



OPEN ACCESS

EDITED BY

José María Ponce-Ortega,
Michoacana University of San Nicolás de
Hidalgo, Mexico

REVIEWED BY

Francisco Javier López-Flores,
Michoacana University of San Nicolás de
Hidalgo, Mexico

Fernano Israel Gómez-Castro,
University of Guanajuato, Mexico
Rogelio Ochoa-Barragan,
Michoacana University of San Nicolás de
Hidalgo, Mexico

*CORRESPONDENCE

Fengqi You,
✉ fengqi.you@cornell.edu

RECEIVED 02 July 2024

ACCEPTED 12 August 2024

PUBLISHED 29 August 2024

CITATION

Decardi-Nelson B, Alshehri AS and You F (2024)
Generative artificial intelligence in chemical
engineering spans multiple scales.
Front. Chem. Eng. 6:1458156.
doi: 10.3389/fceng.2024.1458156

COPYRIGHT

© 2024 Decardi-Nelson, Alshehri and You. This
is an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Generative artificial intelligence in chemical engineering spans multiple scales

Benjamin Decardi-Nelson^{1,2}, Abdulelah S. Alshehri^{3,4} and
Fengqi You^{1,2,3*}

¹Systems Engineering, Cornell University, Ithaca, NY, United States, ²Cornell University AI for Science Institute, Cornell University, Ithaca, NY, United States, ³Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, United States, ⁴Department of Chemical Engineering, College of Engineering, King Saud University, Riyadh, Saudi Arabia

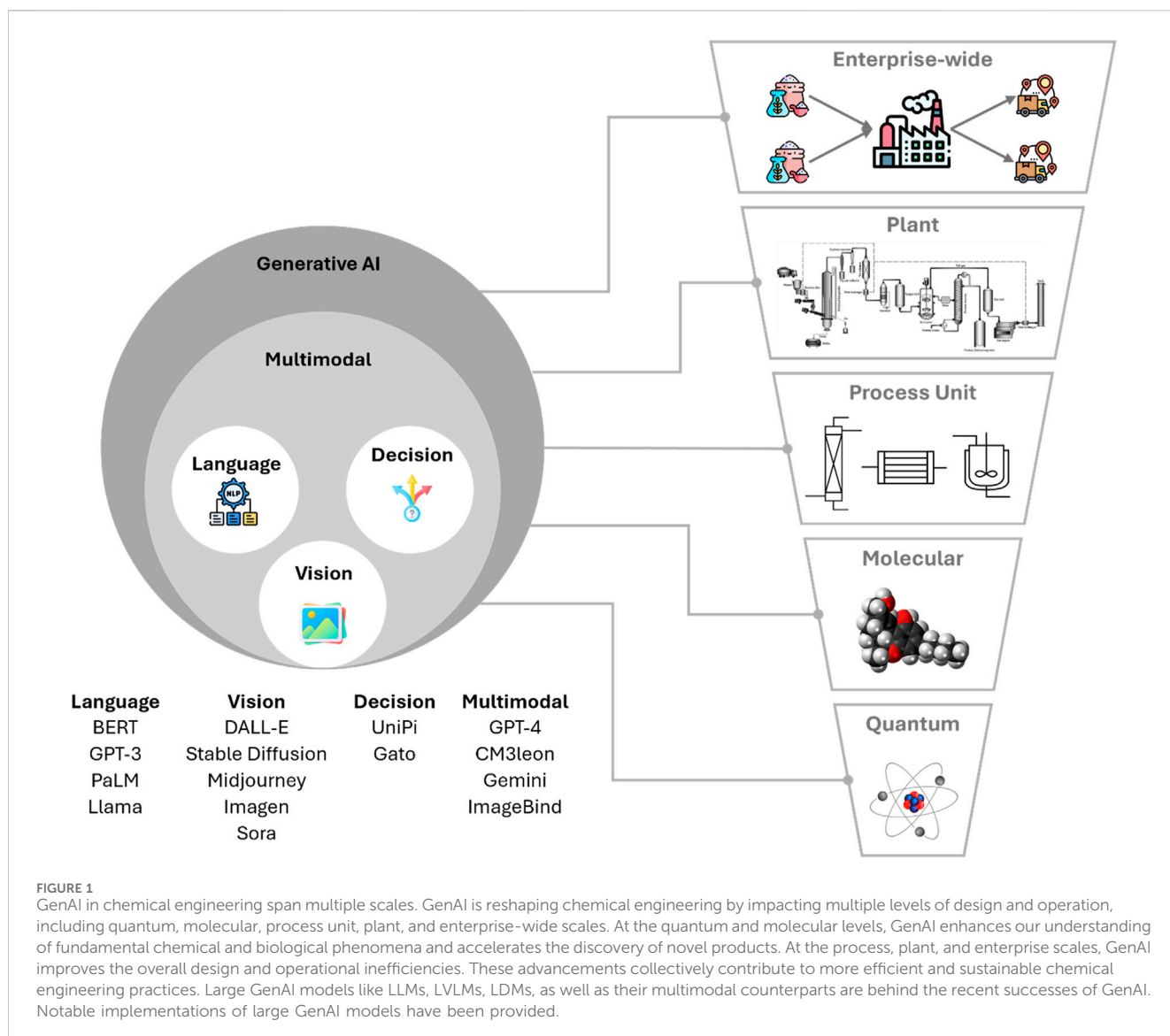
Recent advances in generative artificial intelligence (GenAI), particularly large language models (LLMs), are profoundly impacting many fields. In chemical engineering, GenAI plays a pivotal role in the design, scale-up, and optimization of chemical and biochemical processes. The natural language understanding capabilities of LLMs enable the interpretation of complex chemical and biological data. Given the rapid developments of GenAI, this paper explores the extensive applications of GenAI in multiscale chemical engineering, spanning from quantum mechanics to macro-level optimization. At quantum and molecular levels, GenAI accelerates the discovery of novel products and enhances the understanding of fundamental phenomena. At larger scales, GenAI improves process design and operational efficiency, contributing to sustainable practices. We present several examples to demonstrate the role of GenAI, including its impact on nanomaterial hardness enhancement, novel catalyst generation, protein design, and the development of autonomous experimental platforms. This multiscale integration demonstrates the potential of GenAI to address complex challenges, drive innovation, and foster advancements in chemical engineering.

KEYWORDS

artificial intelligence, AI, generative learning, quantum-chemical calculations, materials, process engineering

Introduction

Generative artificial intelligence (GenAI) has enabled several recent developments in various fields (Decardi-Nelson et al., 2024; Gangwal and Lavecchia, 2024; Preuss et al., 2024; Subramanian et al., 2024). A notable example is the few-shot learning capability of GenAI tools like ChatGPT, which can understand and interpret natural language (Wu et al., 2023). GenAI refers to artificial intelligence (AI) models that generate new data that resembles a given set of input data. Recently, large GenAI models with extensive parameters have gained significant attention for their ability to perform a wide range of tasks including natural language processing (NLP), image generation, and complex decision-making. These models include large language models (LLMs) (Zhao et al., 2023), large vision-language models (LVLMs) (Zhang et al., 2024), and large decision models (LDMs) (Zhang, 2023) (see Figure 1). Typically, these GenAI models are built using deep learning models, such as generative adversarial networks (GANs) (Goodfellow et al., 2020), autoencoders (Kingma and Welling, 2013), autoregressive (Vaswani et al., 2017), diffusion (Ho et al., 2020), and



flow-based models (Chen et al., 2019). For instance, ChatGPT is an LLM powered by a Transformer model (Vaswani et al., 2017), which is an autoregressive model. The recent success of GenAI across multiple disciplines highlights the need to explore its potential in chemical engineering.

In modern chemical engineering, which involves the design, scale-up, and optimization of chemical and biological processes, the impact of GenAI across multiple scales is equally significant. In this context, text-based representations of chemical and biological processes can be considered as codified unstructured languages to describe domain knowledge, which parallels with general NLP tasks. As discussed earlier, GenAI goes beyond NLP, to encompass mechanisms that generate data in an adversarial manner (such as GANs), or ones that mimic diffusion and flows, each uniquely equipped to capture the underlying data patterns and generate novel instances. While GenAI in process design has been previously discussed (Schweidtmann, 2024), here we emphasize that the applications of GenAI in chemical engineering extend significantly beyond such confines, poised to address a spectrum

of multiscale chemical engineering problems from quantum mechanics to macro-level optimization (see Figure 1) (Decardi-Nelson et al., 2024).

Generative AI in multiscale chemical engineering

In molecular and materials design, the integration of GenAI techniques is inspiring a multiscale design approach, from atomic-scale interactions to macroscopic phenomena (Alshehri and You, 2021). A notable example application of GenAI in tooth enamel design has demonstrated its effectiveness in enhancing nanomaterial hardness through non-destructive methods, facilitating bioinspired engineering solutions using a generative adversarial model (Goodfellow et al., 2020) with deep image regression (Lew et al., 2023). Another example in catalysis is the application of generative variational autoencoder (Kingma and Welling, 2013), inspired by interatomic insights from density functional theory (DFT) data, to

TABLE 1 Example applications of GenAI techniques across chemical engineering disciplines.

Discipline	Generation example	Generative technique
Biomolecular Engineering	Functional protein (Ingraham et al., 2023)	Diffusion models (Ho et al., 2020)
Environmental Analysis	Data retrieval (Luo et al., 2023)	Autoregressive models (Vaswani et al., 2017)
Reaction Engineering	Transition state (Duan et al., 2023)	Diffusion models (Ho et al., 2020)
Process Engineering	Chemical flowsheet (Vogel et al., 2023)	Autoregressive models (Vaswani et al., 2017)
Process Control	Soft sensing (Wang and Yan, 2024)	Generative adversarial networks (Goodfellow et al., 2020)
Soft Matter	Polymer dielectrics (Liu et al., 2023)	Flow models (Chen et al., 2019)
Separations	Metal-organic frameworks (Yao et al., 2021)	Variational autoencoder (Kingma and Welling, 2013)
Transport Phenomena	Multiphase flow (Wang et al., 2022)	Generative adversarial networks (Goodfellow et al., 2020)
Thermodynamics	Phase morphologies (Chen and Dorfman, 2023)	Generative adversarial networks (Goodfellow et al., 2020)

facilitate the generation of novel catalysts with optimized binding energies via latent space representation and deep learning-based regression (Schilter et al., 2023). These innovations span multiple areas of materials design, including drug discovery (Decardi-Nelson et al., 2024), functional biomaterials (Gartner et al., 2024), among others (Alshehri and You, 2022). This multiscale integration combining imaging techniques, quantum chemistry calculations, and molecular dynamics simulations, and empirical data using GenAI, offers improvements in molecular and materials design, bridging the gap to broader chemical product and process scales (Gartner et al., 2024).

Another aspect of chemical engineering where GenAI is significantly impacting is protein design. The Chroma (Ingraham et al., 2023) generative model samples novel protein structures and steers the design process towards desired functionalities. Incorporating a diffusion-based framework (Ho et al., 2020), the Chroma generative model captures the complex statistical distributions of natural proteins, transforming them into simpler distributions through a series of infinitesimal, constraint-biased steps, enabling the design of novel protein structures that meet specific functional requirements (Ingraham et al., 2023). These developments extend across the biomolecular domain with novel enzymes and nucleic acids, carrying promising advances for bioengineering and therapeutic innovations in medicine (Langer and Peppas, 2024), and more broadly in biomanufacturing.

On a macroscale, the integration of GenAI into robotic experimentation platforms, exemplified by GPT-Lab, can potentially transform the planning and execution of chemical experiments (Qin et al., 2023). As an (Analysis - Retrieval - Mining - Feedback - Execution) workflow, GPT-Lab employs a GPT-4 (OpenAI, 2024) as the generative model to analyze and synthesize experimental parameters, integrating these with robotic platforms for the autonomous execution of chemical syntheses (Qin et al., 2023). By mining literature for experimental parameters and validating outcomes through high-throughput synthesis, GenAI has brought us closer to achieving full-process autonomy in self-driven laboratories.

Beyond GenAI's roles in design and optimization, interpretable GenAI enhances our scientific understanding of complex phenomena within complex fluids and interfacial science, such as

the nature of disorder in domain boundaries (Dan et al., 2023). However, this important aspect of chemical engineering was not discussed in the literature (Schweidtmann, 2024). Utilizing a diffusion-based approach (Ho et al., 2020), the hybrid generative model synthesizes domain boundary structures by employing a limited Markovian dataset to algorithmically predict and scale structural motifs from atomistic to mesoscopic levels, thus uncovering critical, previously unobserved configurations that enhance our understanding and design of functional materials (Dan et al., 2023).

Table 1 illustrates the diverse applications of GenAI across prominent chemical engineering disciplines. Despite the limited examples, they underscore the expansive potential and broad applicability of various generative techniques within the branched and complex landscape of chemical engineering.

Challenges and opportunities

Despite the promising potential of GenAI in chemical engineering across multiple scales, their use and implementation come with significant challenges and limitations. Successfully addressing these challenges will require international collaboration among all stakeholders, including researchers from relevant disciplines, industrial practitioners, and regulatory authorities.

One of the foremost issues is the quality and availability of data. GenAI models, such as LLMs, need vast amounts of high-quality, domain-specific data to train effectively (Whang et al., 2023). In chemical engineering, such data is often proprietary, sparse, or inconsistent (Chiang et al., 2017), complicating the development of robust GenAI models. This challenge presents an opportunity for the entire community to collaborate in establishing standard data representations and open data-sharing platforms, thus facilitating the development and application of chemical engineering-specific GenAI models.

Another major limitation is the interpretability of GenAI models. Many large GenAI models often hallucinate (Rawte et al., 2003), and provide little to no insight into how they arrive at specific solutions (Ross et al., 2021). This lack of transparency can be a significant barrier to adoption in the safety-critical applications

often encountered in chemical engineering. Therefore, there is a need to develop benchmarks and metrics tailored to the needs of chemical engineering, requiring input from regulators, researchers, and industry. Additionally, integrating well-established first principles modeling in chemical engineering with GenAI can enhance their interpretability and trustworthiness (Takeishi and Kalousis, 2021).

Lastly, the ethical and regulatory implications of deploying GenAI in chemical engineering cannot be overlooked. Issues such as data privacy, security, and ethical considerations surrounding the autonomous nature of GenAI systems need to be carefully addressed (Huang et al., 2024). Regulatory bodies, researchers, and industrial practitioners must collaborate to establish guidelines on data use, security, and ethical issues.

Outlook

These multiscale successes demonstrate the potential of GenAI in chemical engineering. This potential extends beyond individual examples, offering novel solutions to the complex, multiscale challenges at the forefront of research in the field (Torrente-Murciano et al., 2024). At various scales, from molecular engineering to enterprise-wide supply chain (Grossmann, 2005), GenAI can enable the design and optimization of chemical and biological processes across multiple scales with high precision and efficiency. Particularly promising areas include foundation models that can be adapted to diverse chemical engineering tasks, multimodal systems that integrate heterogeneous data types (e.g., textual, visual, and experimental data), and language models that enhance data retrieval and knowledge extraction processes in chemical systems. Additionally, GenAI can facilitate advanced task learning and the development of autonomous experimental robotic systems, thereby accelerating the cycle of hypothesis generation, testing, and validation in chemical research and development. The integration of GenAI across the multiple scales and facets of chemical engineering holds the promise of significantly advancing the field, driving innovation, and fostering sustainable industrial practices.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

References

- OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., et al. (2023). *GPT-4 technical report. arXiv [cs.LG]*. Retrieved from: <http://arxiv.org/abs/2303.08774>.
- Alshehri, A. S., and You, F. (2021). Paradigm shift: the promise of deep learning in molecular systems engineering and design. *Front. Chem. Eng.* 3, 700717. doi:10.3389/fceng.2021.700717
- Alshehri, A. S., and You, F. (2022). Deep learning to catalyze inverse molecular design. *Chem. Eng. J.* 444, 136669. doi:10.1016/j.cej.2022.136669
- Chen, P., and Dorfman, K. D. (2023). Gaming self-consistent field theory: generative block polymer phase discovery. *Proc. Natl. Acad. Sci.* 120 (45), e2308698120. doi:10.1073/pnas.2308698120
- Chen, R. T., Behrmann, J., Duvenaud, D. K., and Jacobsen, J.-H. (2019). Residual flows for invertible generative modeling. *Adv. Neural Inf. Process. Syst.*, 32.
- Chiang, L., Lu, B., and Castillo, I. (2017). Big data analytics in chemical engineering. *Annu. Rev. Chem. Biomol. Eng.* 8, 63–85. doi:10.1146/annurev-chembioeng-060816-101555
- Dan, J., Waqar, M., Erofeev, I., Yao, K., Wang, J., Pennycook, S. J., et al. (2023). A multiscale generative model to understand disorder in domain boundaries. *Sci. Adv.* 9 (42), eadj0904. doi:10.1126/sciadv.adj0904
- Decardi-Nelson, B., Alshehri, A. S., Ajagekar, A., and You, F. (2024). Generative AI and process systems engineering: the next frontier. *Comput. and Chem. Eng.* 187 108723. doi:10.1016/j.compchemeng.2024.108723
- Duan, C., Du, Y., Jia, H., and Kulik, H. J. (2023). Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nat. Comput. Sci.* 3 (12), 1045–1055. doi:10.1038/s43588-023-00563-7

Author contributions

BD-N: Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft, Writing—review and editing. AA: Investigation, Writing—review and editing. FY: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. BD-N. acknowledges the partial support from Schmidt Futures via an Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship to Cornell University.

Acknowledgments

BD-N. acknowledges the partial support from Schmidt Futures via an Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship to Cornell University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Gangwal, A., and Lavecchia, A. (2024). Unleashing the power of generative AI in drug discovery. *Drug Discov. Today* 29, 103992. doi:10.1016/j.drudis.2024.103992
- Gartner, T. E., Ferguson, A. L., and Debenedetti, P. G. (2024). Data-driven molecular design and simulation in modern chemical engineering. *Nat. Chem. Eng.* 1 (1), 6–9. doi:10.1038/s44286-023-00010-4
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2020). Generative adversarial networks. *Commun. ACM* 63 (11), 139–144. doi:10.1145/3422622
- Grossmann, I. (2005). Enterprise-wide optimization: a new frontier in process systems engineering. *AIChE J.* 51 (7), 1846–1857. doi:10.1002/aic.10617
- Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Adv. neural Inf. Process. Syst.* 33, 6840–6851.
- Huang, K., Ponnappalli, J., Tantsura, J., and Shin, K. T. (2024). “Navigating the GenAI security landscape,” in *Generative AI security: theories and practices*. Editors K. Huang, Y. Wang, B. Goertzel, Y. Li, S. Wright, and J. Ponnappalli (Nature Switzerland: Springer), 31–58.
- Ingraham, J. B., Baranov, M., Costello, Z., Barber, K. W., Wang, W., Ismail, A., et al. (2023). Illuminating protein space with a programmable generative model. *Nature* 623 (7989), 1070–1078. doi:10.1038/s41586-023-06728-8
- Kingma, D. P., and Welling, M. (2013). Auto-encoding variational bayes. *arXiv Prepr. arXiv:1312.6114*.
- Langer, R., and Peppas, N. A. (2024). A bright future in medicine for chemical engineering. *Nat. Chem. Eng.* 1 (1), 10–12. doi:10.1038/s44286-023-00016-y
- Lew, A. J., Stifler, C. A., Cantamesa, A., Tits, A., Ruffoni, D., Gilbert, P. U., et al. (2023). Deep learning virtual indenter maps nanoscale hardness rapidly and non-destructively, revealing mechanism and enhancing bioinspired design. *Matter* 6 (6), 1975–1991. doi:10.1016/j.matt.2023.03.031
- Liu, D.-F., Zhang, Y.-X., Dong, W.-Z., Feng, Q.-K., Zhong, S.-L., and Dang, Z.-M. (2023). High-temperature polymer dielectrics designed using an invertible molecular graph generative model. *J. Chem. Inf. Model.* 63 (24), 7669–7675. doi:10.1021/acs.jcim.3c01572
- Luo, B., Liu, J., Deng, Z., Yuan, C., Yang, Q., Xiao, L., et al. (2023). *AutoPCF: a novel automatic product carbon footprint estimation framework based on large language models* in Proceedings of the AAAI Symposium Series 2 (1), 102–106. doi:10.1609/aaais.v21i.27656
- Preuss, N., Alshehri, A. S., and You, F. (2024). Large language models for life cycle assessments: opportunities, challenges, and risks. *J. Clean. Prod.* 466, 142824. doi:10.1016/j.jclepro.2024.142824
- Qin, X., Song, M., Chen, Y., Ai, Z., and Jiang, J. (2023). GPT-lab: next generation of optimal chemistry discovery by GPT driven robotic lab. *arXiv Prepr. arXiv:230916721*. doi:10.48550/arXiv.2309.16721
- Rawte, V., Sheth, A., and Das, A. (2003) *A survey of hallucination in large foundation models. arXiv preprint arXiv:230905922*. 2023.
- Ross, A., Chen, N., Hang, E. Z., Glassman, E. L., and Doshi-Velez, F. (2021) “Evaluating the interpretability of generative models by interactive reconstruction,” in *Presented at: proceedings of the 2021 CHI conference on human factors in computing systems*. Yokohama, Japan.
- Schilter, O., Vaucher, A., Schwaller, P., and Laino, T. (2023). Designing catalysts with deep generative models and computational data. A case study for Suzuki cross coupling reactions. *Digit. Discov.* 2 (3), 728–735. doi:10.1039/D2DD000125J
- Schweidtmann, A. M. (2024). Generative artificial intelligence in chemical engineering. *Nat. Chem. Eng.* 1 (3), 193. doi:10.1038/s44286-024-00041-5
- Subramanian, A., Gao, W., Barzilay, R., Grossman, J. C., Jaakkola, T., Jegelka, S., et al. (2024). *Closing the execution gap in generative AI for chemicals and materials: freeways or safeguards*. An MIT Exploration of Generative AI.
- Takeishi, N., and Kalousis, A. (2021). *Physics-integrated variational autoencoders for robust and interpretable generative modeling*.
- Torrente-Murciano, L., Dunn, J. B., Christofides, P. D., Keasling, J. D., Glotzer, S. C., Lee, S. Y., et al. (2024). The forefront of chemical engineering research. *Nat. Chem. Eng.* 1 (1), 18–27. doi:10.1038/s44286-023-00017-x
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). *Advances in Neural Information Processing Systems* 30. Retrieved from: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91bd053c1c4a845aa-Paper.pdf.
- Vogel, G., Schulze Balhorn, L., and Schweidtmann, A. M. (2023). Learning from flowsheets: a generative transformer model for autocompletion of flowsheets. *Comput. and Chem. Eng.* 171, 108162. doi:10.1016/j.compchemeng.2023.108162
- Wang, Y., and Yan, P. (2024). RegGAN: a virtual sample generative network for developing soft sensors with small data. *ACS omega* 9 (5), 5954–5965. doi:10.1021/acsomega.3c09762
- Wang, Z., Jeong, H., Gan, Y., Pereira, J.-M., Gu, Y., and Sauret, E. (2022). Pore-scale modeling of multiphase flow in porous media using a conditional generative adversarial network (cGAN). *Phys. Fluids* 34 (12), 123325. doi:10.1063/5.0133054
- Whang, S. E., Roh, Y., Song, H., and Lee, J.-G. (2023). Data collection and quality challenges in deep learning: a data-centric AI perspective. *VLDB J.* 32 (4), 791–813. doi:10.1007/s00778-022-00775-9
- Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., et al. (2023). A brief overview of ChatGPT: the history, status quo and potential future development. *IEEE/CAA J. Automatica Sinica* 10 (5), 1122–1136. doi:10.1109/JAS.2023.123618
- Yao, Z., Sánchez-Lengeling, B., Bobbitt, N. S., Bucior, B. J., Kumar, S. G. H., Collins, S. P., et al. (2021). Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nat. Mach. Intell.* 3 (1), 76–86. doi:10.1038/s42256-020-00271-1
- Zhang, J., Huang, J., Jin, S., and Lu, S. (2024). Vision-Language models for vision tasks: a survey. *IEEE Trans. Pattern Analysis Mach. Intell.* 46 (8), 5625–5644. doi:10.1109/TPAMI.2024.3369699
- Zhang, W. (2023) “Large decision models,” in Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23. Editor E. Elkind, 7062–7067. doi:10.24963/ijcai.2023/808
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., et al. (2023) *A Survey of Large Language Models*. arXiv [Cs.CL]. Retrieved from: <http://arxiv.org/abs/2303.18223>.