# Leveraging artificial intelligence for advanced glass science and discovery

By Anurag Sachan, Hargun Singh Grover, and N. M. Anoop Krishnan

In modern times, the demand for materials with highly engineered functionalities has spurred the need for rapid innovation. This demand is accentuated by advancements in the fields of digital manufacturing, 3D printing, and internet of things platforms, where technological disruption is happening at a rapid pace.

Traditional trial-and-error methods of materials discovery struggle to meet this demand. So, researchers are adopting artificial intelligence (AI) and machine learning (ML) techniques to speed up the development process.<sup>1,2</sup>

AI and ML systems leverage vast datasets and advanced algorithms to identify novel compositions and structures in a much shorter timeframe and with greater accuracy than traditional methods. In addition, these technologies can quickly analyze vast amounts of data on structure-property relationships, which enables the development of engineered materials with properties optimized for various applications.

Glass, a disordered material obtained by the fast quenching of liquids, presents an ideal candidate for data-driven modeling due to several key factors.<sup>3</sup> First, their formation is highly versatile, as almost all elements or combinations thereof can create a glass when cooled at the required rate. Second, unlike crystalline materials, glass properties are dictated primarily by composition and processing conditions owing to their disordered structure. This feature allows for continuous tuning of the compositions, facilitating tailored design. Finally, there exists extensive experimental data on glass properties, which is ideal for modeling.



Figure 1. Three-step cyclic process of AI-based materials discovery.

Consequently, the glass community has increasingly embraced AI and ML approaches to address various challenges in glass development, including property prediction, tailored design, understanding underlying physics, and expediting modeling processes. This paper highlights one such system, Python for Glass Genomics (PyGGi),<sup>4</sup> and describes how it aims to help predict and optimize the composition-property relationships in glasses.

#### AI and ML for materials discovery

In the realm of materials science, understanding and predicting the complex relationships between composition, structure, and properties is pivotal for developing innovative materials. AI and ML systems accomplish this understanding through a three-step cyclic process (Figure 1).

The first step involves curating the information that has been created thus far by researchers and published in the scientific literature. Natural language processing algorithms can play a major role in this regard by extracting data from scholarly journals and books and organizing the information within large-scale databases.<sup>5,6</sup> This process allows for data extraction and dissemination at an unprecedented scale and pace compared to manual data curation.

The second step involves exploiting this information to predict new materials. Specifically, understanding the intricate patterns in the data and decode these patterns to discover tailored materials. Here, again, ML models combined with optimization can play a crucial role in accelerated materials discovery.<sup>7,8</sup>

The final step involves the actualization of the material, i.e., fabricating the computer-designed material in a laboratory. This step often involves high-throughput experiments, which can be automated using robotics, planning, and AL.<sup>9</sup> Such approaches are now possible, more than ever, due to the advent of large language models.<sup>10</sup>

While work remains to support the adoption of these methods in industry, the scientific studies referenced above and others provide strong evidence to conclude that AI and ML systems hold huge promise to disrupt and accelerate materials modeling, design, discovery, and manufacturing.

#### Obtaining data for computer-driven discovery

In the domain of glass science, engineering, and technology, two significant challenges obstruct the adoption of AI and ML systems: (1) the need for high-fidelity experimental datasets and (2) the automation of data extraction from literature.

The first challenge hinges on the requirement for consistent and reliable datasets, encompassing properties such as density, elastic moduli, hardness, glass transition temperature, and liquidus temperature, among others. Current databases, such as Interglad<sup>11</sup> and SciGlass,<sup>12</sup> suffer from inconsistencies and outliers, while proprietary data from companies such as Corning Inc. (Corning, N.Y.) remain inaccessible to the broader research community.

Addressing this challenge calls for an international collaborative effort to establish a universally accepted experimental glass property database allowing for the development of high-quality data-driven models. Initiatives such as PyGGi, described in the next section, offer platforms for sharing data that align with the FAIR data principle (findability, accessibility, interoperability, and reusability) and thus foster collaboration within the glass community.

The second challenge revolves around automating the extraction of data from literature, a task currently reliant on manual curation, which is inefficient for the vast amount of potentially relevant information available. Recent advancements in AI present opportunities to streamline this process, as seen in other fields such as materials science with the ChemDataExtractor toolkit.<sup>13</sup> However, the complex representations of glass compositions pose a unique challenge, necessitating the development of a glass-specific information extraction system to accurately parse and extract data from literature.

While natural language processing algorithms have shown promise in other domains,<sup>5</sup> a tailored approach for glass science literature currently is lacking, representing a crucial hurdle to overcome for the automated extraction of datasets from the literature. Addressing this challenge would significantly enhance the accessibility and utilization of valuable information within the glass research community, ultimately accelerating progress in glass science, engineering, and technology.

There have been recent efforts to combine and exploit existing databases such as SciGlass and Interglad to develop AI algorithms that can extract information from the literature. DiSCoMaT, which uses Interglad data along with data from other papers to create tables of compositions in an automated fashion, is one such model.<sup>6</sup> However, evaluating the performance of these models at scale for creating large databases is still an open problem.

Further, there have been several works that employ ML to predict properties of glasses.<sup>14</sup> However, most of these models are limited to a small range of compositions or properties. In addition, these models are either closed source and not accessible to a larger crowd or require high-level knowledge of computation and programming to use.

Thus, a SciGlass- or Interglad-like package that people can use on their own desktop for applications ranging from undergraduate education to development of commercial glass composition is a need of the hour. One such package that is developed with the aim to democratize glass discovery and make knowledge accessible to a wide range of audiences is Python for Glass Genomics (PyGGi).

#### Python for Glass Genomics

PyGGi is a pioneering software package developed by Substantial Artificial Intelligence (New Delhi, India), a startup aimed at accelerating and democratizing materials discovery. The software, which leverages machine learning algorithms to predict and optimize the composition–property relationships in glasses, comprises three main packages:

- PyGGi Bank facilitates exploration of a vast composition– property database, aiding users in retrieving glass compositions based on selected compounds and properties.
- **PyGGi Seer** employs data-driven techniques to predict up to 25 different properties for various glass compositions, crucial for material development.
- **PyGGi Zen** focuses on composition optimization, assisting users in discovering new glass compositions meeting specific property requirements.

The data-driven models developed for PyGGi are trained on extensive datasets of more than 300,000 glass compositions, encompassing more than 180 compounds and 25 different properties. By doing so, PyGGi can handle the nonlinear behavior exhibited by glasses as a function of their composition, making it a robust tool for material innovation.

Properties in the PyGGi database include optical, mechanical, electrical, and physical characteristics, among others. Compositions include oxides (e.g., SiO<sub>2</sub>, B<sub>2</sub>O<sub>3</sub>, Al<sub>2</sub>O<sub>3</sub>), halides (e.g., LiF, NaF, MgF<sub>2</sub>), and other compounds.

PyGGi is meant to be useful in both academic and industrial settings. From a teaching perspective, PyGGi is an excellent tool to explore different glass compositions and their properties. The glass selection chart allows a student to choose glass compositions with targeted properties. Further, the ternary diagram gives insights into the nonlinear and complex behavior of glasses. This feature is equally useful for an industry professional working on glasses for a variety of applications to either finetune their glass compositions or to discover new glass compositions with targeted properties, as demonstrated in the next section.

#### Case study

To demonstrate the capabilities of PyGGi, the authors used it to develop an alumina-based phosphate glass with high refractive index and Abbe number for low dispersion applications.

#### Reason for study

Due to its excellent optical properties, glass has been used extensively as a component in communication technologies, including as touch screens for smartphones and other displays, core and cladding materials for fiber optic cables, and substrates for liquid crystal displays. There are several factors to consider when designing glasses for these purposes.

## Leveraging artificial intelligence for advanced glass science and discovery

Glass can have high dispersion, resulting in colors corresponding to different wavelengths with different focal points.

Smartphone screens and other display glasses are manufactured using traditional melt cooling processes followed by expensive finishing operations, such as grinding and polishing. Therefore, a high temperature is required to form the glass, requiring additional post-treatment at a lower temperature to make it usable.

To address these challenges, glasses with excellent optical properties and a low glass transition temperature are desirable. To this extent, phosphate glasses are an attractive candidate due to their unique structure and optical and thermodynamic properties.

#### Step-by-step process

<u>Step 1:</u> Search the PyGGi Bank glass database to find similar glasses available in literature. PyGGi Bank can be searched iteratively to look at the Abbe number and refractive index of different kinds of phosphate glasses.

<u>Step 2:</u> Based on the required characteristics of the glass and the glass compositions searched in the first step, the next step is to choose the components. The following components were chosen:  $B_2O_3$ ,  $Al_2O_3$ , MgO, CaO, BaO, Li<sub>2</sub>O, ZnO, La<sub>2</sub>O<sub>3</sub>, Gd<sub>2</sub>O<sub>3</sub> and P<sub>2</sub>O<sub>5</sub>.

<u>Step 3:</u> Run an optimization in PyGGi Zen with the following parameters (Figure 2):

- 1. Methodology: Genetic algorithm
- 2. Target property: Abbe number (maximized)
- 3. Components: Use the ones selected in Step 2
- 4. Constraints on components:  $60\% \le P_2O_5 \le 100\%$
- 5. Constraints on other properties: Refractive index  $\geq 1.55$

<u>Step 4:</u> Analyze the results obtained in Step 3. To optimize the results, run the algorithm iteratively in PyGGi Zen while adding constraints to the various glass components. Table 1 shows compositions obtained after running the optimization.

<u>Step 5:</u> After satisfactory results are obtained, optimized glass compositions from Step 4 are downloaded. These compositions are then input to PyGGi Zen and other properties of interest such as hardness, refractive index, glass transition temperature, and liquidus temperature are predicted (Table 2).

<u>Step 6:</u> The results obtained in Steps 4 and 5 can be graphically analyzed using the tools available in PyGGi Zen and PyGGi Seer.

#### Outlook

It goes without saying that AI and ML systems are revolutionizing different phases of human life. While the impact in areas such as computer science, computational modeling, automation, and coding are more visible, it is impacting other sectors such as materials, construction and manufacturing, and medicine as well, albeit at a slower pace.

In the realm of glasses, modeling advancements in the last two to three decades have accelerated innovation in this field at an incredible rate compared to the material's development over the last 5,000 years. In addition, packages such as PyGGi have allowed researchers and companies with limited



Figure 2. Screenshots showing how the parameters and glass components decided in Step 2 are input into a) the desktop and b) web versions of PyGGi Zen.

R&D resources to contribute innovations by reducing reliance on the expensive and tedious trial-and-error discovery process. Furthermore, such approaches cut the environmental footprint of glass development during the discovery, manufacturing, and deployment period by supporting the identification of glass compositions with low carbon footprints and process optimization.

Overall, embracing a collaborative human–AI approach to advanced glass science and discovery enables a future where novel materials and technologies are adopted quickly and efficiently through inclusive, sustainable, and scalable development.

#### Acknowledgments

The authors would like to thank Lisa McDonald for her thoughtful edits and suggestions, which significantly improved the manuscript.

#### About the authors

Anurag Sachan is head of marketing and business development at Substantial Artificial Intelligence (New Delhi, India). Hargun Singh Grover is co-founder and CEO of Substantial Artificial Intelligence and Ph.D. student at the Indian Institute of Technology Delhi. N. M. Anoop Krishnan is associate professor at the Indian Institute of Technology Delhi. Contact Krishnan at krishnan@iitd.ac.in or Grover at hargun@substantial.ai.

Sample #	<b>B</b> <sub>2</sub> <b>0</b> <sub>3</sub>	Al <sub>2</sub> 0 <sub>3</sub>	MgO	CaO	BaO	Li <sub>2</sub> 0	Zn0	La <sub>2</sub> 0 <sub>3</sub>	Gd <sub>2</sub> O <sub>3</sub>	P <sub>2</sub> 0 <sub>5</sub>
1	15.54	3.39	0.92	2.72	15.78	3.38	0.32	0.57	1.38	56
2	15.54	3.39	1	3.05	11.75	3.38	0.32	0.57	0.97	60.03
3	16.82	3.01	0.8	3.72	14.61	2.35	0	0.47	1.15	57.07
4	17.14	3.31	1.13	3.72	13.75	0	0	0.47	1.15	59.33
5	20.86	3.31	0.8	0.53	13.75	0	0	0.47	1.15	59.13
6	18.54	0.39	1.33	2.72	11.75	3.38	0.32	0.57	0.97	60.03
7	17.32	2.01	0.8	3.72	14.61	2.35	0	0.47	1.15	57.57
8	17.32	2.51	0.8	3.72	14.11	2.35	0	0.47	1.15	57.57
9	17.32	2.51	0	3.72	14.11	2.35	0	0.47	1.15	58.37
10	21.54	2.39	1.4	2.72	9.9	0.38	0.32	0.57	0	60.78

### Table 1. Optimized glass compositions (wt.%).

## Table 2. Properties of optimized glass compositions.

Sample #	Refractive index	Abbe number	Glass transition temperature (K)	Hardness (GPa)	Liquidus temperature (K)
1	1.559625	66.39965	728.69	5.17	1131.3
2	1.551692	67.01077	723.75	5.18	1137.73
3	1.557878	66.55641	731.31	5.13	130.61
4	1.554092	66.74953	760.63	5.18	1139.35
5	1.548762	67.06789	749.45	4.96	1129.12
6	1.552433	66.36196	715.77	5.13	1129.78
7	1.557966	66.3882	726.69	5.12	1128.95
8	1.557154	66.50497	728.85	5.11	1130.28
9	1.556245	66.63231	724.2	5.09	1128.41
10	1.54524	67.4689	740.76	5.07	1135.72

## Leveraging artificial intelligence for advanced glass science and discovery

#### References

The Americar Ceramic Society

<sup>1</sup>Krishnan, N. M. A., Kodamana, H., and Bhattoo, R., Machine Learning for Materials Discovery. Springer, 2024.

<sup>2</sup>Jain, A., et al., "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials* 2013, 1(1): 011002.

<sup>3</sup>Zaki, M., Jan, A., Krishnan, N. M. A., and Mauro, J. C., "Glassomics: An omics approach toward understanding glasses through modeling, simulations, and artificial intelligence," *MRS Bulletin* 2023, **48**: 1026–1039.

<sup>4</sup>Python for Glass Genomics (PyGGi), https://pyggi.substantial.ai

<sup>5</sup>Sayeed, H.M., Mohanty, T. and Sparks, T.D. "Annotating materials science text: A semi-automated approach for crafting outputs with Gemini Pro," *Integr. Mater. Manuf. Innov.* 2024, **13**: 445–452.

<sup>6</sup>Gupta, T., Zaki, M., Khatsuriya, D., Hira, K., and Krishnan, N. M. A., "DiSCoMaT: Distantly Supervised Composition Extraction from Tables in Materials Science Articles," *Proceedings of the* 61<sup>st</sup> Annual Meeting of the Association for Computational Linguistics 2023, 1: 13465–13483.

<sup>7</sup>Ravinder, B., Bishnoi, S., Zaki, M., and Krishnan, N. M. A., "Understanding the compositional control on electrical, mechanical, optical, and physical properties of inorganic glasses with interpretable machine learning," *Acta Materialia* 2023, **242**: 118439.

<sup>8</sup>Liu, H., Fu, Z., Yang, K., Xu, X., and Bauchy, M., "Machine learning for glass science and engineering: A review," *Journal of Non-Crystalline Solids* 2021, **557**: 119419.

<sup>9</sup>Miret, S. and Krishnan, N. M. A., "Are LLMs ready for real-world materials discovery?" *Condensed Matter, Materials Science* 2024 (Preprint, under review).

<sup>10</sup>Boiko, D. A., MacKnight, R., Kline, B., and Gomes, G., "Autonomous chemical research with large language models," *Nature* 2023, **624**(7992): 570–578.

<sup>11</sup>INTERGLAD ver. 8, https://www.newglass.jp/interglad\_n/gaiyo/ info\_e.html

<sup>12</sup>SciGlass, http://www.akosgmbh.de/sciglass/sciglass.htm

<sup>13</sup>Swain, M. C., and Cole, J. M., "ChemDataExtractor: A toolkit for automated extraction of chemical information from the scientific literature," *J. Chem. Inf. Model* 2016, **56**(10): 1894–1904.

<sup>14</sup>Ravinder et al., "Artificial intelligence and machine learning in glass science and technology: 21 challenges for the 21<sup>st</sup> century," *International Journal of Applied Glass Science* 2021, **12**(3): 277–292.

# Find your path with The American Ceramic Society (ACerS)

Join fellow graduate-level Global research students who have a Graduate primary interest in ceramics esearche The key to career success for and glass. Network engineers, scientists, researchers, \$30 USD or manufacturers in the ceramic and glass materials fields. \$120 USD Associate Membership Individual Membership Material Advantage Connect with emerging Student ceramic experts to expand The premier membership Program your network while making option for full-time students the transition from student to involved in materials studies. the professional workforce. \$30 USD First year complimentary 2<sup>nd</sup> year \$40 USD Level up Lifetime Retired Emeritus **Corporate Partnership** \$2.000 USD \$40 USD Complimentary Prices vary with ACerS Avoid future dues increases Available for those who are Available for those who are Gain exposure to a global community of and the need to renew 65 years or older, have five 65 years or older and have ceramic and glass professionals, utilize each year. or more years of continuous 35 or more years of continuous the vast technical resources available.

ACerS membership, and are

retired.

ACerS membership.

Where do you see yourself? ACerS will help you get there!

View a full list of ACerS membership benefits at ceramics.org/memberbenefits. Sign up for your membership with The American Ceramic Society today at ceramics.org/join.



all while supporting your industry.

Learn more ceramics.org/corporate