Generative AI for Materials Discovery and SSbD-Driven Material Selection: From Inverse Design to Knowledge Extraction for Faster Validation

Santiago Muiños-Landin [1], Christian Eike Precker[1]
[1]AIMEN Technology Centre, Polígono Industrial de Cataboi SUR-PPI-2 (Sector 2) Parcela 3, O Porriño,
36418, Spain,
santiago.muinos@aimen.es

Generative Artificial Intelligence is rapidly transforming materials science by enabling accelerated exploration of chemical and compositional spaces, sustainability-driven decision making, and improved integration of heterogeneous knowledge into modelling and validation workflows. In this talk, we present the approach of AIMEN Technology Centre to AI for materials, covering an end-to-end pipeline that connects candidate generation, SSbD-oriented screening, and data-centric validation acceleration for industrial manufacturing scenarios.

First, we address inverse molecular and materials design using generative models to propose novel candidate compounds and compositions that satisfy targeted functional requirements. These approaches enable efficient navigation of high-dimensional design spaces and reduce the reliance on expensive trial-and-error experimentation. We highlight how conditional generative strategies can support the early-stage identification of promising candidates for advanced materials, accelerating discovery and down-selection processes [1].

Second, we demonstrate how generative and evolutionary optimization can be leveraged within Safe and Sustainable by Design (SSbD) frameworks to support the identification and ranking of candidate alternative materials and chemicals. By combining generative models with multi-objective optimization techniques, we enable the exploration of candidate solutions that remain compliant with regulatory constraints while preserving key performance requirements in manufacturing processes. This perspective is particularly relevant for industrial decision making, where sustainability and safety criteria must be integrated from the earliest design stages [2].

Finally, we introduce a data-centric GenAI workflow for automated knowledge extraction from heterogeneous materials literature and databases. Using open-source Large Language Models and Retrieval-Augmented Generation (RAG) strategies, we extract relevant material properties from unstructured text and complex tabular formats, standardize the extracted values through automated unit conversion into SI units, and structure the output to support traceable simulation model calibration and validation. This capability addresses the challenge of incomplete and scattered material property information, improving the efficiency of digital validation workflows and reducing the time required to build reliable datasets [3]. Additionally, we discuss how Natural Language Processing and GenAI-based methods can support broader lifecycle-oriented materials insight and optimization strategies in industrial contexts [4].

By combining generative design, SSbD-driven candidate screening, and LLM-based materials data acquisition, the AIMEN Technology Centre framework bridges AI-driven discovery and real-world qualification, supporting faster and more sustainable development of advanced materials and manufacturing solutions.

References

[1] Gregores Coto, A., Eike Precker, C., Andersson, T., Laukkanen, A., Suhonen, T., Rey Rodriguez, P., & Muíños-Landín, S. (2023). The use of generative models to speed up the discovery of materials. *Computer Methods in Materials Science*, 23(1), 13–26.

[2] Gregores Coto, A. G., Precker, C. E., Hernando Rodríguez, L., & Muiños Landín, S. (2025). Generative and Evolutionary Approaches for Safe and Sustainable by Design Material Development: A Case Study on Acid Alternatives for Advanced Material Coatings. In *Proceedings of the 11th International Conference on Computing and Artificial Intelligence (ICCAI 2025)*, Kyoto, Japan, pp. 736–745.

[3] Pérez, I., Suárez, L., Novas, G., & Muiños-Landin, S. (2025). *Enhancing Data Acquisition in Manufacturing: Leveraging LLMs for Material Property Databases*. Approved Abstract, ProgramMaster.

[4] Couñago, I. P., Casabiell, L. S., Gregores-Coto, A., Precker, C. E., & Muiños-Landín, S. (2025). Enhancing Product Lifecycle Efficiency: Harnessing Natural Language Processing for Materials Insight and Optimization. In K. Alexopoulos, S. Makris, & P. Stavropoulos (Eds.), *Advances in Artificial Intelligence in Manufacturing II*. ESAIM 2024. Lecture Notes in Mechanical Engineering. Springer, Cham.