we can get from density functional theory (DFT) calculations



Bartel, *Journal of Materials Science*, 2022

Materials with low ΔE_d are more likely to be synthesizable



When designing new materials, they should either:

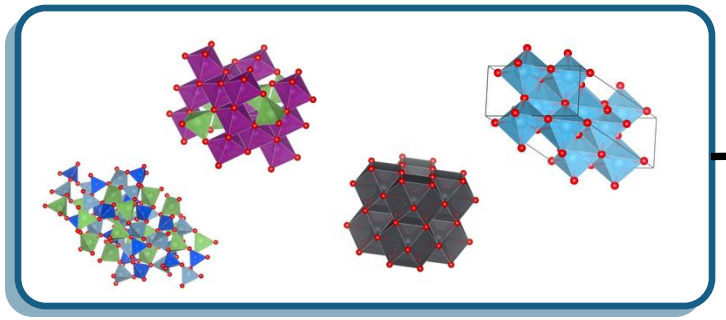
- **Be stable ($\Delta E_d \leq 0$)**
- **Have a small ΔE_d**

Figure adapted from Sun *et al.*, Sci. Adv. (2016)

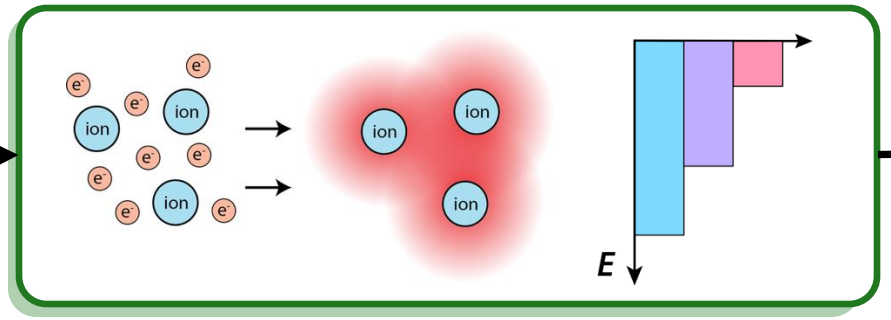
A framework to assess generative AI models using ΔE_d

CEMS Chemical Engineering & Materials Science

Materials generation



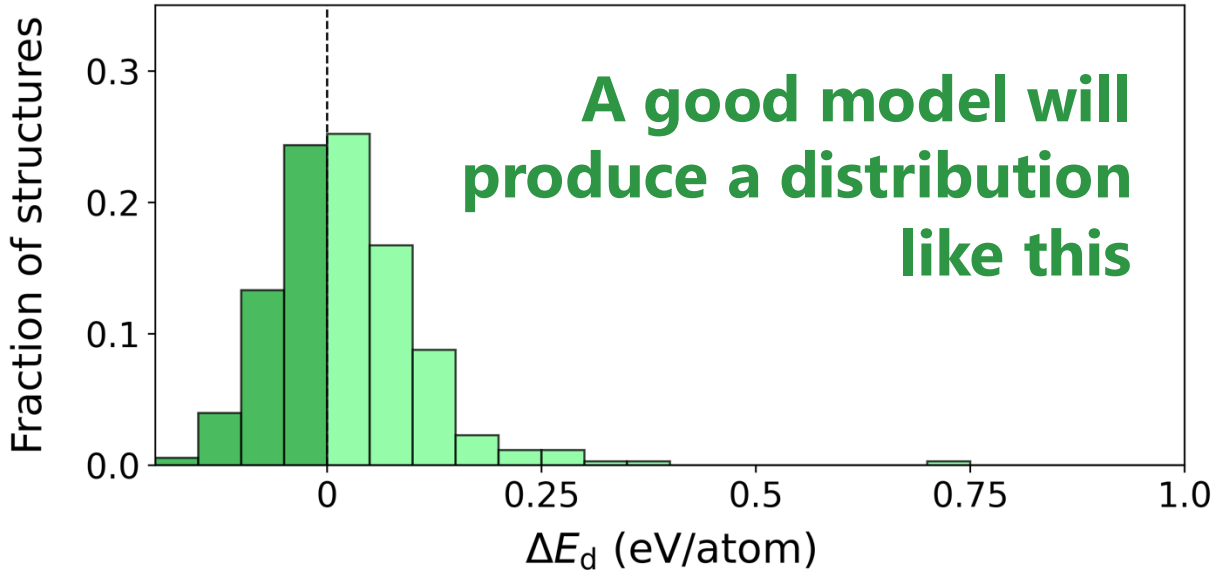
DFT energy calculations

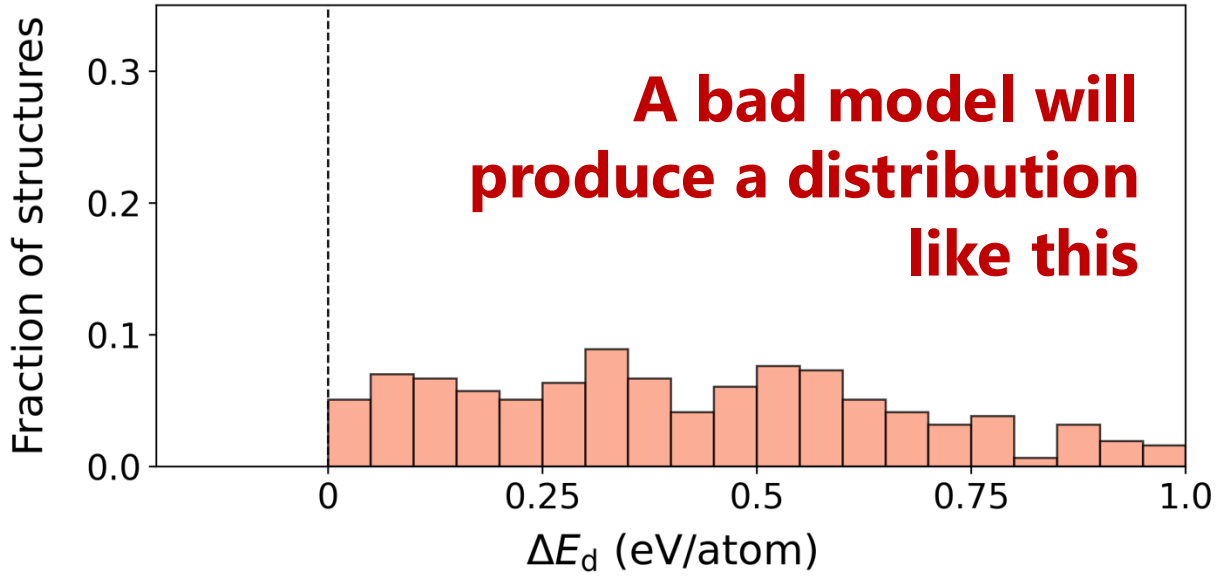


Comparison to MP

~200k compounds

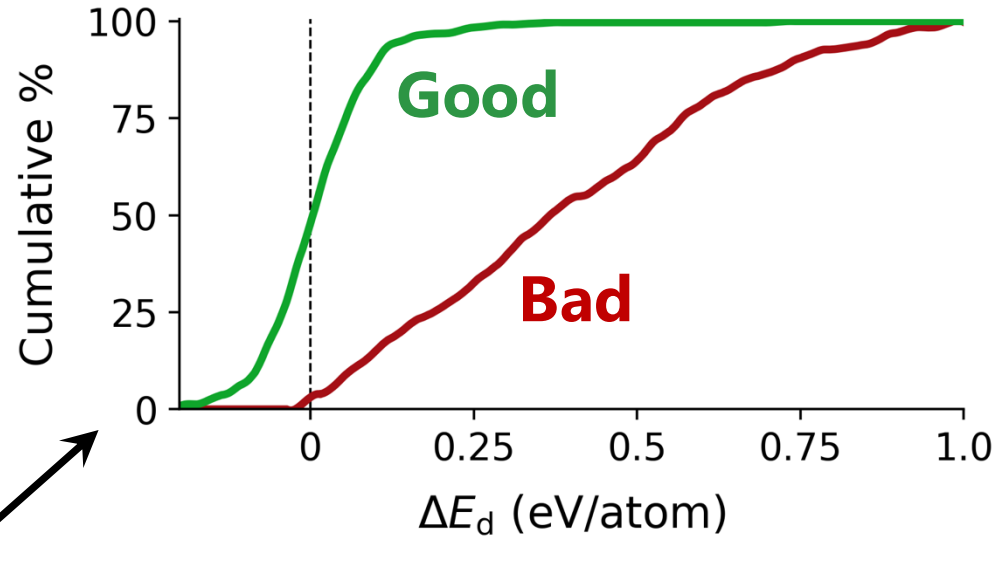
MatGen-Baselines

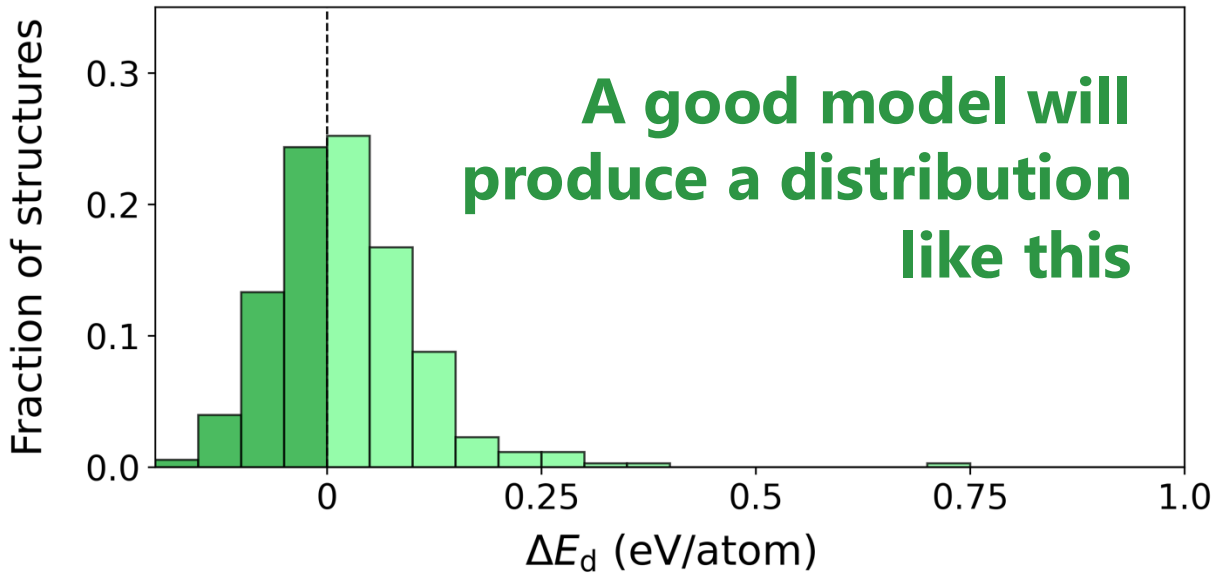


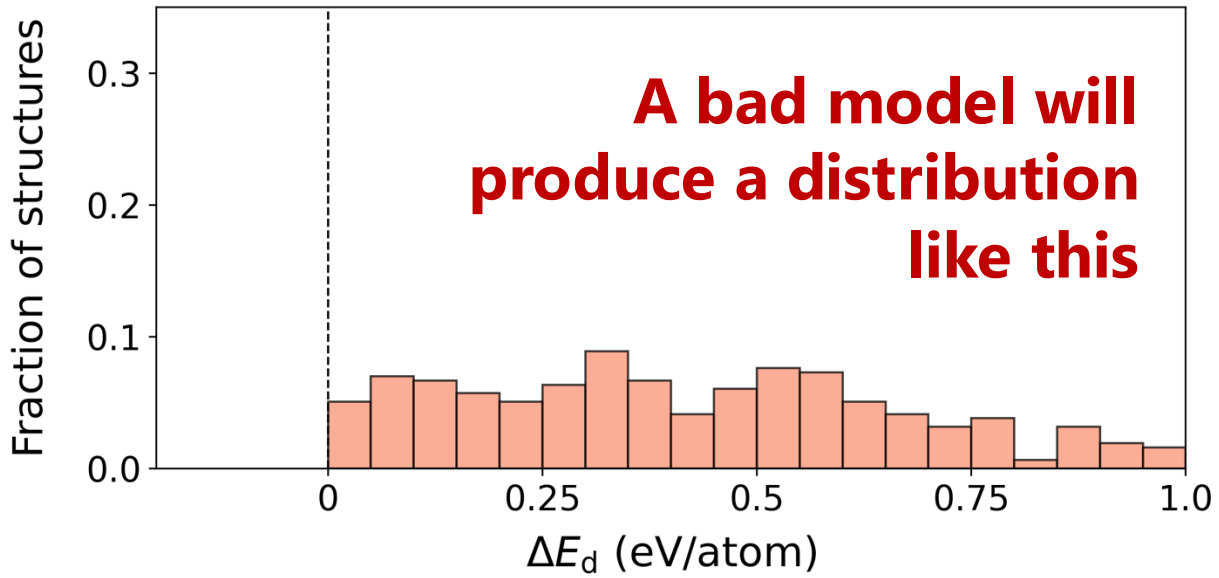


pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

A framework to assess generative AI models using ΔE_d





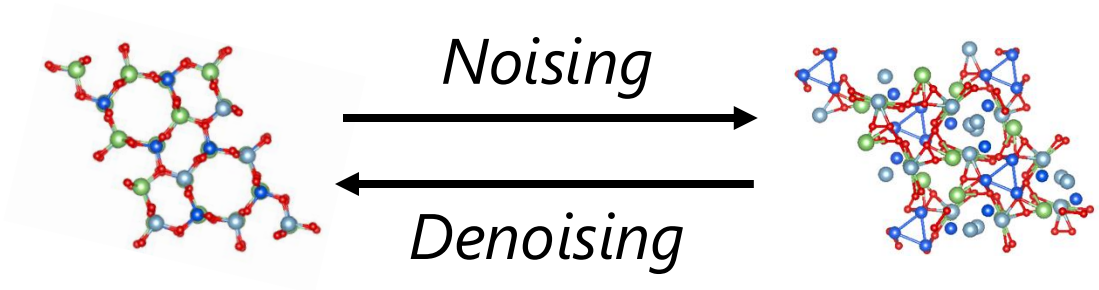


pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

We assessed three types of generative models

CEMS
Chemical Engineering
& Materials Science

CDVAE: diffusion model



FTCP: autoencoder

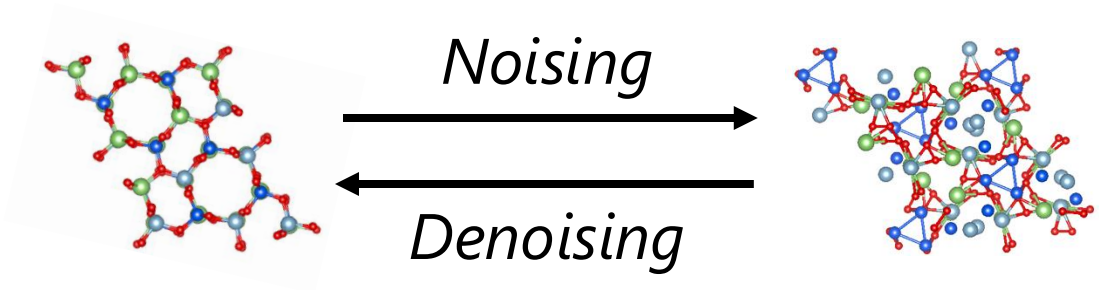
CrystaLLM: language model

```
# generated using pymatgen
data_BaTiO3
_symmetry_space_group_name_H-M   Pm-3m
_cell_length_a   4.64914500
_cell_length_b   4.64914500
_cell_length_c   4.64914500
_cell_angle_alpha   90.00000000
_cell_angle_beta   90.00000000
_cell_angle_gamma   90.00000000
_symmetry_Int_Tables_number   221
_chemical_formula_structural   BaTiO3
_chemical_formula_sum   'Ba1 Ti1 O3'
_cell_volume   100.48917348
_cell_formula_units_Z   1
loop_
 _symmetry_equiv_pos_site_id
 _symmetry_equiv_pos_as_xyz
  1  'x, y, z'
  2  '-x, -y, -z'
  3  '-y, x, z'
  4  'y, -x, -z'
  5  '-x, -y, z'
  6  'x, y, -z'
  7  'y, -x, z'
  8  '-y, x, -z'
loop_
 _atom_type_symbol
 _atom_type_oxidation_number
 Ti4+ 4.0
 O2- -2.0
loop_
 _atom_site_type_symbol
 _atom_site_label
 _atom_site_symmetry_multiplicity
 _atom_site_fract_x
 _atom_site_fract_y
 _atom_site_fract_z
 _atom_site_occupancy
 Ti4+ Ti0 1 0.00000000 0.00000000 0.00000000 1
 Ti4+ Ti1 1 0.50000000 0.50000000 0.50000000 1
 O2- Ti2 3 0.00000000 0.00000000 0.50000000 1
```

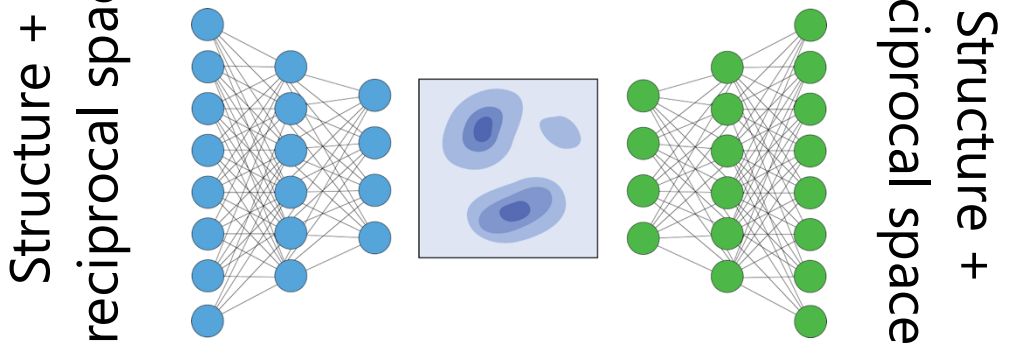
We assessed three types of generative models

CEMS
Chemical Engineering
& Materials Science

CDVAE: diffusion model



FTCP: autoencoder



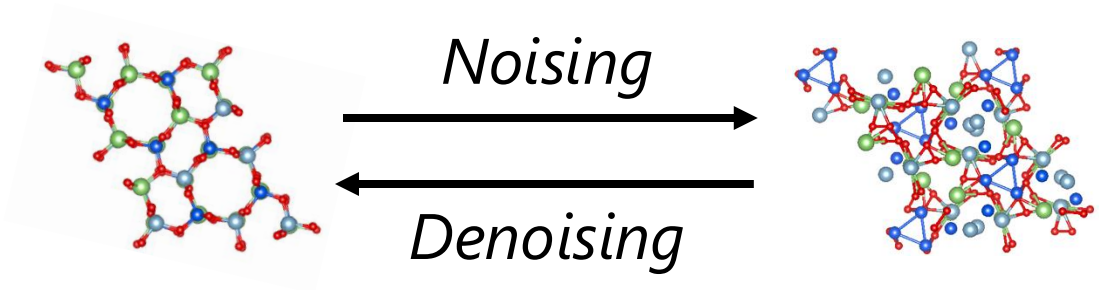
CrystaLLM: language model

```
# generated using pymatgen
data_BaTiO3
_symmetry_space_group_name_H-M   Pm-3m
_cell_length_a   4.64914500
_cell_length_b   4.64914500
_cell_length_c   4.64914500
_cell_angle_alpha   90.00000000
_cell_angle_beta   90.00000000
_cell_angle_gamma   90.00000000
_symmetry_Int_Tables_number   221
_chemical_formula_structural   BaTiO3
_chemical_formula_sum   'Ba1 Ti1 O3'
_cell_volume   100.48917348
_cell_formula_units_Z   1
loop_
 _symmetry_equiv_pos_site_id
 _symmetry_equiv_pos_as_xyz
  1  'x, y, z'
  2  '-x, -y, -z'
  3  '-y, x, z'
  4  'y, -x, -z'
  5  '-x, -y, z'
  6  'x, y, -z'
  7  'y, -x, z'
  8  '-y, x, -z'
loop_
 _atom_type_symbol
 _atom_type_oxidation_number
 Ti4+  4.0
 O2-  -2.0
loop_
 _atom_site_type_symbol
 _atom_site_label
 _atom_site_symmetry_multiplicity
 _atom_site_fract_x
 _atom_site_fract_y
 _atom_site_fract_z
 _atom_site_occupancy
 Ti4+  Ti0  1  0.00000000  0.00000000  0.00000000  1
 Ti4+  Ti1  1  0.50000000  0.50000000  0.50000000  1
 O2-  Ti2  3  0.00000000  0.00000000  0.50000000  1
```

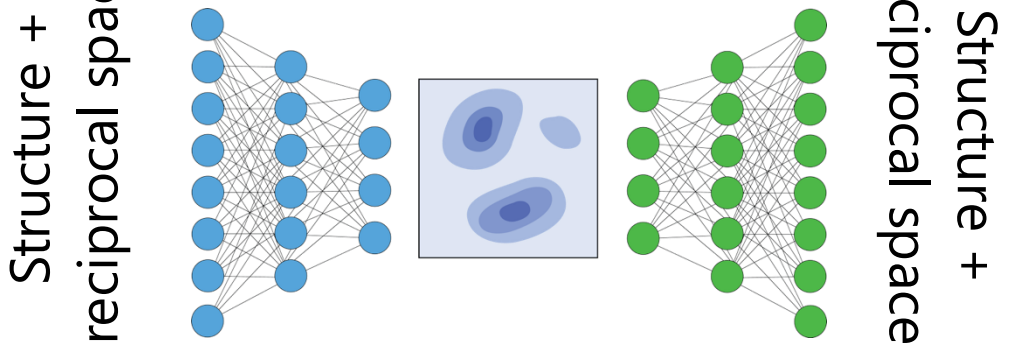
We assessed three types of generative models

CEMS
Chemical Engineering
& Materials Science

CDVAE: diffusion model



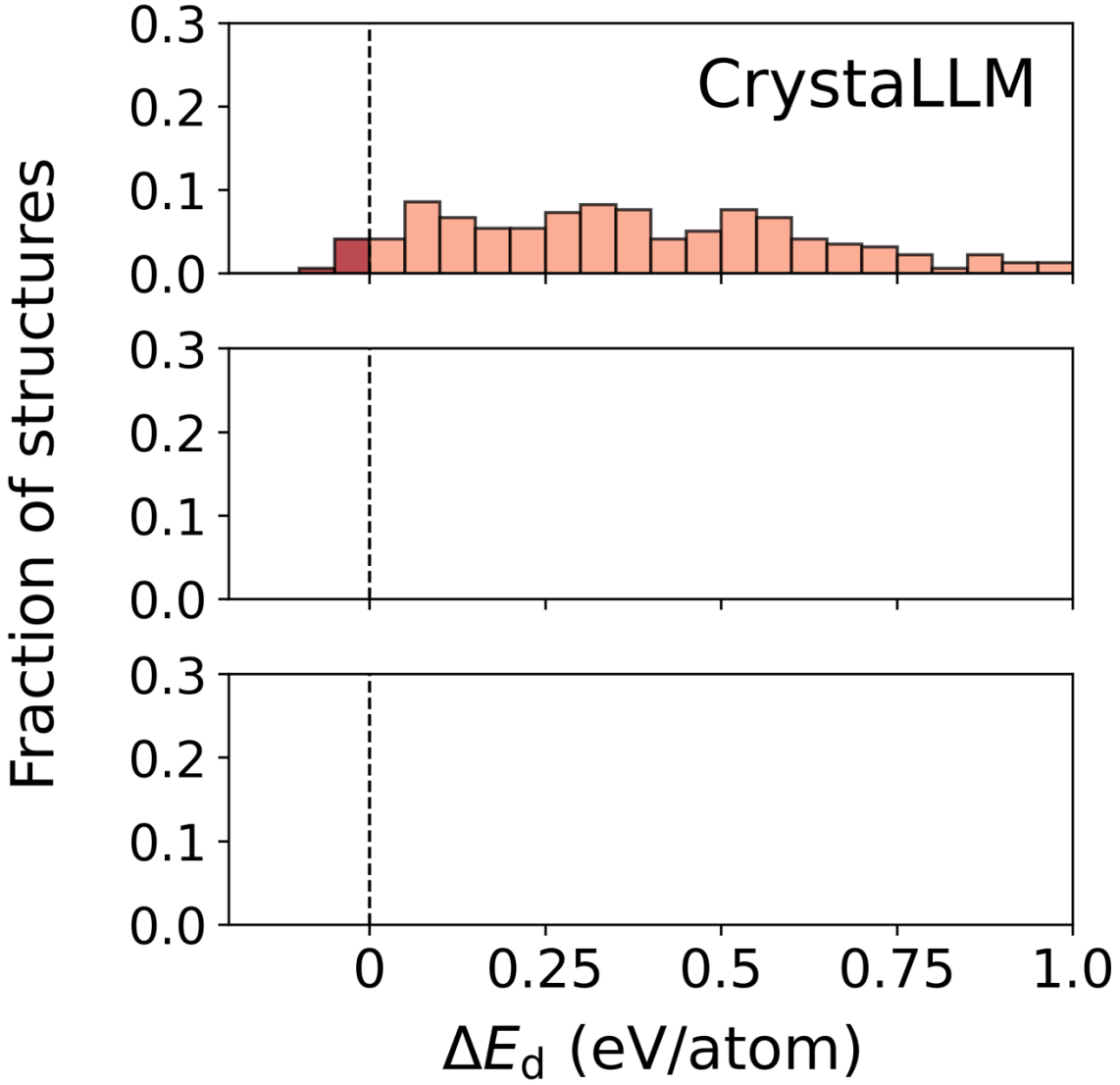
FTCP: autoencoder



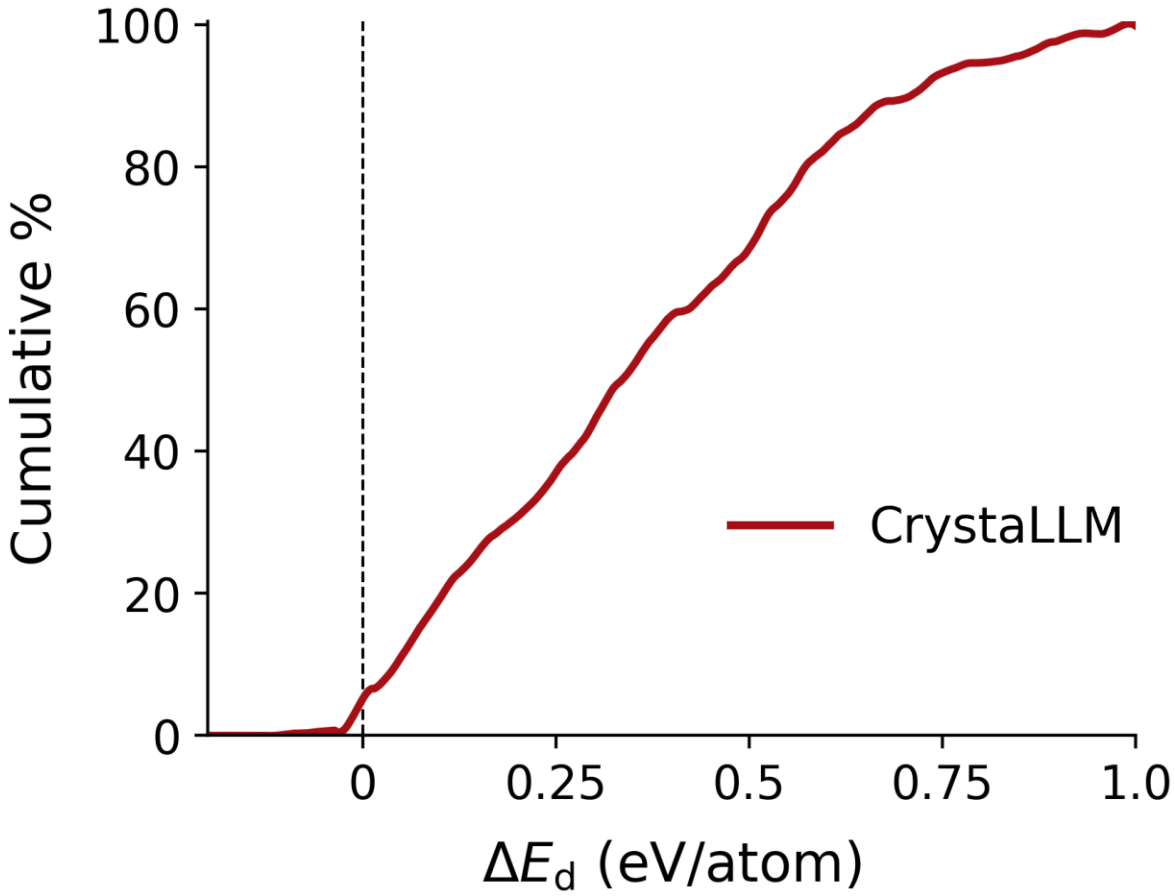
CrystaLLM: language model

```
# generated using pymatgen
data_BaTiO3
_symmetry_space_group_name_H-M  Pm-3m
_cell_length_a  4.64914500
_cell_length_b  4.64914500
_cell_length_c  4.64914500
_cell_angle_alpha  90.00000000
_cell_angle_beta  90.00000000
_cell_angle_gamma  90.00000000
_symmetry_Int_Tables_number  221
_chemical_formula_structural  BaTiO3
_chemical_formula_sum  'Ba1 Ti1 O3'
_cell_volume  100.48917348
_cell_formula_units_Z  1
loop_
 _symmetry_equiv_pos_site_id
 _symmetry_equiv_pos_as_xyz
 1 'x, y, z'
 2 '-x, -y, -z'
 3 '-y, x, z'
 4 'y, -x, -z'
 5 '-x, -y, z'
 6 'x, y, -z'
 7 'y, -x, z'
 8 '-y, x, -z'
loop_
 _atom_type_symbol
 _atom_type_oxidation_number
 Ti4+ 4.0
 O2- -2.0
loop_
 _atom_site_type_symbol
 _atom_site_label
 _atom_site_symmetry_multiplicity
 _atom_site_fract_x
 _atom_site_fract_y
 _atom_site_fract_z
 _atom_site_occupancy
 Ti4+ Ti0 1 0.00000000 0.00000000 0.00000000 1
 Ti4+ Ti1 1 0.50000000 0.50000000 0.50000000 1
 O2- Ti2 3 0.00000000 0.00000000 0.50000000 1
```

Results show varied performance among distinct models

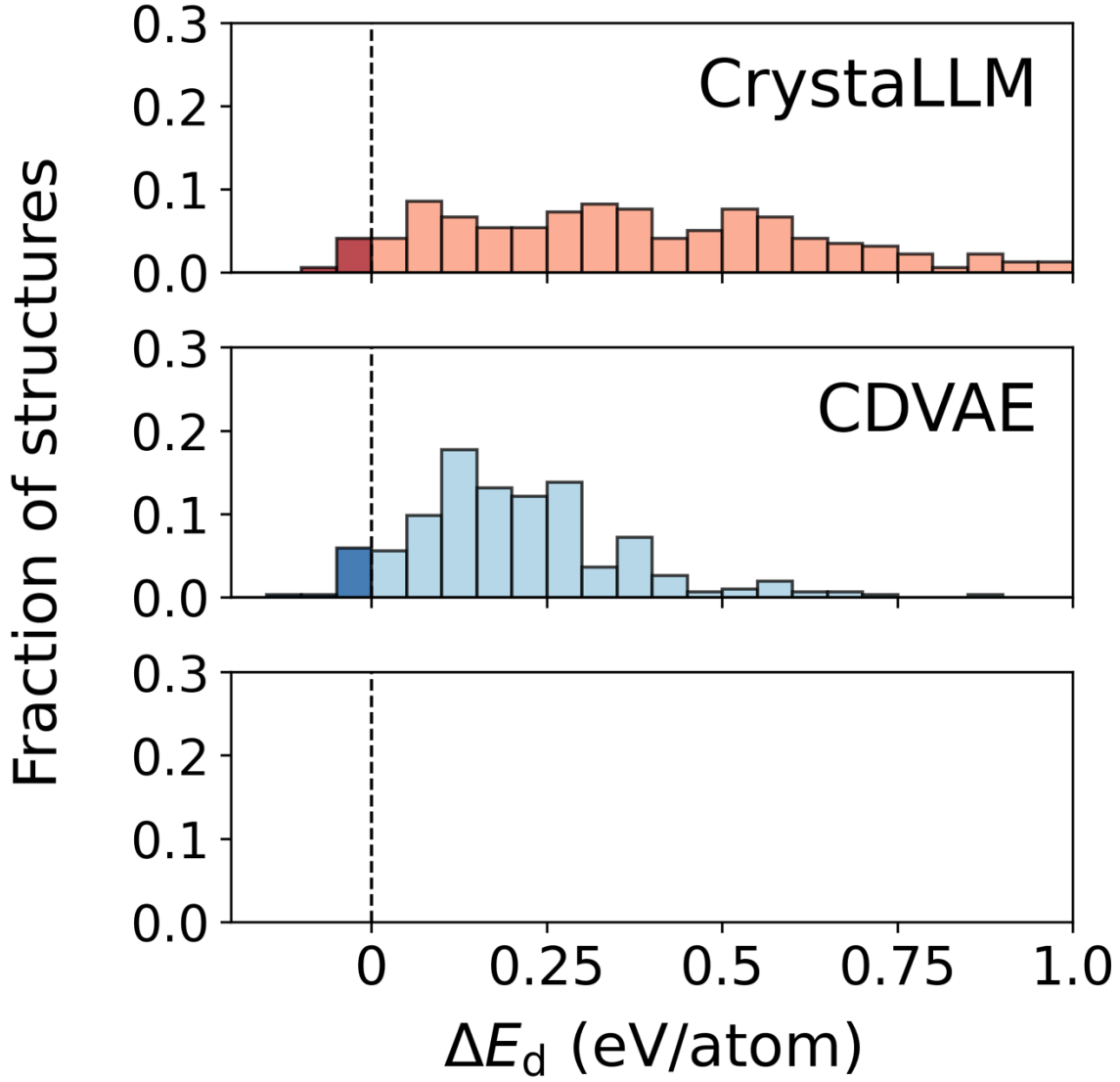


Results based on **500 materials** generated from each method

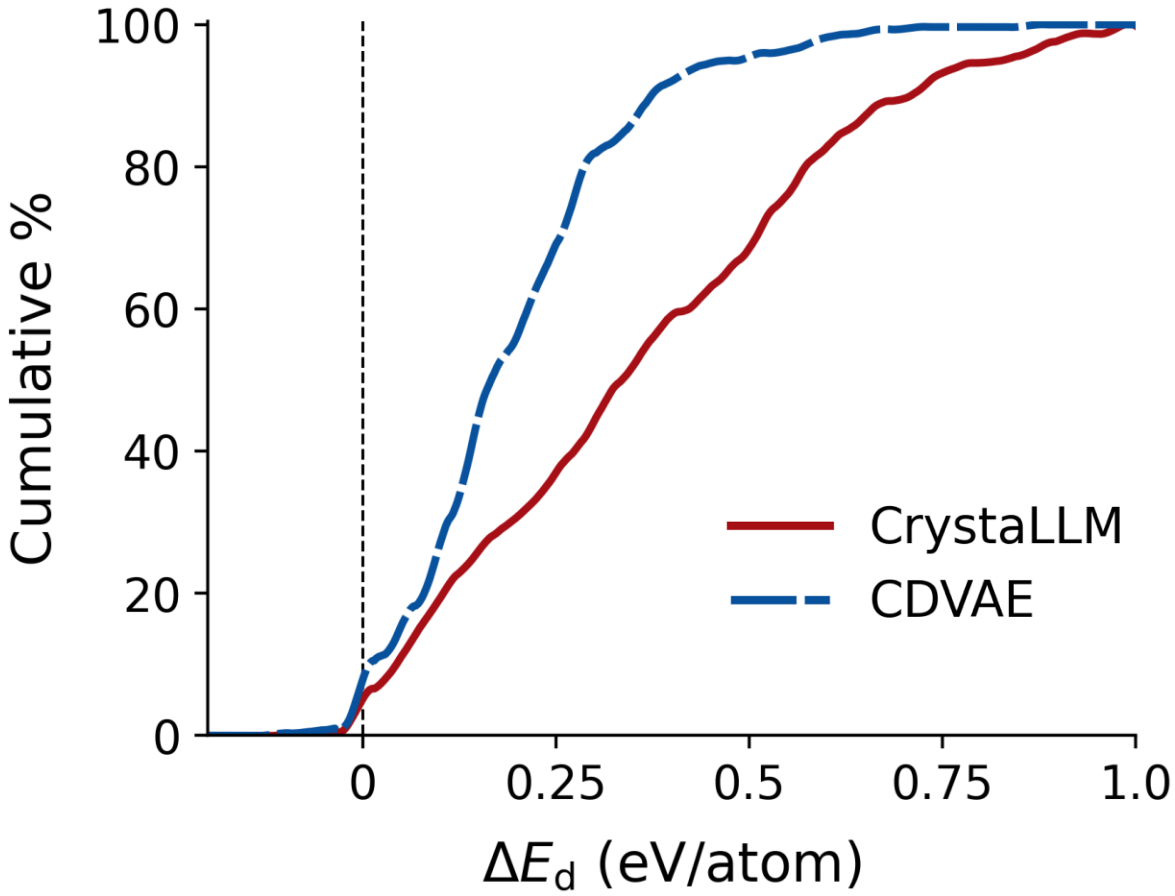


pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

Results show varied performance among distinct models

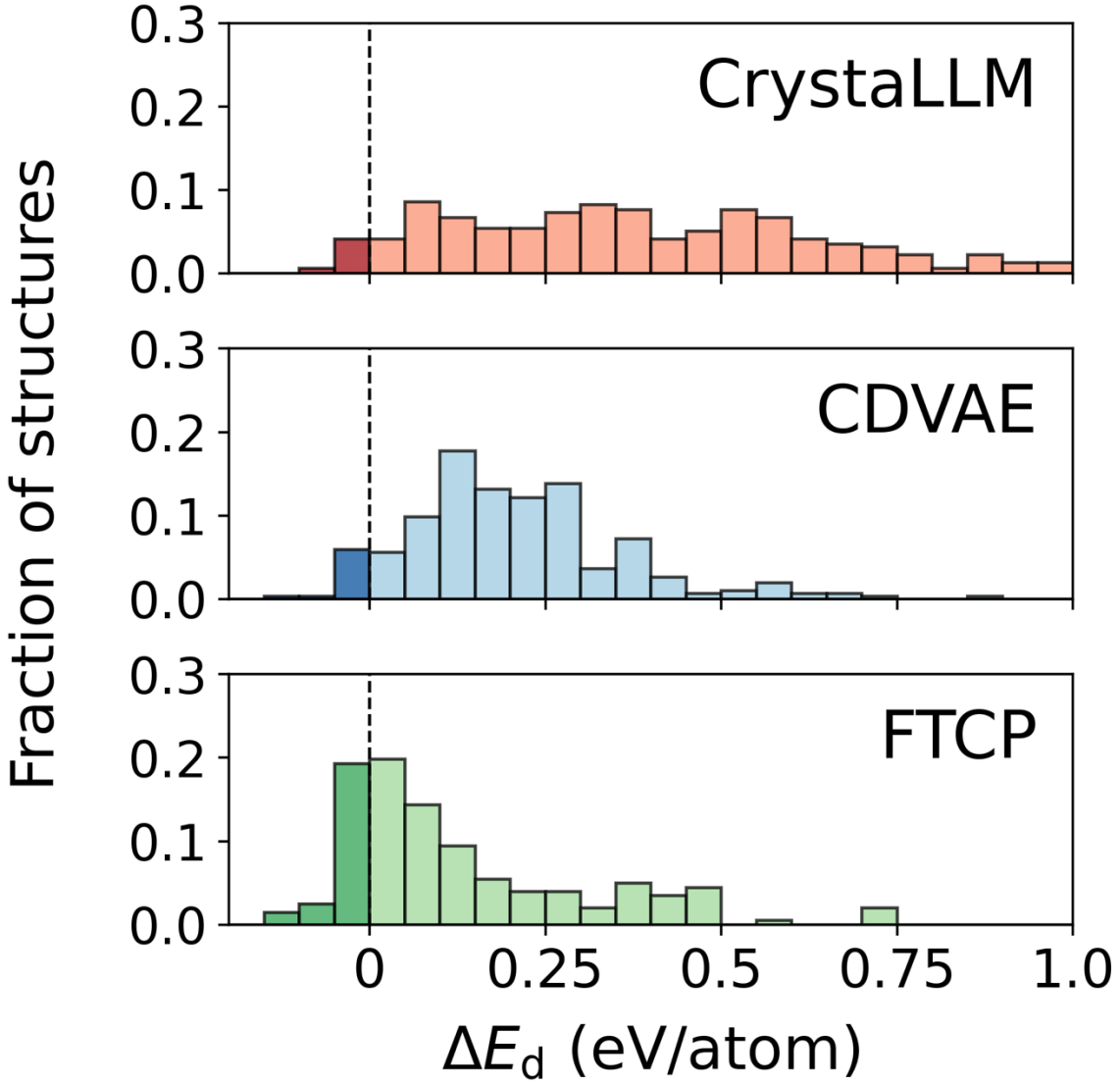


Results based on **500 materials** generated from each method

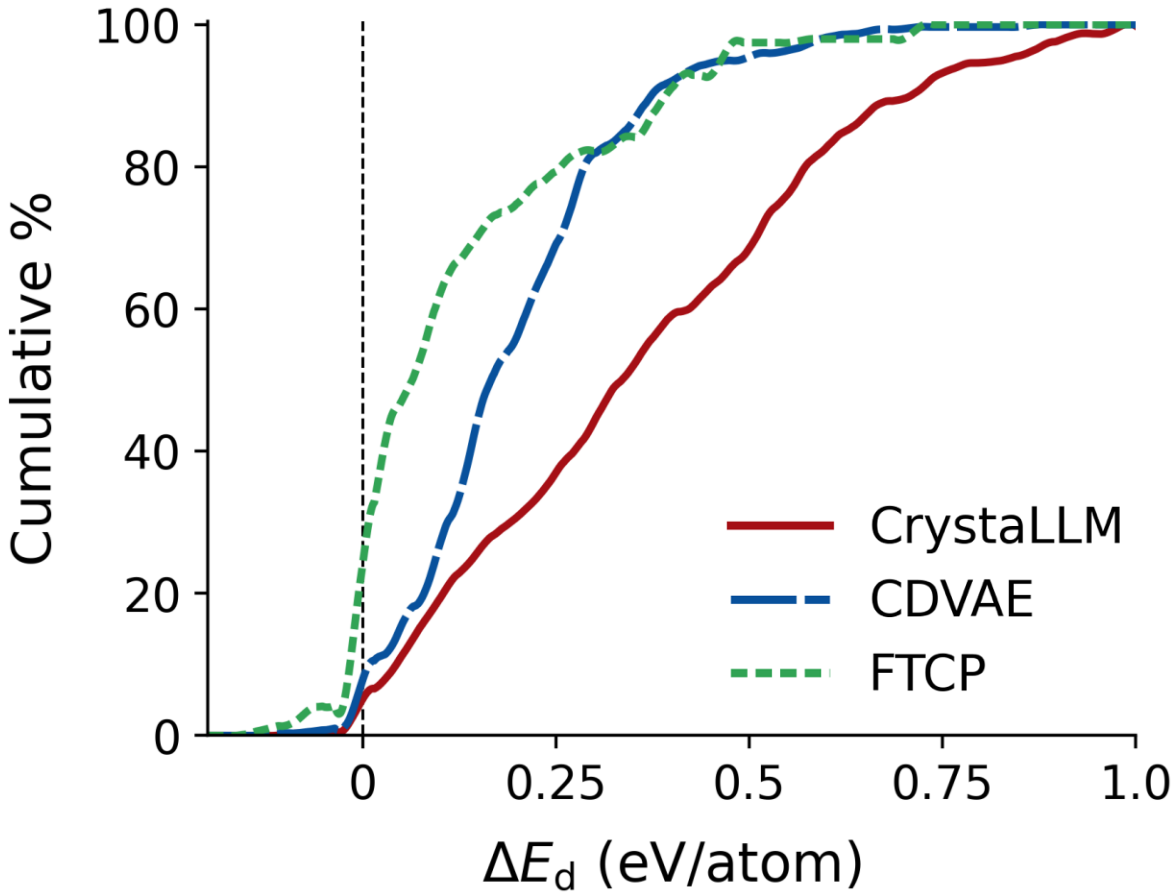


pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

Results show varied performance among distinct models



Results based on **500 materials** generated from each method



pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

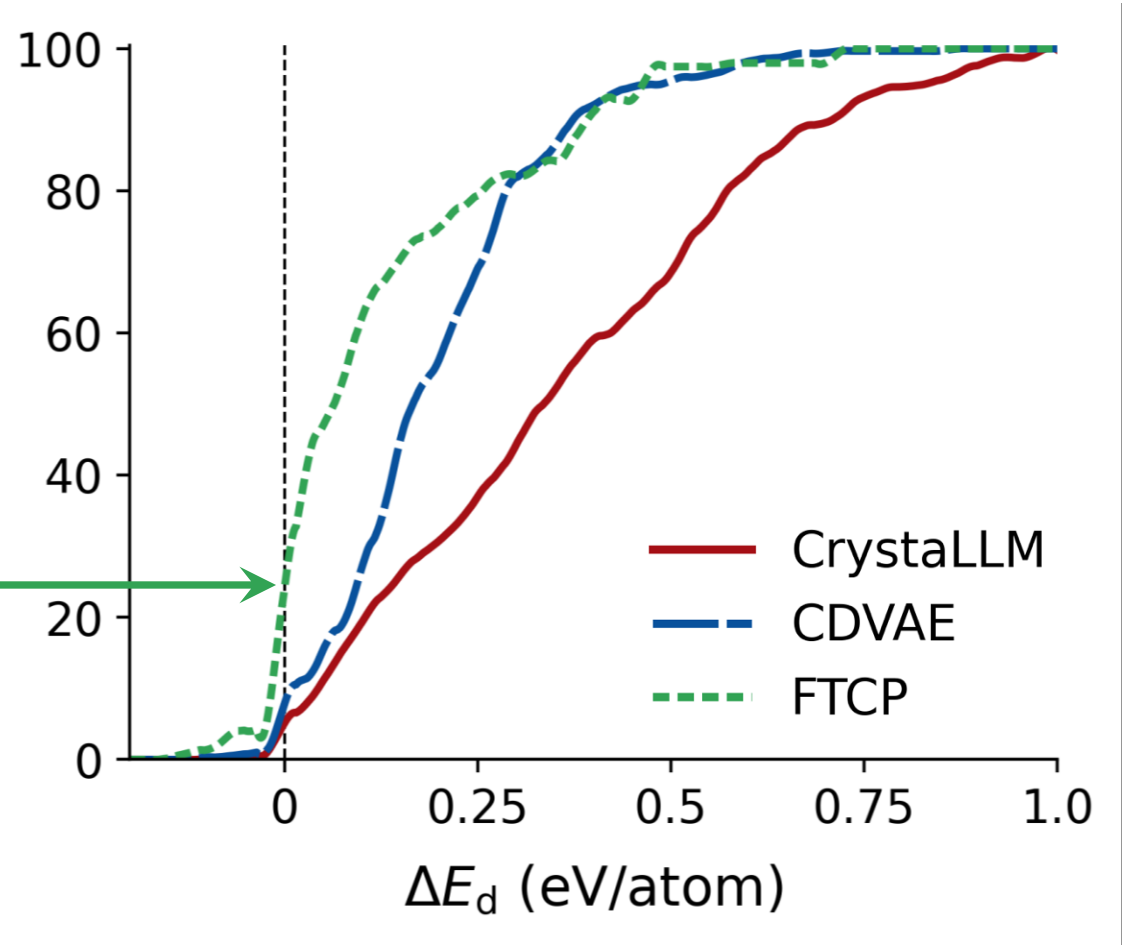
Results show varied performance among distinct models

CEMS
Chemical Engineering
& Materials Science

Is this good? Better than traditional approaches?

We need baselines!

Even the best method has a stability rate of only 26.8%



pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

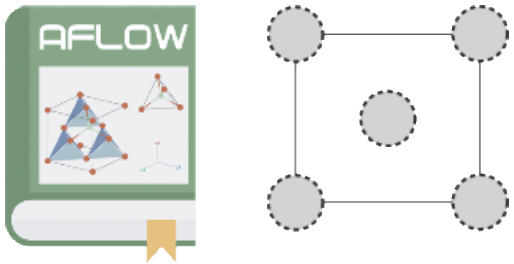
Template-based methods are used as baselines

CEMS Chemical Engineering & Materials Science

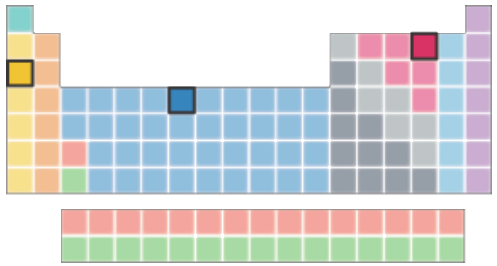
New materials are often designed ***through analogy*** to known materials

EX: $Li_7La_3\mathbf{Ti_2}O_{12}$ (LLTO) → $Li_7La_3\mathbf{Zr_2}O_{12}$ (LLZO)

Known structures



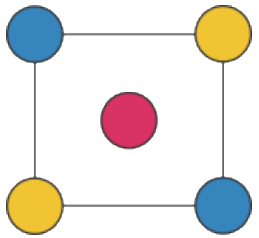
New compositions



Random selection

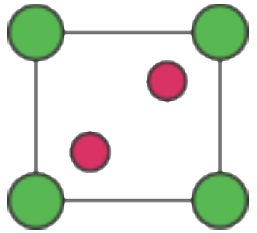
Ensure charge balance

e.g., $A^+B^{3+}O_3^{2-}$

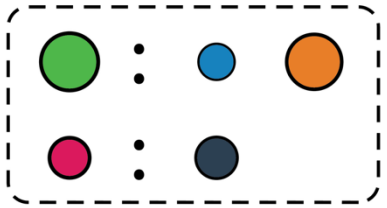


Choose known materials ***that are stable***

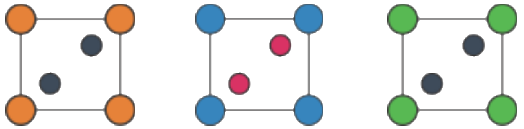
MATERIALS PROJECT



Then perform **data-driven ion exchange**

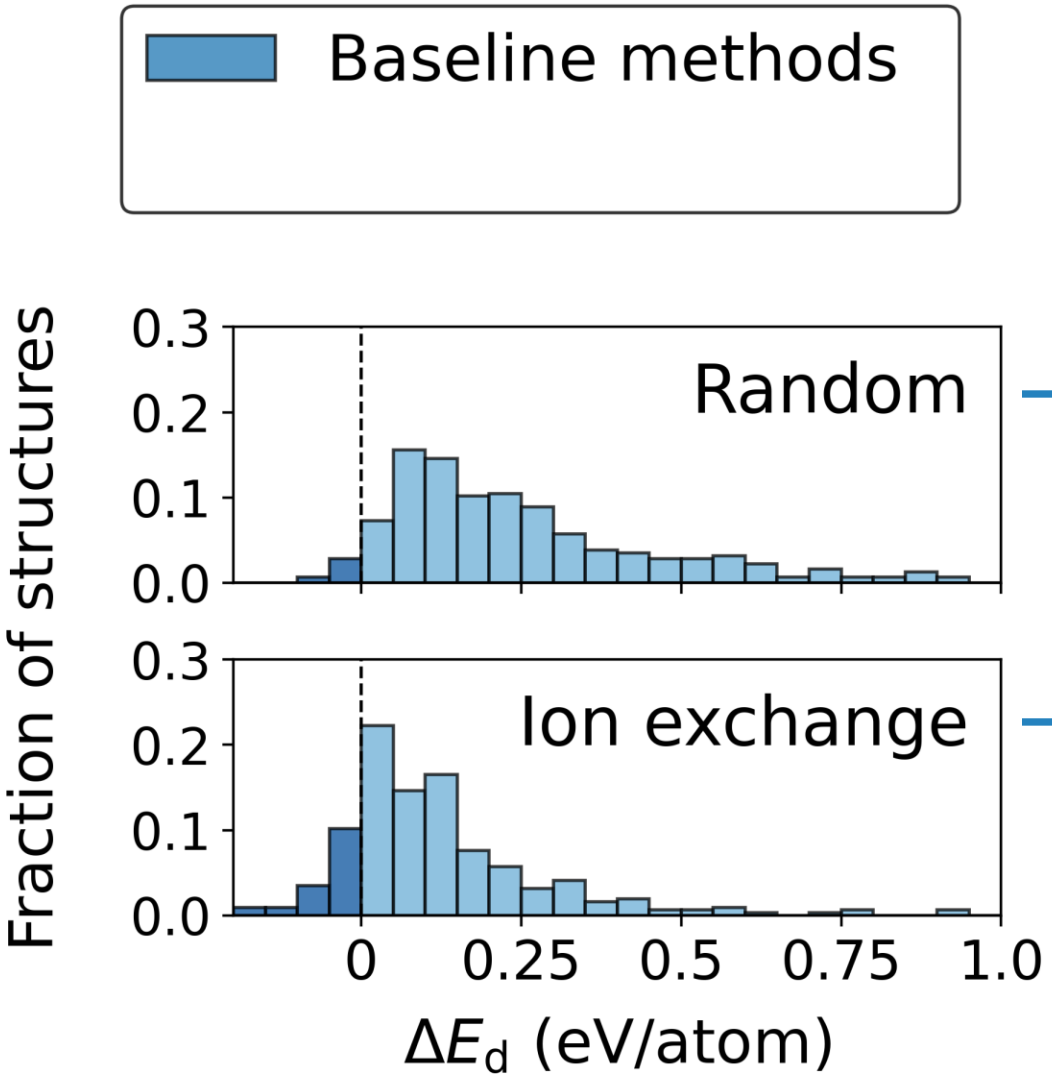


e.g., Li^+ for Na^+



pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

Results show that baseline methods perform comparably well



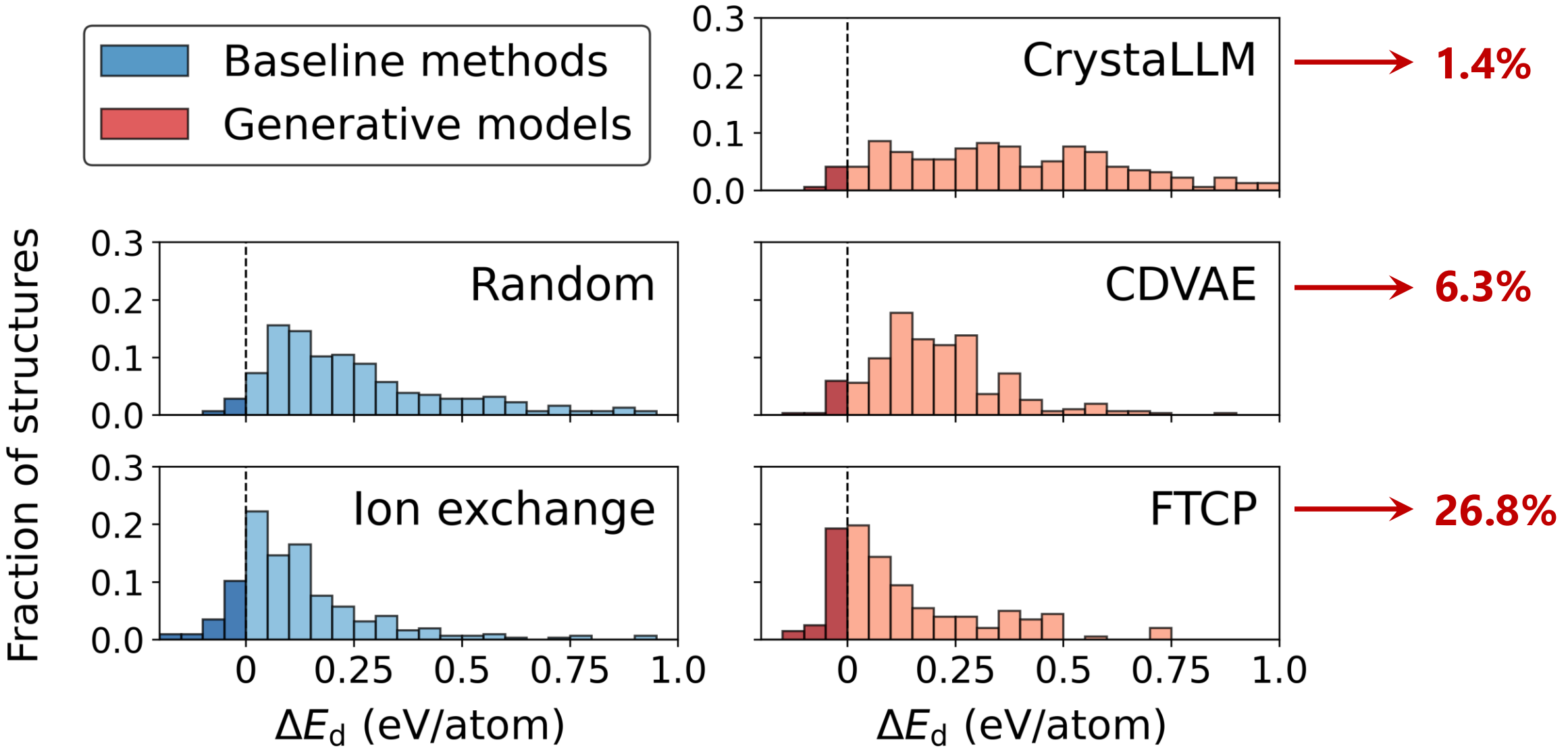
Results based on **500 materials** generated from each method

Stability rate of 4.3%

Stability rate of 19.1%

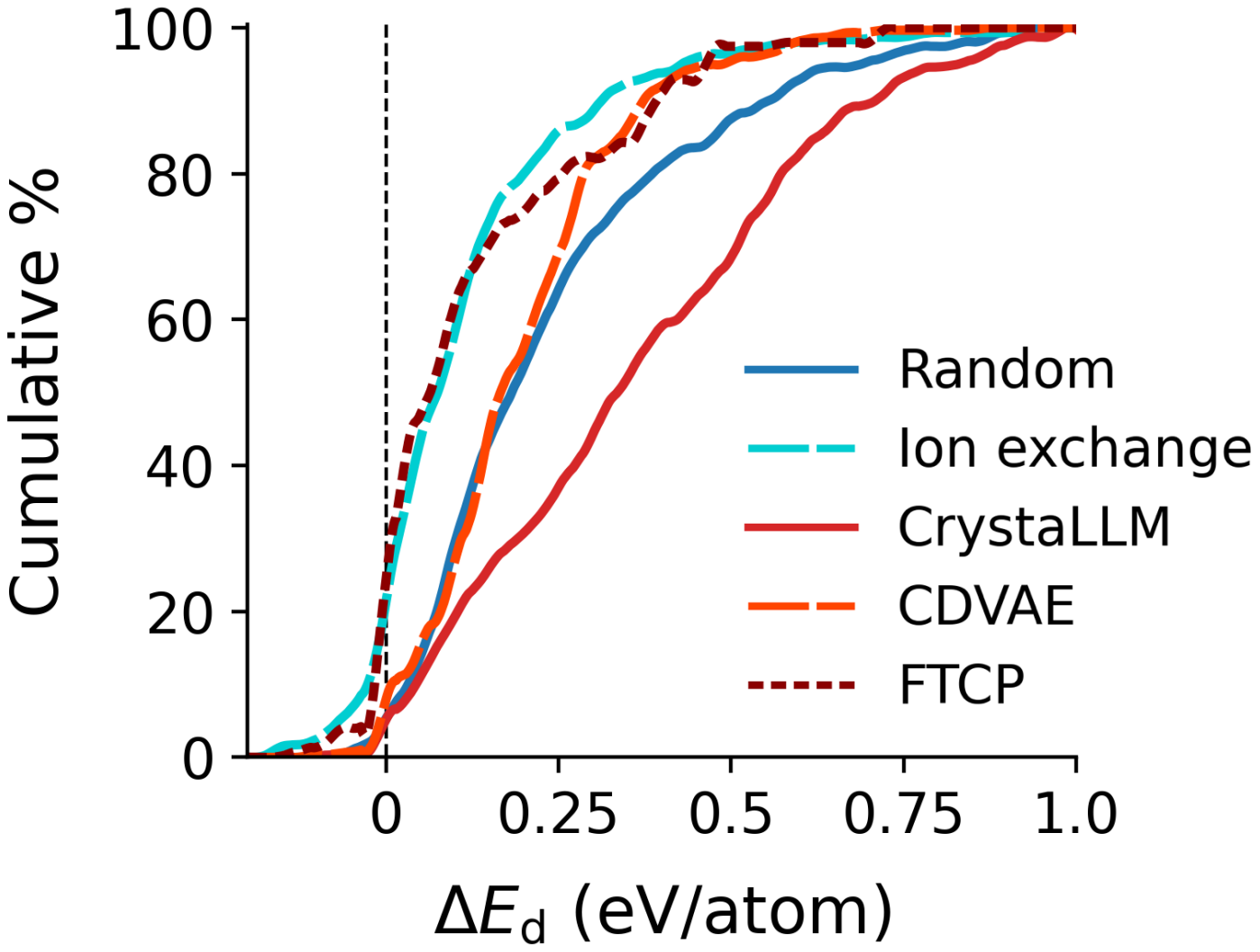
pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

Results show that baseline methods perform comparably well



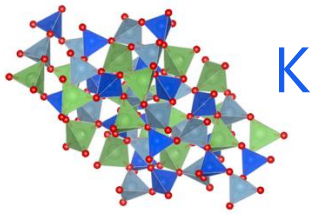
pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

Results show that baseline methods perform comparably well



In terms of stability, **generative models don't have a clear advantage** over more traditional methods…

What about **novelty**?



Known?

pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

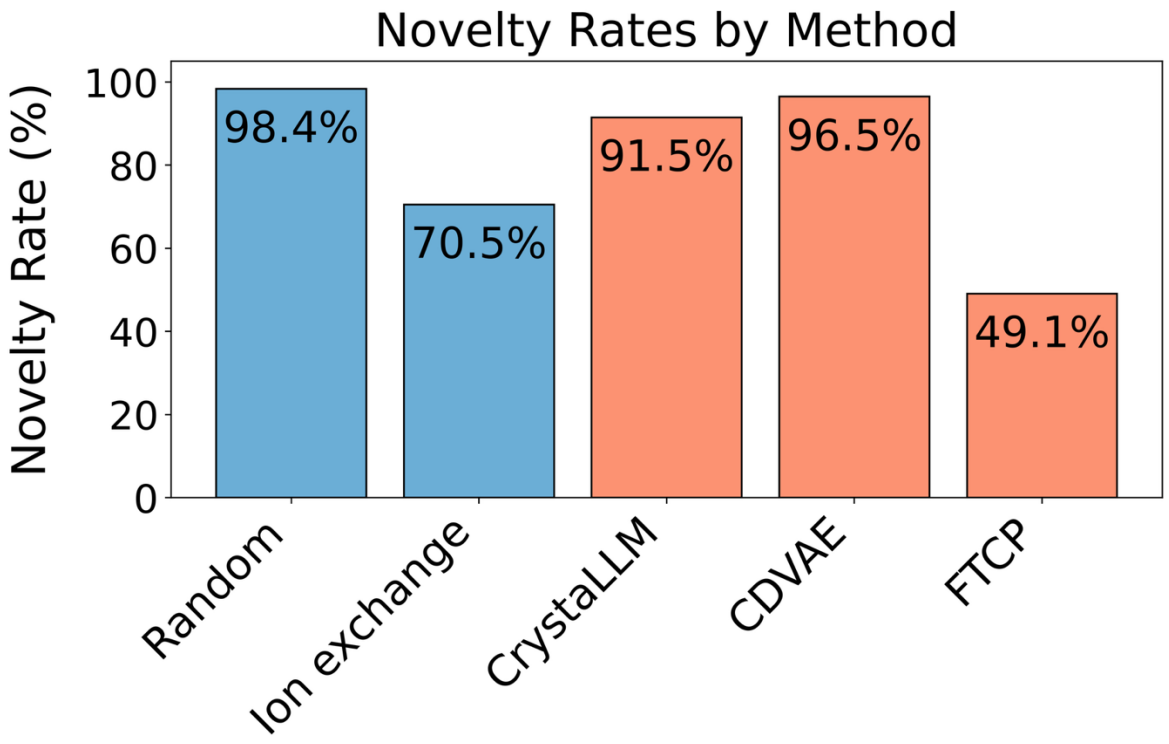
Assessing novelty by comparing with open databases

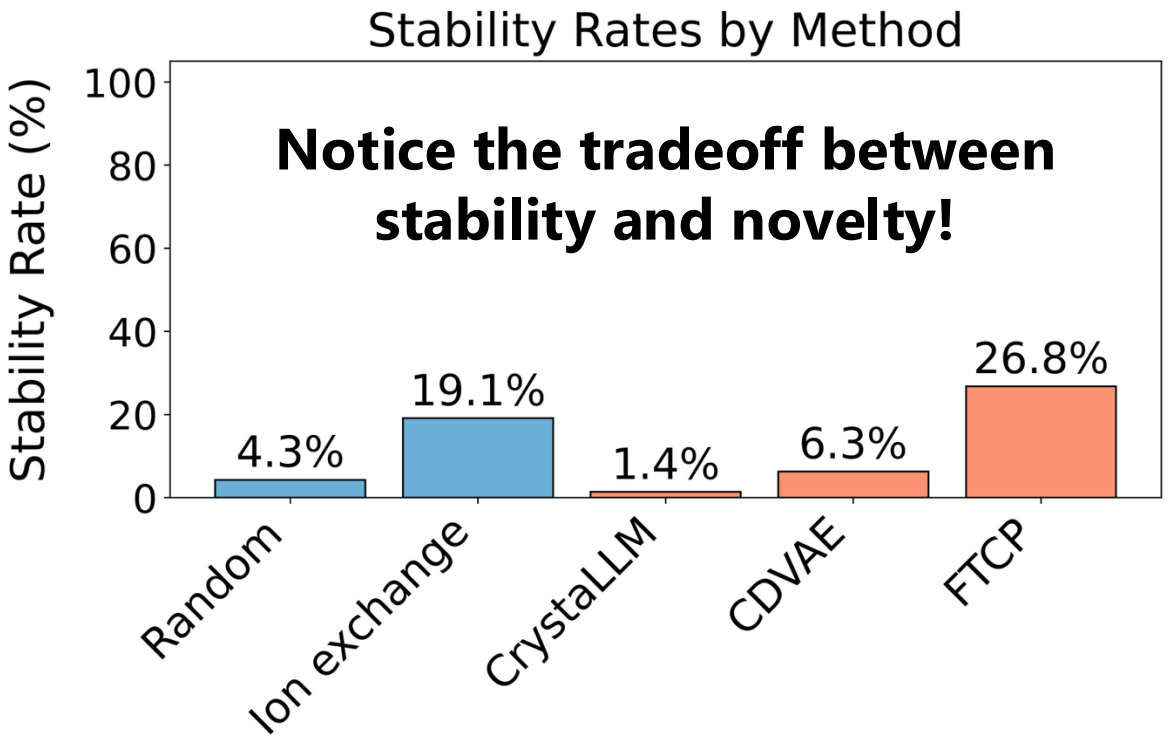
CEMS Chemical Engineering & Materials Science

Definition 1: *Novel* materials → being outside of the training data

MATERIALS PROJECT

A material is "novel" if it does **not match any of the known 200k materials in MP**





pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

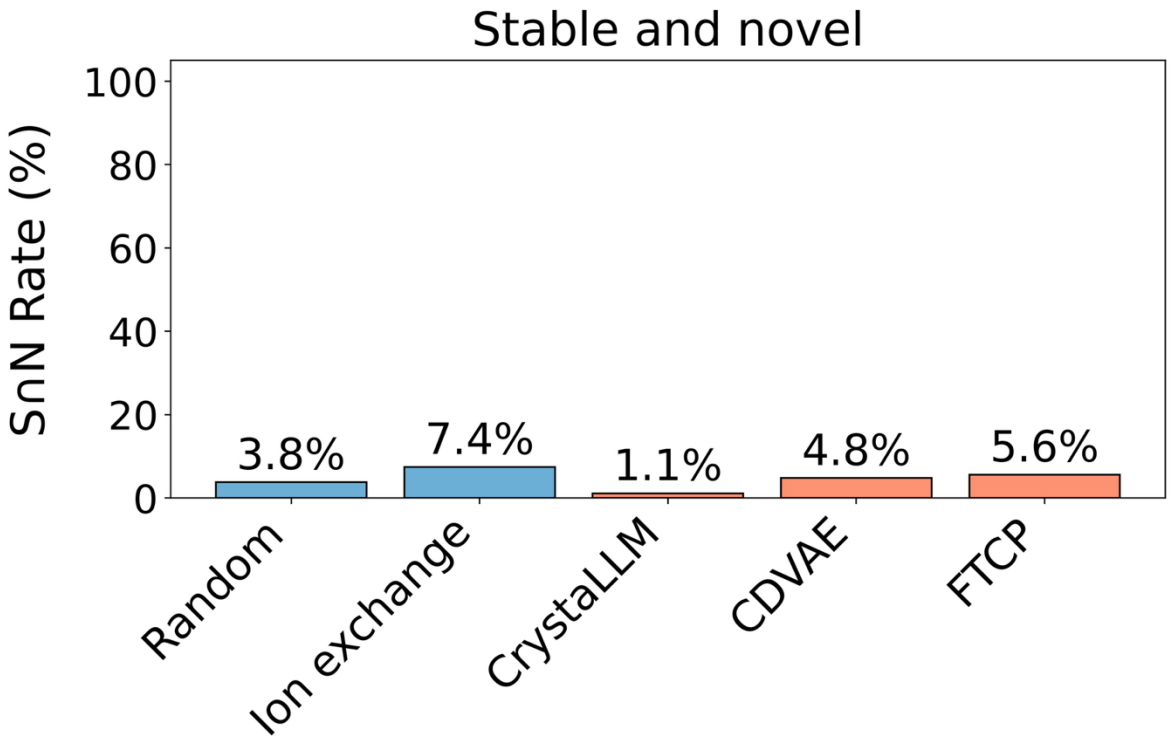
Assessing novelty by comparing with open databases

CEMS Chemical Engineering & Materials Science

Definition 1: *Novel* materials → being outside of the training data

MATERIALS PROJECT

A material is "novel" if it does **not match any of the known 200k materials in MP**



Ion exchange produces the most new materials that are stable

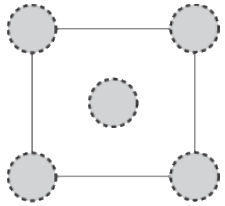
...**Outperforming AI!**

pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

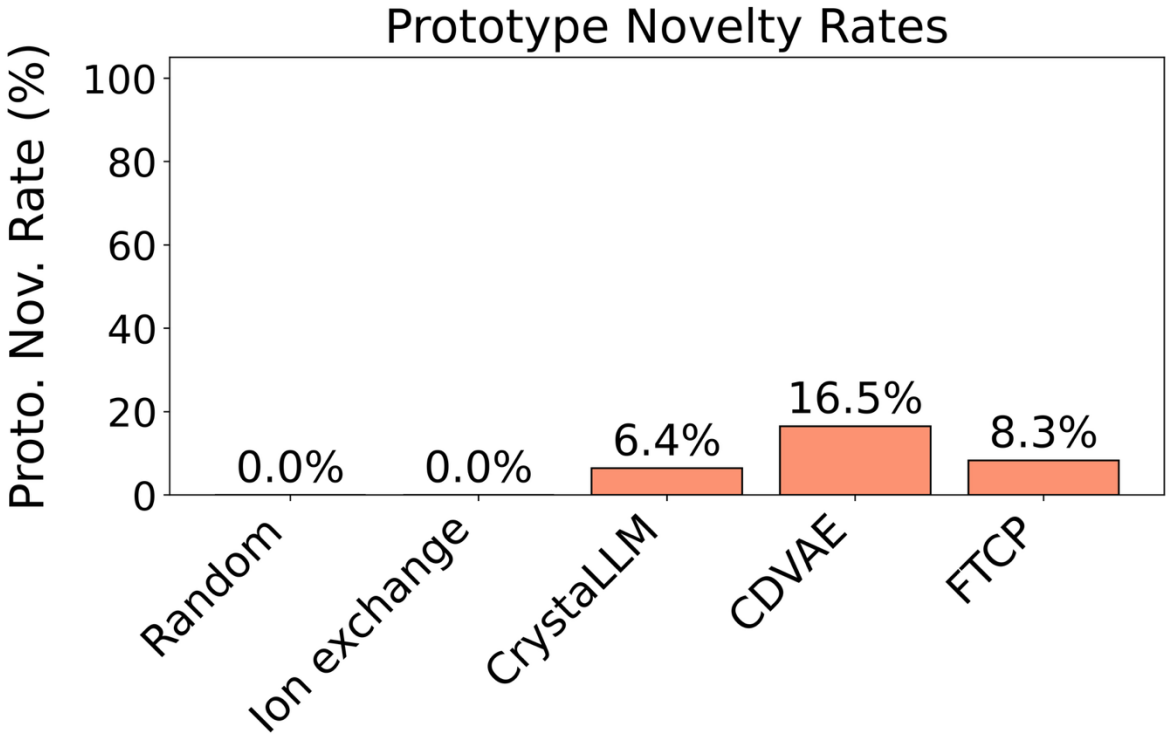
Assessing novelty by comparing with open databases

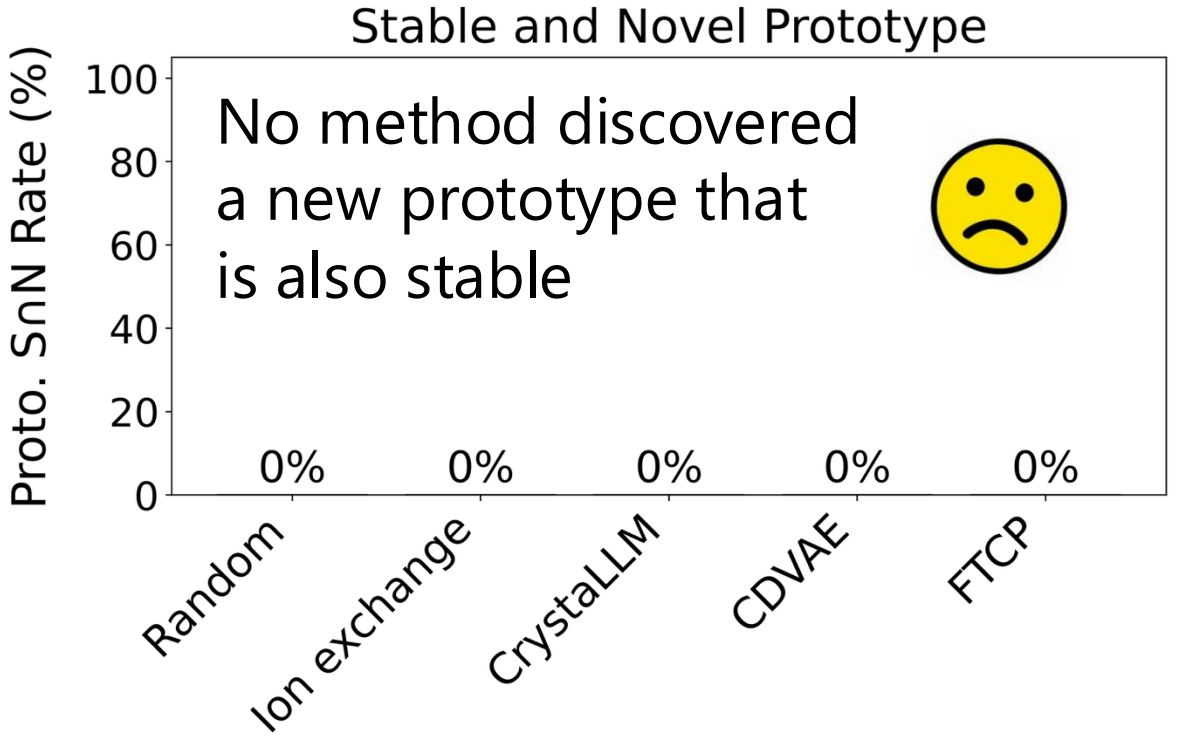
CEMS Chemical Engineering & Materials Science

Definition 2: *Prototype novelty* → an entirely new crystal structure



A material is "prototype-novel" if its **structure cannot be mapped to a known prototype in AFLOW**

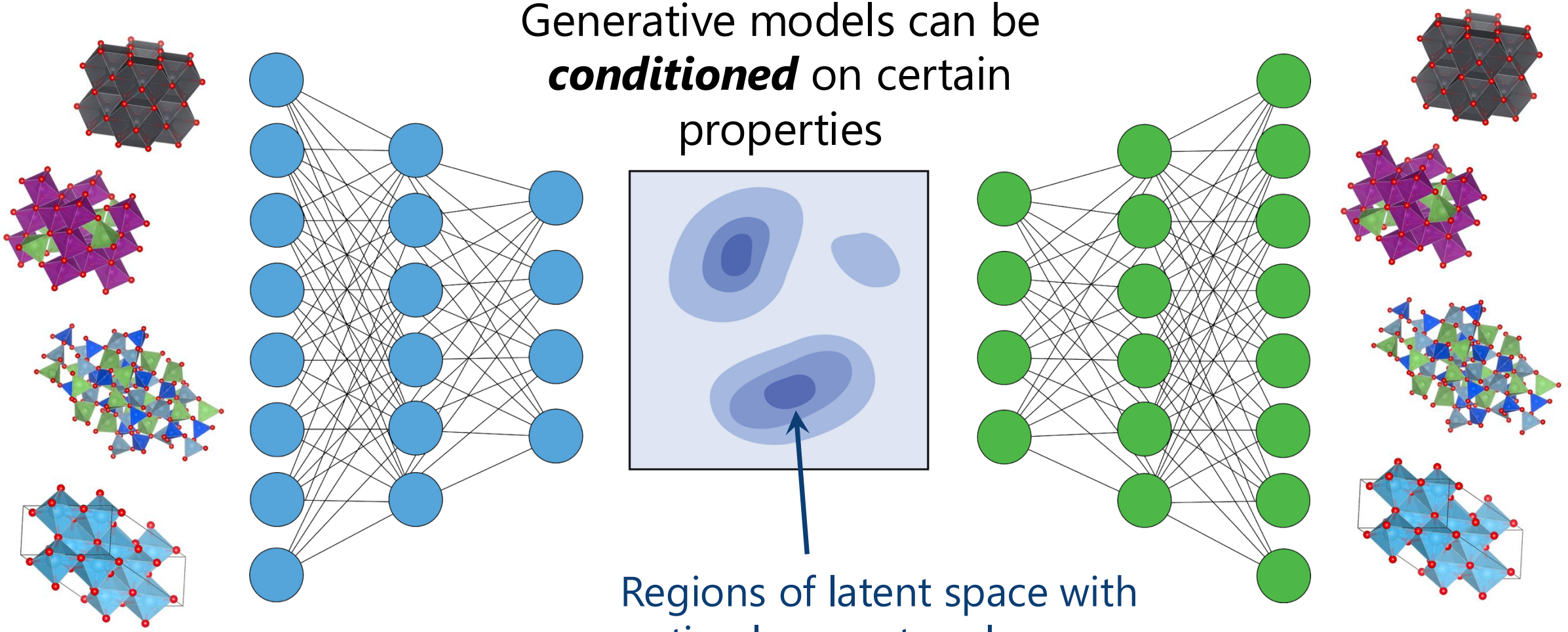




pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

Beyond stability and novelty: what about functionality?

Generative models can be ***conditioned*** on certain properties

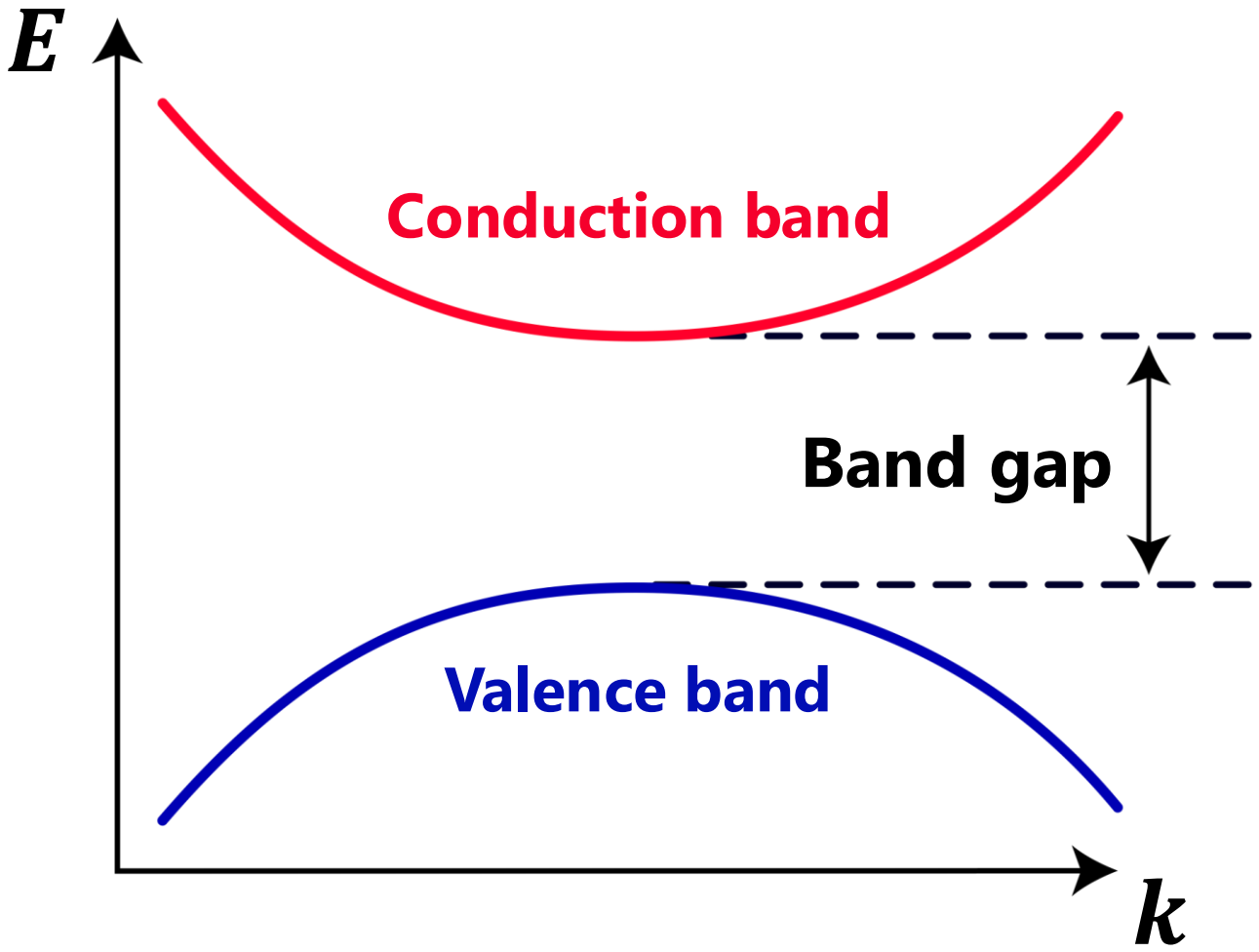


Regions of latent space with optimal property values

pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

We evaluated two properties: band gap and bulk modulus

Target #1: band gap of 3 eV

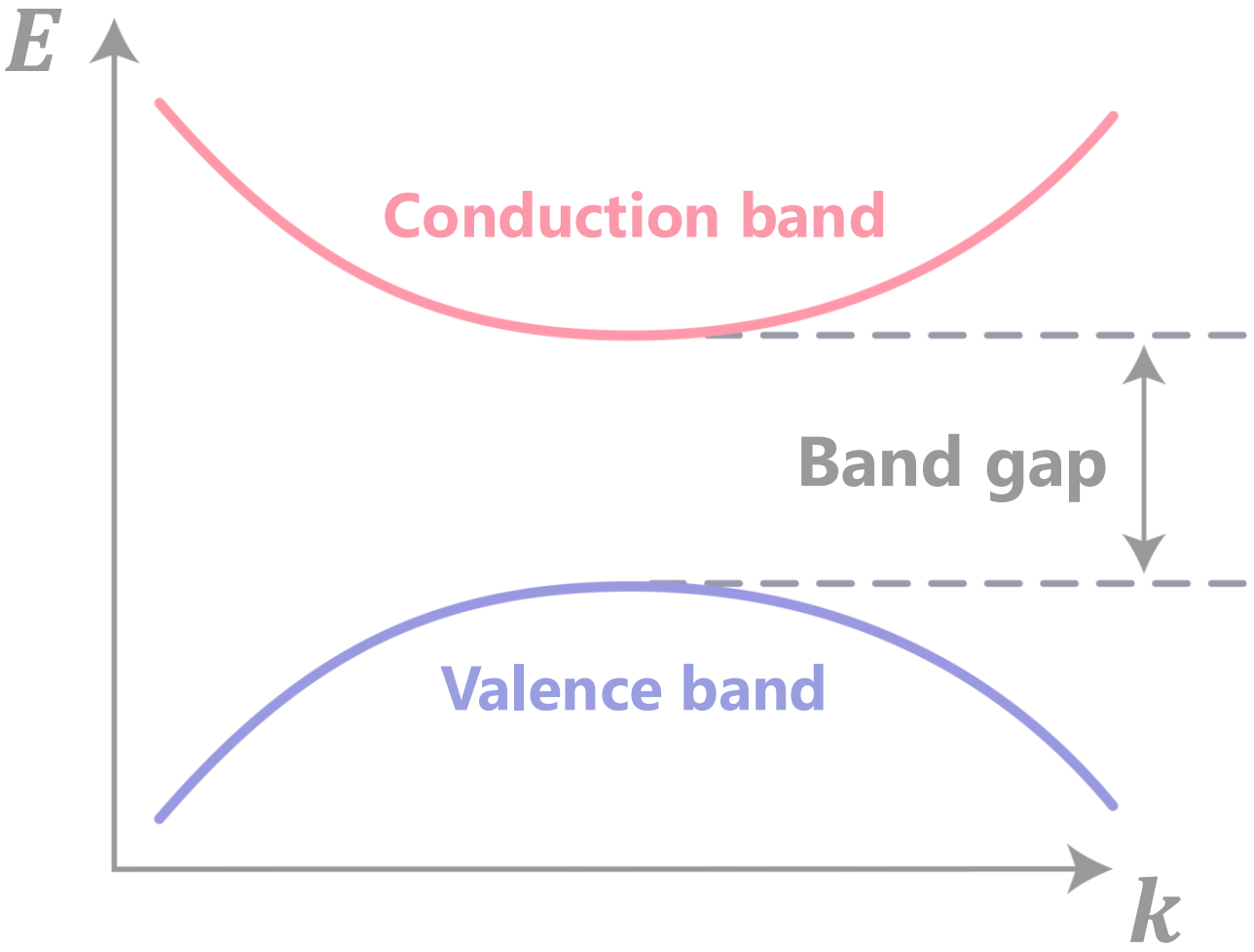


pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

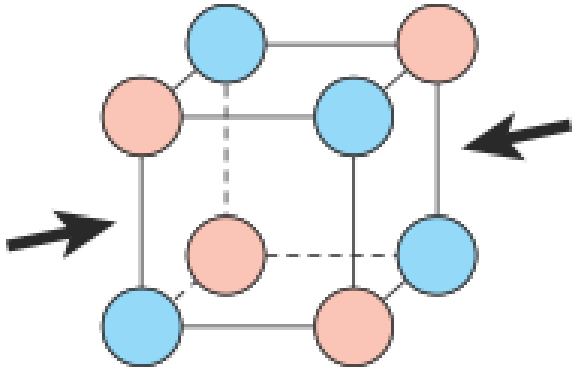
We evaluated two properties: band gap and bulk modulus

CEMS Chemical Engineering & Materials Science

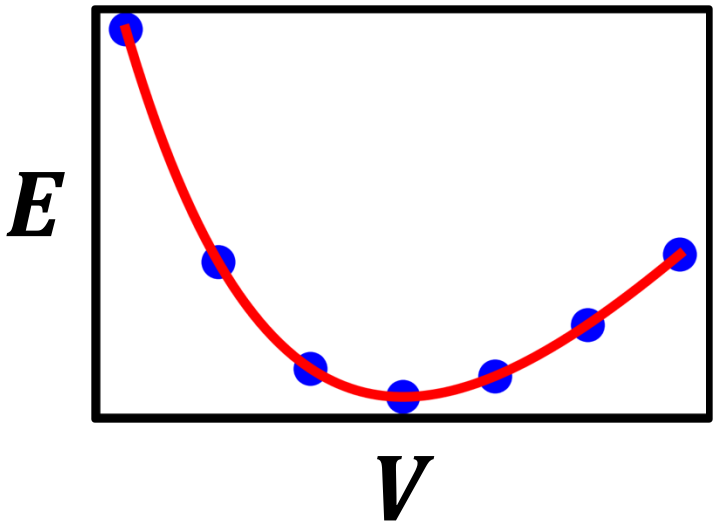
Target #1: band gap of 3 eV



Target #2: bulk modulus > 300 GPa

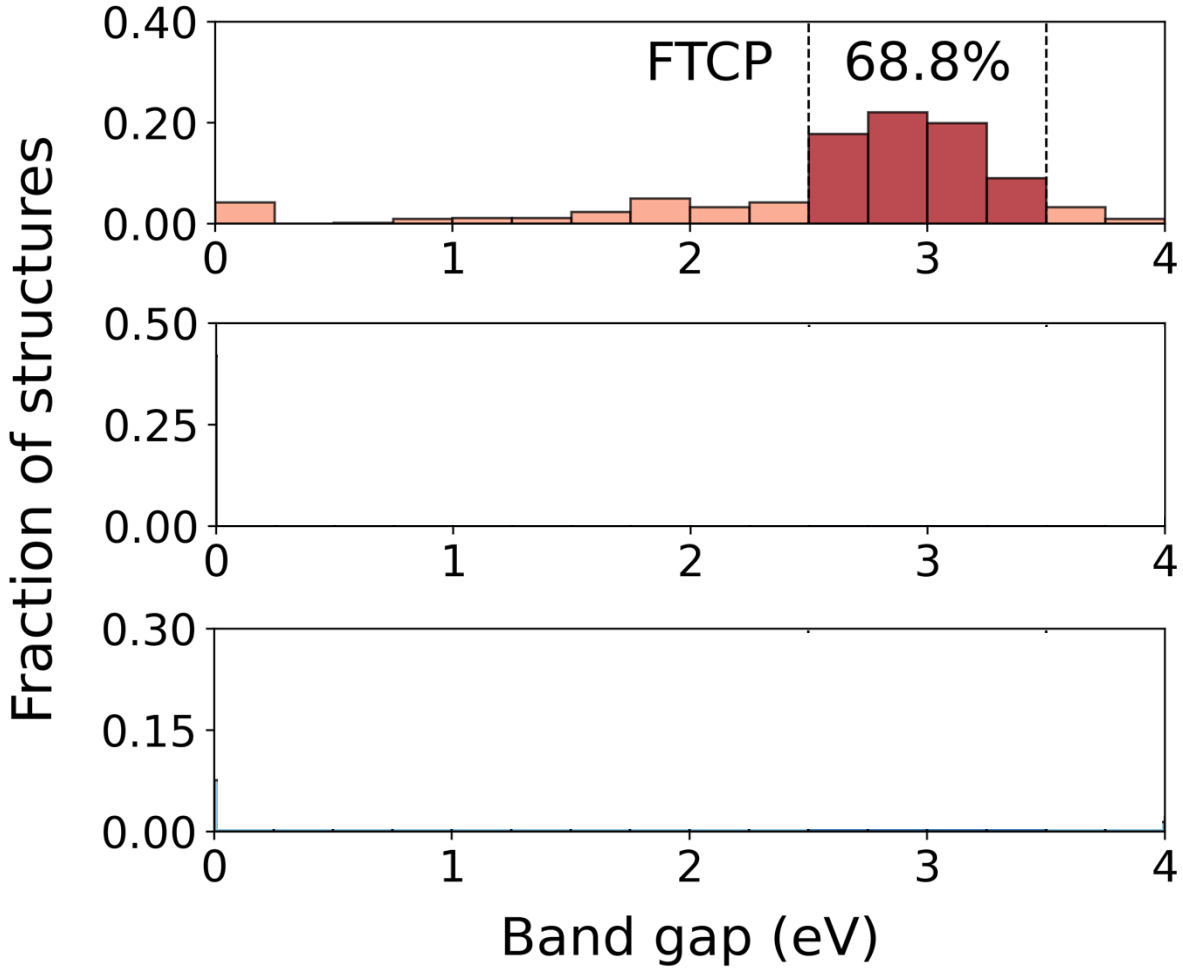


Large bulk modulus means highly resistant to compression



pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

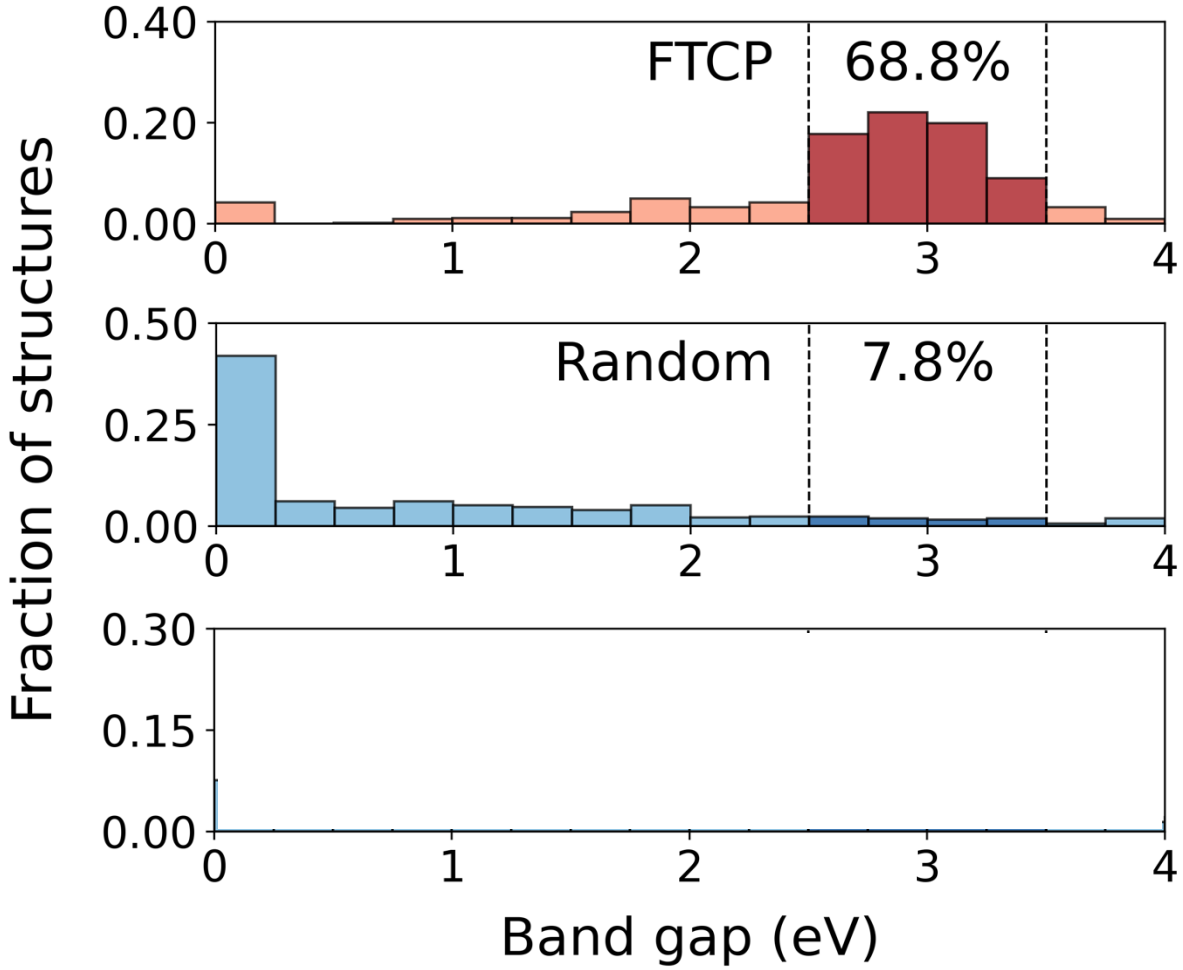
Results show that generative AI excels on (some) properties



68.8% of materials proposed by the **best generative model (FTCP)** exhibit a band gap within 0.5 eV of the target

pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

Results show that generative AI excels on (some) properties

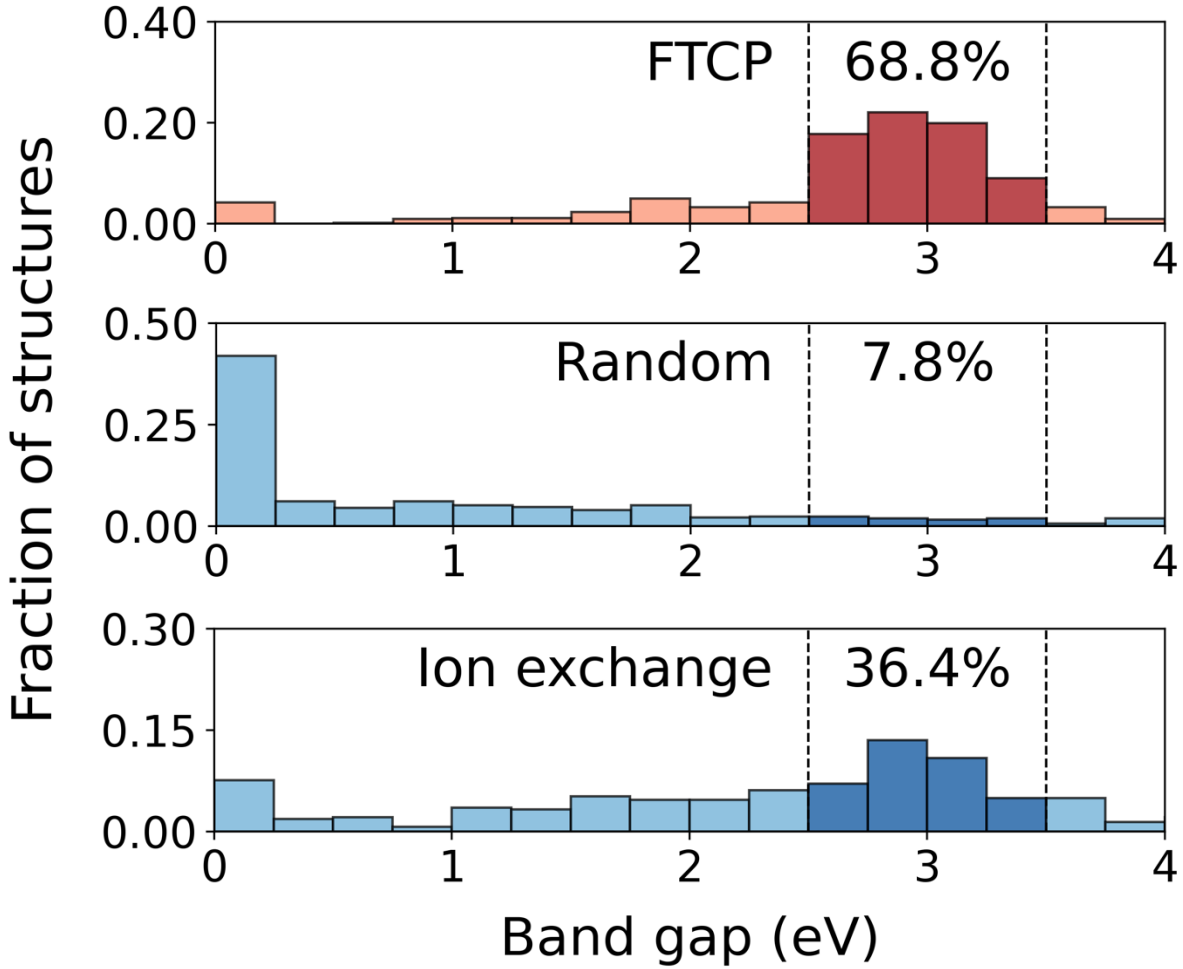


68.8% of materials proposed by the **best generative model (FTCP)** exhibit a band gap within 0.5 eV of the target

Random sampling is much worse: only 7.8% of materials satisfy the same constraint

pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

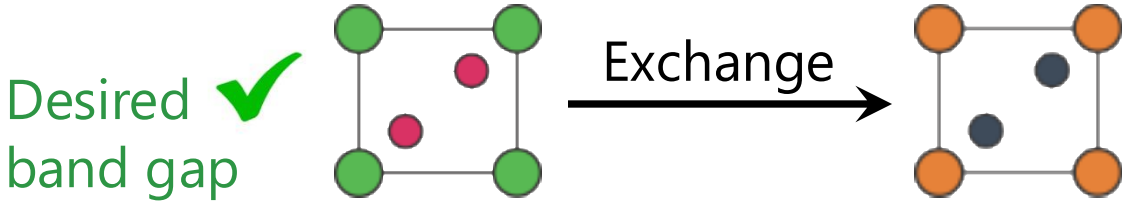
Results show that generative AI excels on (some) properties



68.8% of materials proposed by the **best generative model (FTCP)** exhibit a band gap within 0.5 eV of the target

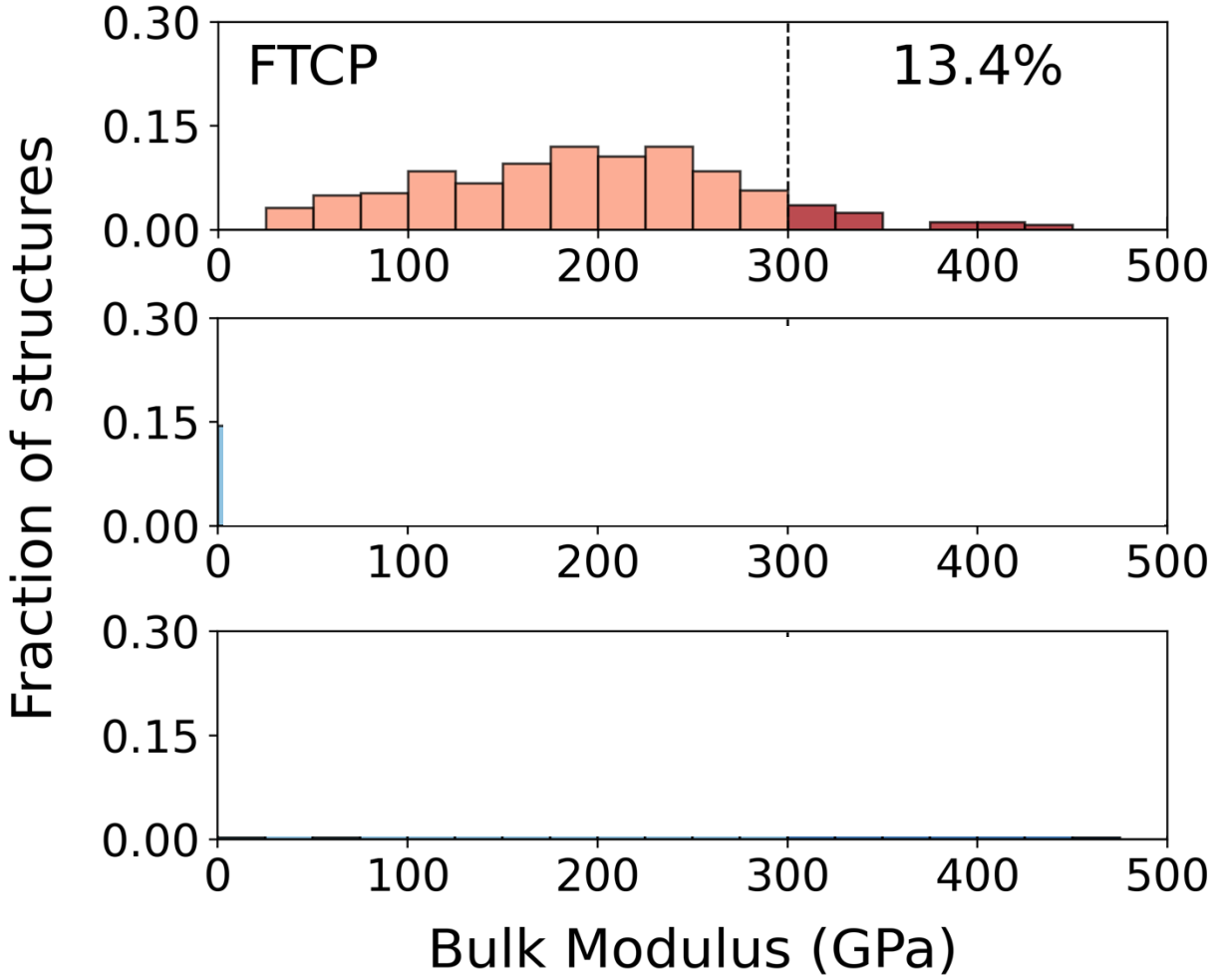
Random sampling is much worse: only 7.8% of materials satisfy the same constraint

Ion exchange works okay (36.4%) but is not nearly as effective as the generative AI (FTCP)



pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

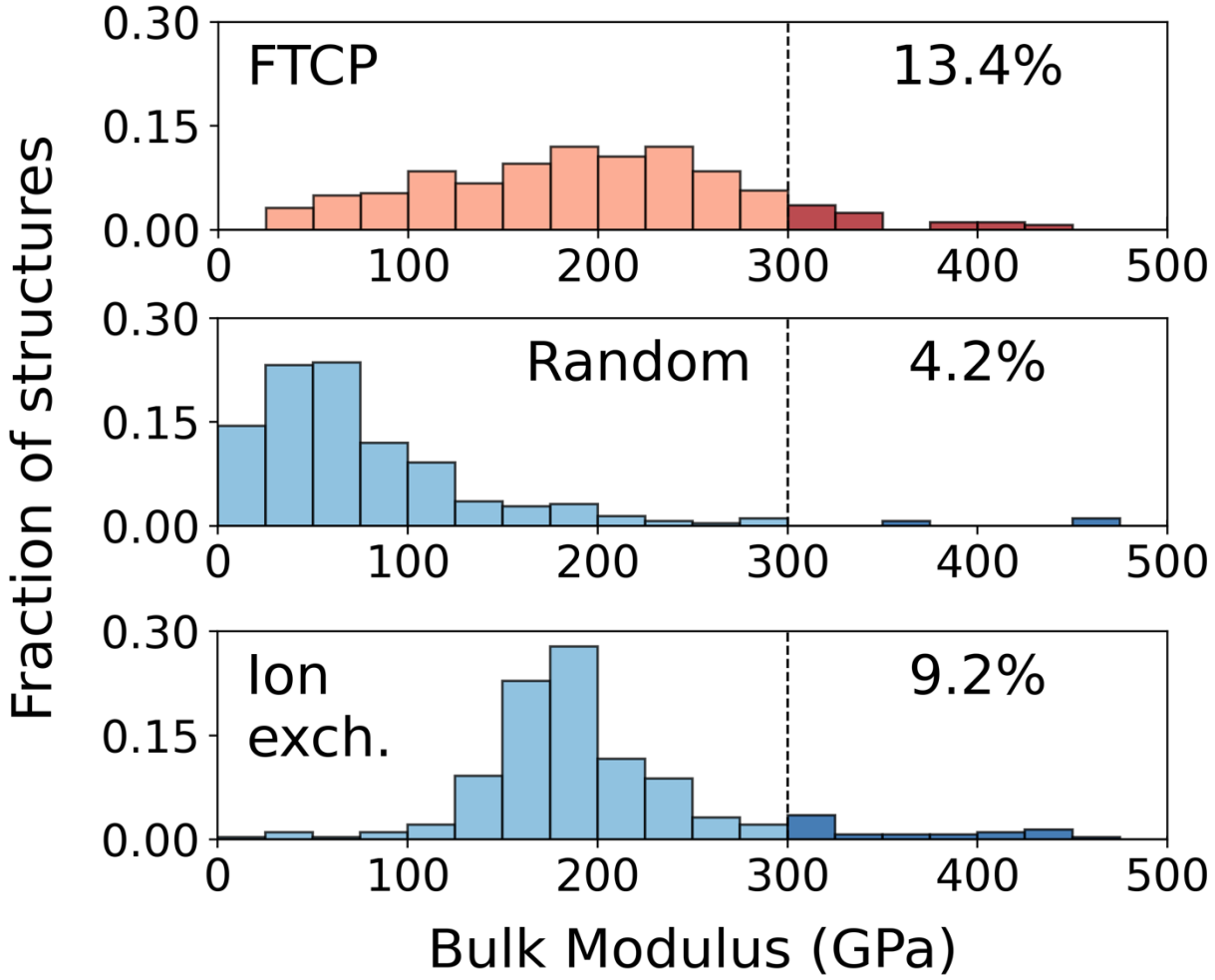
Results show that generative AI excels on (some) properties



Only 13.4% of materials proposed by the **best generative model (FTCP)** exhibit a bulk modulus > 300 GPa

pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

Results show that generative AI excels on (some) properties



Only 13.4% of materials proposed by the **best generative model (FTCP)** exhibit a bulk modulus > 300 GPa

Random sampling and ion exchange both perform even worse…

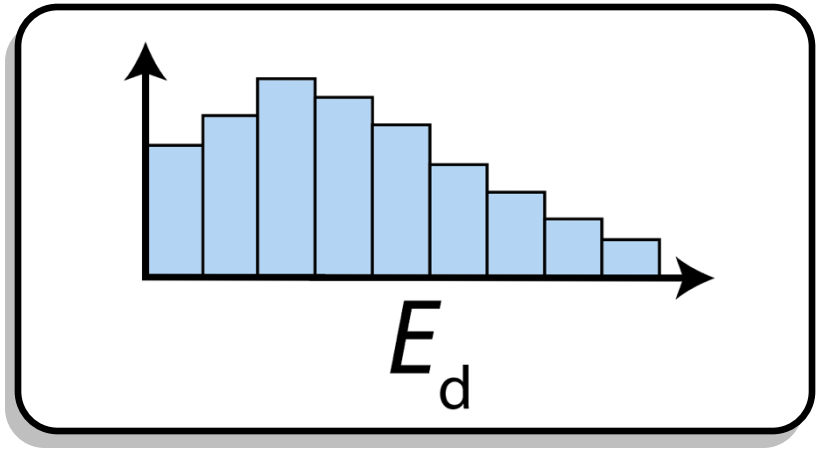
Extreme properties are difficult due to a lack of relevant data

MATERIALS PROJECT

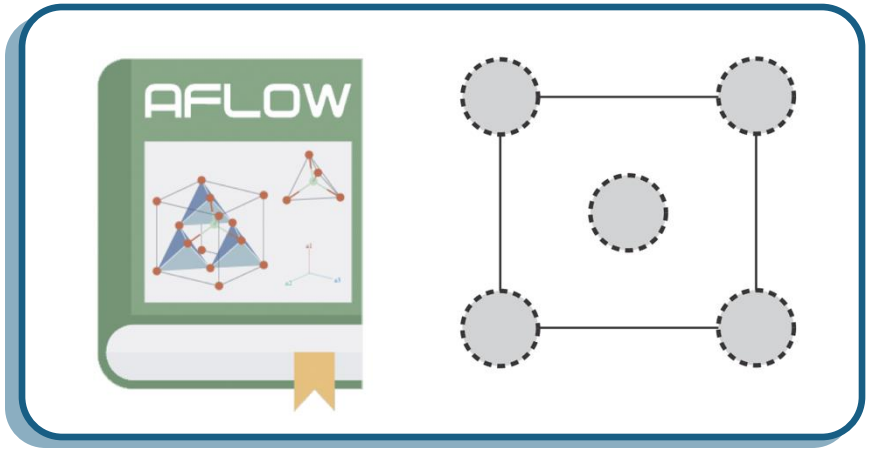
< 1% materials with $B > 300$ GPa

pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

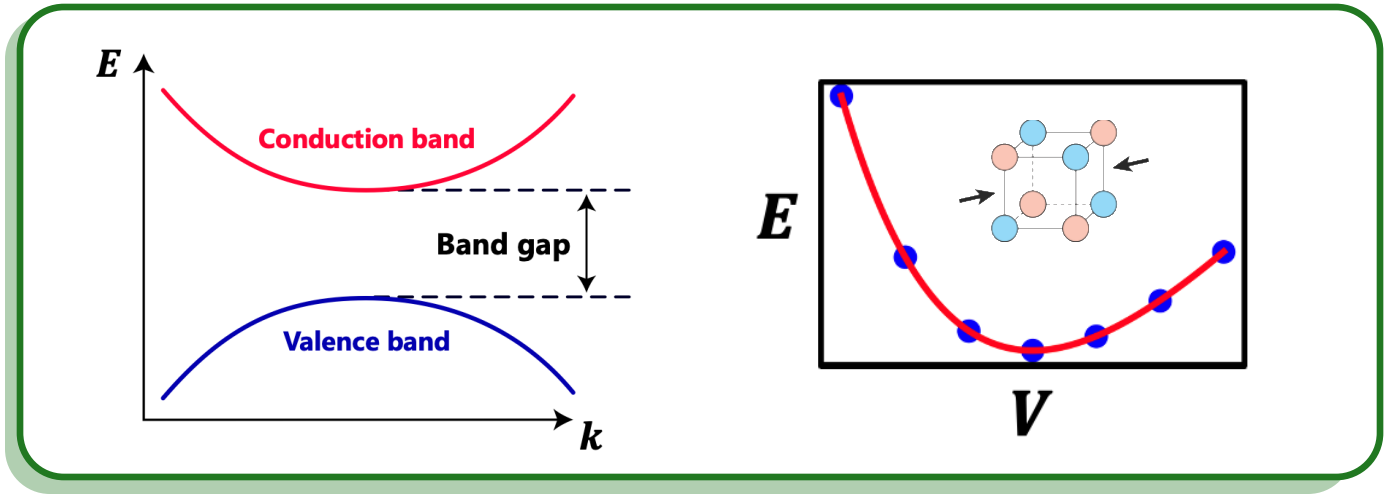
Balancing stability, novelty, and targeted properties



Thermodynamic stability



Structural novelty

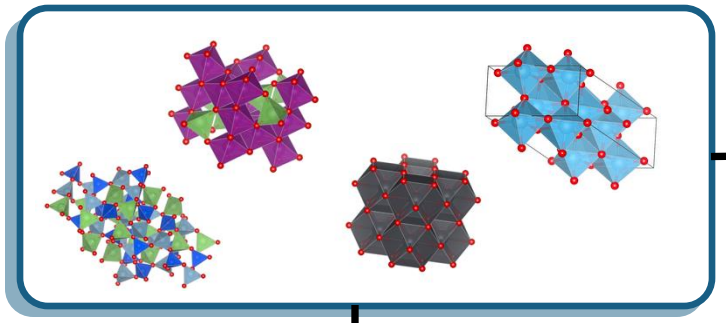


Properties to achieve desired functionality

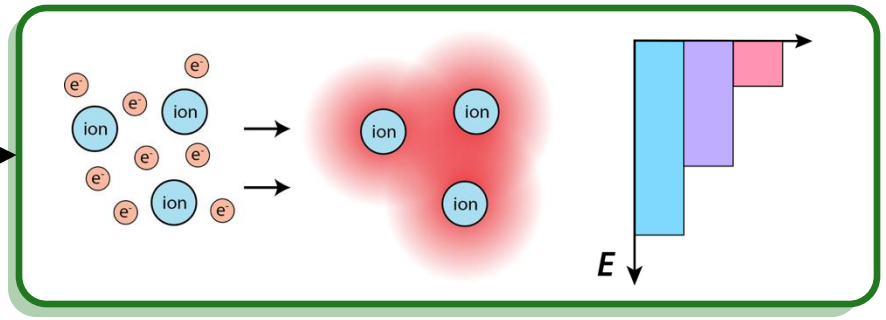
Can we achieve a balance through ML-based filtration?

CEMS Chemical Engineering & Materials Science

Materials generation

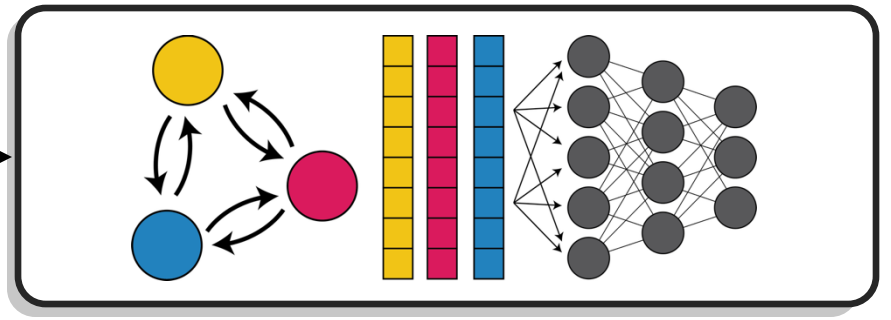


DFT calculations



Can take hours-to-days for each material

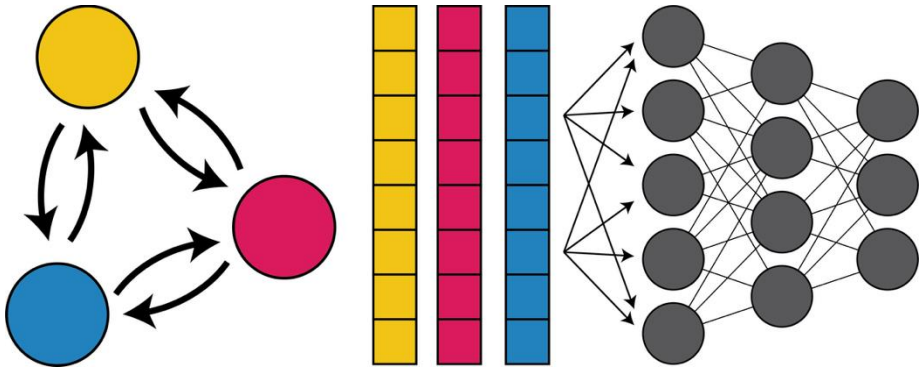
ML predictions



Once trained, models make predictions in a matter of seconds

Graph neural networks are fast and (usually) reliable

Crystal structure

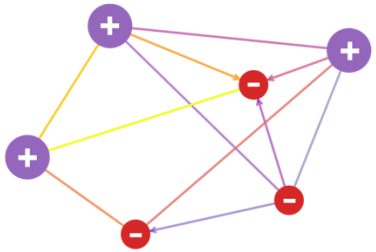


Energy, properties

SevenNet

CGCNN

CHGNet

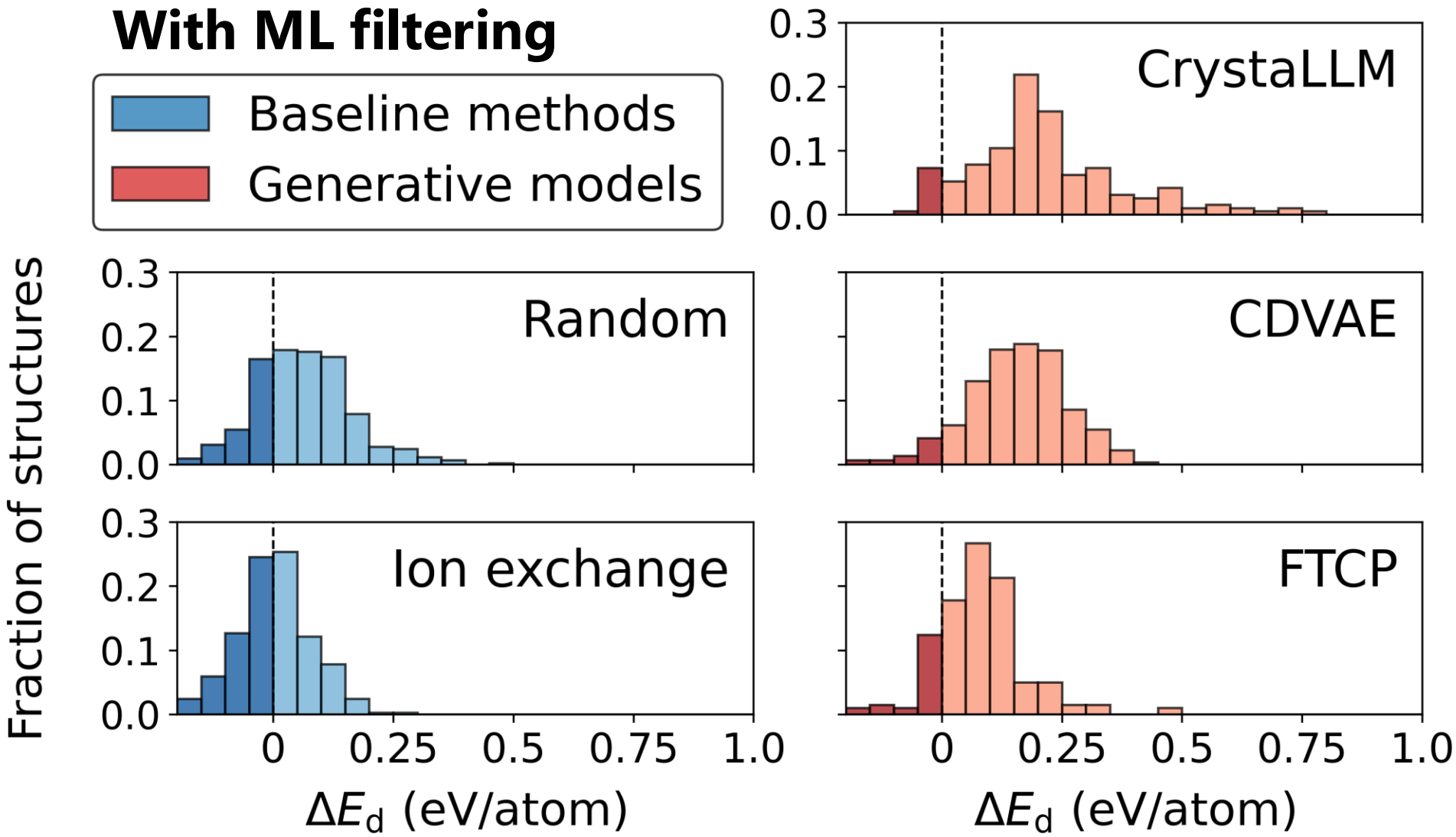


Orbital Materials

Some mean absolute errors:

- **Energy:** 18 meV/atom
- **Band gap:** 0.38 eV
- **Bulk modulus:** 24 GPa

ML filtering leads to improved stability rates for all methods



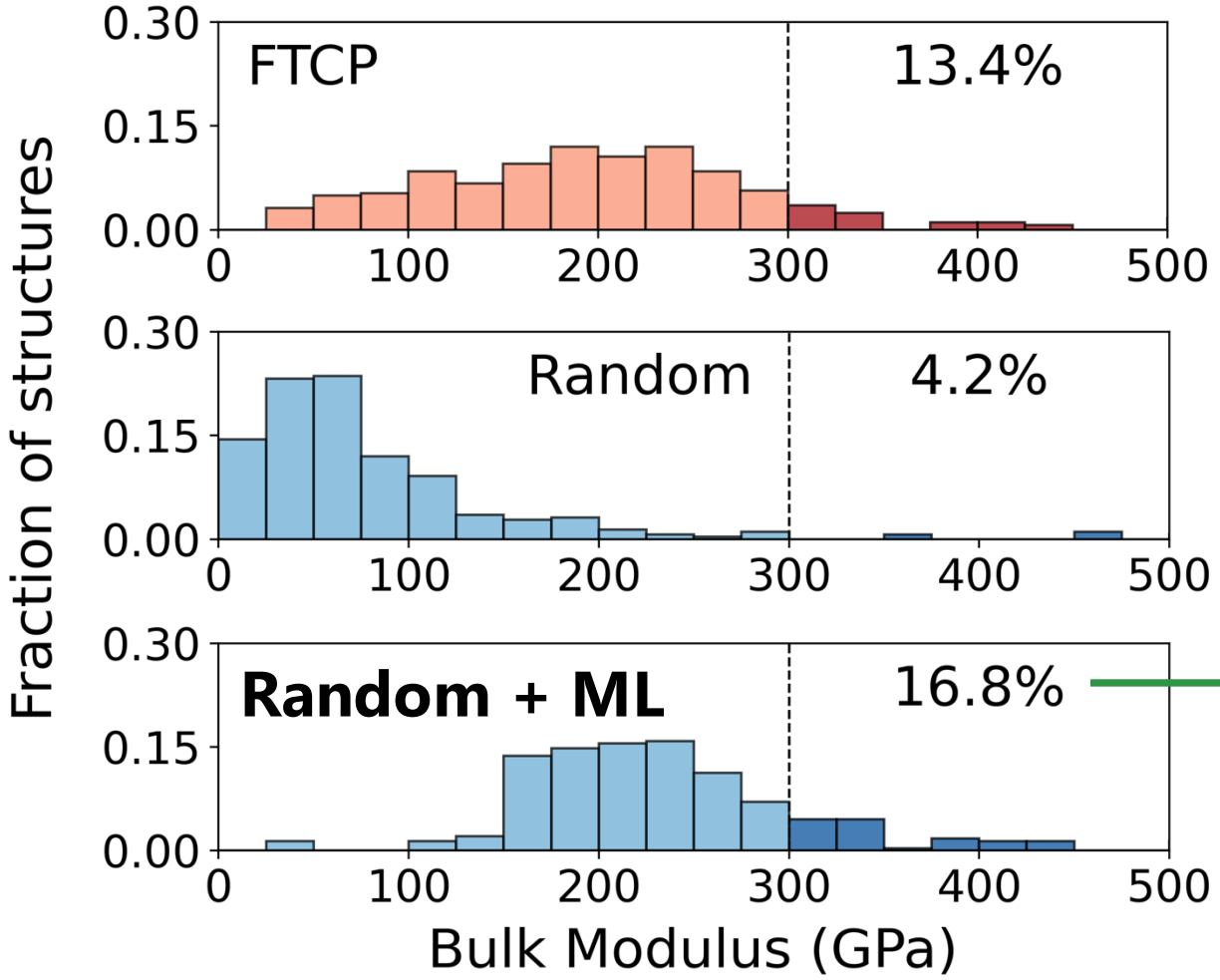
Stability rates now as now as high as 48%

S∩N rates as high as 28%

But novel prototypes remain unstable

pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

ML filtering also improves property targeting



For comparison: these two methods were applied *without* any ML filtering

Using ML to filter by predicted bulk modulus now exceeds the performance of generative AI

pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines

A summary of key takeaways

- **Template-based methods** can be used to generate **stable materials**
- If **structural novelty** is your priority, **generative AI** is the way to go
- **Generative AI** can provide targeted properties, but so can **direct ML**

We can leverage both!

pre-print: Szymanski & Bartel, arXiv:2501.02144 | **code:** github.com/Bartel-Group/matgen_baselines