ACS Sustainable Chemistry & Engineering

pubs.acs.org/journal/ascecg

Editorial

Can Artificial Intelligence and Machine Learning Be Used to Accelerate Sustainable Chemistry and Engineering?

Cite This: *ACS Sustainable Chem. Eng.* 2021, 9, 6126–6129

Read Online

ACCESS | Metrics & More | Article Recommendations

Is the chemical industry on the cusp of a feedstock revolution? It seems likely but only if viable technologies for alternative raw materials—such as biomass, carbon dioxide, and recycled plastics—can be developed. Without new discoveries in catalysis and processing, the chemical industry will continue to rely on traditional petroleum conversions, which often consume enormous amounts of energy, emit greenhouse gases, pollute and consume water, and rely upon precious metals that are scarce or obtained from conflict minerals. New sustainable chemical systems could help mitigate these challenges. However, we need a more strategic approach to decipher these data-rich systems and shed light on new directions of study.

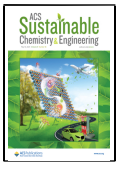
With recent advances in the computer sciences, we now have the power to use artificial intelligence (AI) to logically guide sustainable chemistry research by uncovering complex performance relationships. While AI has impacted areas like bioinformatics and drug discovery, catalysis and sustainable chemistry fields have yet to benefit significantly from the data science revolution.[1] There is enormous potential for high-impact synergies between these fields and computer sciences.[2−4] Take catalysis development for example. For more than a century, researchers have mostly relied on either an Edisonian trial and error approach or empirical evidence to design catalysts and reaction systems, but these efforts are time intensive and hit and miss in terms of results. A plethora of factors affect the performance of a catalyst, including operating conditions, elemental composition of the catalysts (metals, supports, and impurities), morphology of the catalysts (phase, porosity, surface area, conductivity, and more), and reactor configuration and operation. Convoluted by the sheer number of variable combinations, it is exceedingly arduous and time consuming for researchers to make significant advancements through the traditional trial-and-error approach. Leveraging state-of-the-art AI technologies is critical to enable sustainable chemistry and catalysis researchers to mine, organize, and exploit the myriad data sources relevant to reaction innovations (e.g., temperatures, pressures, solvents, metals, supports, molecular makeup, and reactor configurations).

While AI has several subfields (speech processing, vision, and robotics), two areas of most use to the sustainable chemistry and catalysis community are machine learning (ML) and natural language processing (NLP). Machine learning is a subfield of AI that is focused on solving practical problems by building a statistical model of a given data set.[5] The key to machine learning is having a large and high-quality data set in order to build the statistical model. Otherwise, one is susceptible to the computer science adage "garbage in, garbage out". Fortunately, for chemistry and catalysis researchers, the archival literature in chemistry and engineering journals contains vast amounts of high quality experimental and computational data. Unfortunately, this data is not contained in a uniform data set immediately suitable for machine learning algorithms. It is spread out over dozens of publishing houses in hundreds of journals with various html or pdf formats. Nonuniform writing styles further complicate the data. It is possible to automatically extract information from human text using natural language processing, which aims to give meaning to each word of text, so that computers can decipher human language within the context of a sentence or paragraph.

We envision that the future of sustainable chemistry and catalyst design is at the intersection of data-driven research and fundamental mechanistic studies, leveraging both artificial intelligence and human intelligence. We aptly call this design approach CataLST (pronounced catalyst, Figure 1). The steps in CataLST are (1) Catalog the literature with natural language processing and data mining, (2) Learn from this knowledge base using machine learning to uncover new insights, (3) Search for new catalysts and/or reaction conditions with these fundamental insights complemented by machine learning, computational chemistry models, and human intuition, and (4) Test and validate performance experimentally. This discovery model holds the potential to be deployed in a versatile manner: at small scales (e.g., using a limited subset of data from the literature or laboratory) or at larger scales with bigger data sets for groundbreaking research. We envision that researchers employing the CataLST cycle will aid in establishing an "Internet of Catalysis" (IoC) capable of harnessing data to rationally develop new catalysts and processes. The IoC holds the potential to electronically archive curated data from an experiment or mountain of publications, analyze that data, and then point the way to a certain molecular construct or set of conditions that will activate a desired reaction.

For data-driven approaches that are based on experimental data, several areas of research are ripe for development. First, new natural language processing algorithms are needed to be

Received: April 22, 2021
Published: May 10, 2021



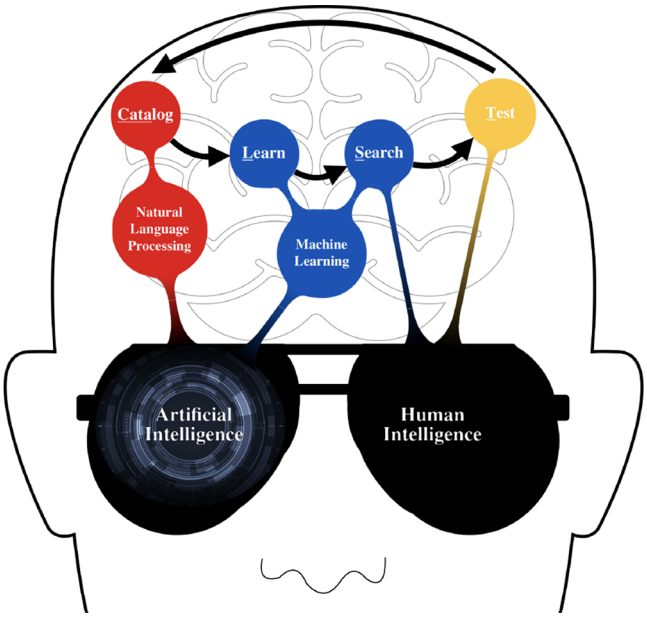


Figure 1. What is the future of sustainable chemistry research? Interfacing artificial intelligence and human intelligence to (1) Catalog the literature with data mining, (2) Learn from this knowledge base using machine learning, and (3) use these insights to Search and Test new systems.

able to extract information from different subfields for sustainable chemistry and engineering. For example, one subtask of natural language processing is named entity recognition, which classifies words or groups of words to predefined categories (Figure 2a). Chemical named entity recognition algorithms have been developed to classify if a particular word is a chemical,[6] but more specialized named entity recognition algorithms are needed to fully extract relevant experimental data from the literature. For example, these algorithms need to identify and distinguish products from reactants, determine a catalyst structure, and accurately elucidate relevant reaction conditions.

New deep learning and predictive analytics algorithms hold the potential to greatly accelerate sustainable chemistry design (Figure 2B). Deep learning, one of the most rapidly growing subfields of machine learning, demonstrates remarkable power in deciphering multiple layers of representations from high-throughput experiment and/or theoretical calculation data without the need of designing and tuning specific feature extractors. By using deeper neural networks that provide a hierarchical representation of the data, deep learning methods, such as convolutional neural networks, recurrent neural networks, and deep generative models (e.g., variational autoencoders and generative adversarial networks), are shown to have much more powerful learning capabilities and thus higher performance and precision for searching for new catalysts. Deep learning techniques can be fed with raw data directly, have automatic features, and are able to learn rapidly for performing various data processing and predictive analysis tasks efficiently. Many research studies have reported the application to chemical molecular generation and property predictions.[7] Deep learning-based generative models are typically used in conjunction with predictive QSPR models to relate learned feature representations of molecular descriptors to target chemical, physical, or biological properties of catalysts. In addition to overperforming other machine learning methods in property predictions, deep learning has recently demonstrated the capability to produce property predictions comparable to density functional theory (DFT) calculations.[8] With the advances of deep learning algorithms, more powerful architectures of generative models, and increasing availability of experimental and computational chemistry data, it is expected that deep learning techniques have great potential for improving the effectiveness and accuracy of AI-based catalyst design for sustainable chemistry applications.

In addition to catalyst development, machine learning can be utilized for solving complex chemical engineering problems. For example, the design and discovery of sustainable pathways often require developing multiscale process systems engineering (PSE) methods addressing systems which are complex and vary across different time and length scales. ML-based models can be utilized for developing highly accurate surrogate models to circumvent the need of representing chemical phenomena via complex and nonlinear relationships.[9] As one example, if we were to solve a complex chemical engineering supply chain problem, we could replace the chemical phenomena occurring inside unit operations with the help of ML-based surrogate models. These could then be integrated within the broader supply chain level to capture both the process-level and supply chain-level aspects. This strategy could significantly improve the tractability of large-scale PSE problems. Another example is CO_2 capture, utilization, and storage (CCUS). The overall cost of CO_2 capture depends on multiple and often contradicting factors at materials, process, and supply chain levels.[10] ML can contribute to all levels. At the materials level, chemistry-informed machine learning can be applied to perform efficient screening of adsorbents, solvents, and membranes for CO_2 capture, natural gas purification, hydrogen storage, and separation. At the process level, reliable estimations of physicochemical, equilibrium, and transport properties are critical. ML can be used to efficiently predict these properties of chemicals and materials for conceptual design, synthesis, and intensification of chemical process flowsheets.[11] ML can be also applied to analyze and improve process economics and safety in the chemical process industry (CPI). ML models can be developed to provide real time assessment of a plant's safety and operability, providing significant value to plant operators. These models can be translated to track various safety metrics (e.g., fire, toxic release, hazard, and risk) and safety-critical material properties.[12] At the supply chain level, key interests are in performing systems analysis for nationwide or regional supply chain structures that would use the most appropriate sources, technologies and materials, transportation networks, and utilization and demand sites.

Industrially, machine learning and artificial intelligence are already starting to impact the drug discovery process. Research is ongoing within large pharmaceutical companies and specialist companies such as Ex Scientia and Benevolent AI to use algorithms to design compounds and predict their properties with the goal of increasing the efficiency of the drug discovery process and thereby the speed to market. From a sustainability perspective, the drug discovery process is often the subject of less attention than the larger-scale drug development process, where there is a far greater probability of any given process transferring into manufacturing. However, the drug discovery process has been estimated to produce between 200,000–2,000,000 kg of waste a year,[13] and so not only do these algorithms hold the promise of improving the speed at which new treatments can get to patients, they should also minimize

https://doi.org/10.1021/acssuschemeng.1c02741
ACS Sustainable Chem. Eng. 2021, 9, 6126–6129

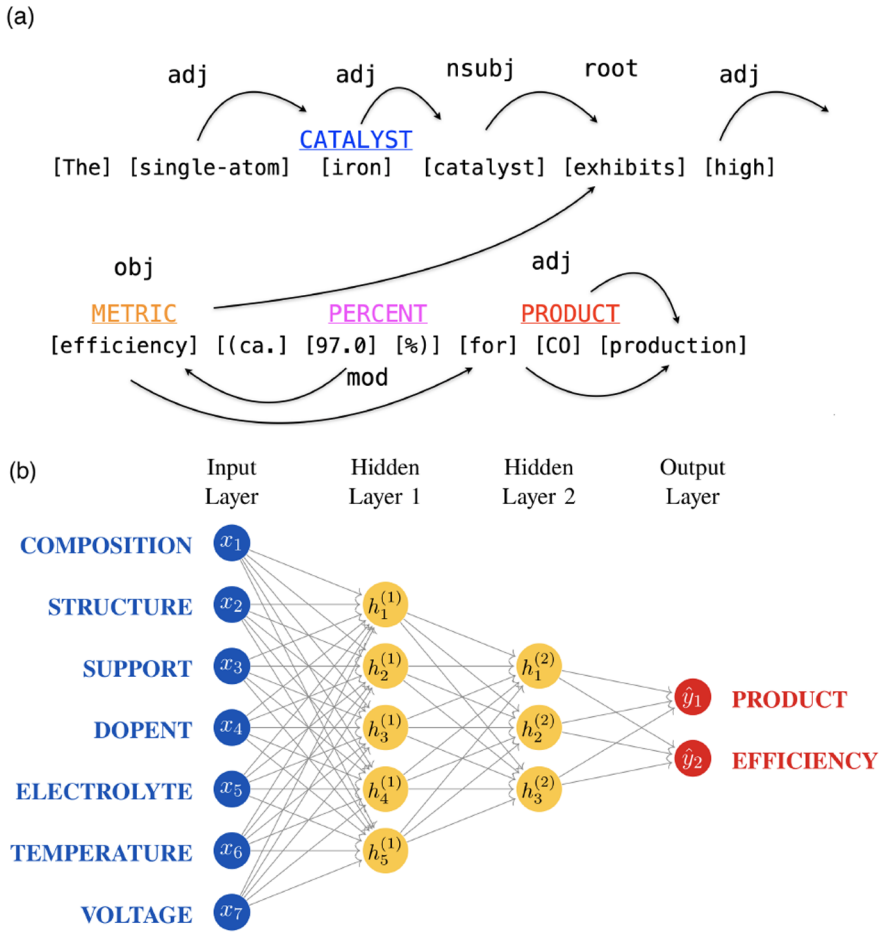


Figure 2. Pictorial representation of natural language processing (a) and machine learning (b). (a) Example sentence showing the named entity recognition tags (colored labels) and part of speech relationships. (b) Artificial neural network that has reaction conditions as the input layer and the product and efficiency as the output layer.

the environmental impact of this phase of the industry. The application of machine learning to synthetic problems has also generated considerable interest and excitement.[14] Again, this has sustainability implications if it can enable shorter, more efficient, higher yielding routes to key targets of interest. Already the technology has developed to such an extent where, in a variant of the Turing test, a panel of chemistry students showed no preference for literature routes to machine-suggested synthetic routes.[15] Additionally efforts are underway, through research partnerships such as the GSK University of Nottingham and University of Strathclyde Prosperity Partnership,[16] to use algorithms to predict the most environmentally benign routes to target molecules. This Prosperity Partnership aims to build an AI-enabled sustainable chemistry community,[17] augmenting tools such as CHEM21[18] by developing and deploying models that are explainable for sustainable chemistry and bringing to researchers' fingertips intuitive and interactive tools, which allow users to define priorities, such as reaction yield, or focus on longer-term considerations, such as the ease with which the chemistry can be scaled up to a chemical engineering process.

To drive all of these innovations, more scientists and engineers need to learn how to integrate AI and ML in their sustainable chemistry investigations. *ACS Sustainable Chemistry & Engineering* (*ACS SCE*) plans to publish a Virtual Special Issue (VSI) later this year titled Advances in Sustainable Chemistry via Artificial Intelligence, featuring methods in which AI has been used in enhancing sustainable chemistry and engineering. The content should preferably address the following aspects: (a) using machine learning algorithms to enhance any aspect of sustainable chemistry and engineering, (b) using novel chemical data extraction or mining techniques, and/or (c) interfacing machine learning with traditional computational chemistry. While all three aspects are welcome, each manuscript submission need not address all aspects.

Please note that the foregoing guidelines are meant as suggestions. Authors are encouraged to add other aspects as appropriate. If you wish to specifically contribute to the VSI in preparation, please email us at your early convenience. We look forward to receiving manuscripts in the important area of translational sustainable chemistry and engineering.

Kevin C. Leonard, Early Career Board orcid.org/0000-0002-0172-3150
Faruque Hasan, Early Career Board orcid.org/0000-0001-9338-6069
Helen F. Sneddon, Editorial Advisory Board
Fengqi You, Editorial Advisory Board orcid.org/0000-0001-9609-4299

AUTHOR INFORMATION

Complete contact information is available at:
https://pubs.acs.org/10.1021/acssuschemeng.1c02741

https://doi.org/10.1021/acssuschemeng.1c02741
ACS Sustainable Chem. Eng. 2021, 9, 6126−6129

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.

REFERENCES

(1) Venkatasubramanian, V. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE J.* **2019**, *65*, 466−478.

(2) Kitchin, J. R. Machine learning in catalysis. *Nature Catalysis* **2018**, *1*, 230−232.

(3) Goldsmith, B. R.; Esterhuizen, J.; Liu, J.-X.; Bartel, C. J.; Sutton, C. Machine learning for heterogeneous catalyst design and discovery. *AIChE J.* **2018**, *64*, 2311−2323.

(4) Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R. Extracting knowledge from data through catalysis informatics. *ACS Catal.* **2018**, *8*, 7403−7429.

(5) Burkov, A. *The Hundred-Page Machine Learning Book*; Andriy Burkov: Canada, 2019; Vol. *1*, pp 3−5.

(6) Swain, M. C.; Cole, J. M. ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature. *J. Chem. Inf. Model.* **2016**, *56*, 1894−1904.

(7) Alshehri, A. S.; Gani, R.; You, F. Deep Learning and Knowledge-Based Methods for Computer-Aided Molecular Design-Toward a Unified Approach: State-of-the-Art and Future Directions. *Comput. Chem. Eng.* **2020**, *141*, 107005.

(8) Jha, D.; Choudhary, K.; Tavazza, F.; Liao, W.-k.; Choudhary, A.; Campbell, C.; Agrawal, A. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. *Nat. Commun.* **2019**, *10*, 1−12.

(9) Grimstad, B.; Andersson, H. ReLU networks as surrogate models in mixed-integer linear programs. *Comput. Chem. Eng.* **2019**, *131*, 106580.

(10) Hasan, M. F.; First, E. L.; Boukouvala, F.; Floudas, C. A. A multi-scale framework for CO_2 capture, utilization, and sequestration: CCUS and CCU. *Comput. Chem. Eng.* **2015**, *81*, 2−21.

(11) Lee, J. H.; Shin, J.; Realff, M. J. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Comput. Chem. Eng.* **2018**, *114*, 111−121.

(12) Ji, C.; Yuan, S.; Jiao, Z.; Huffman, M.; El-Halwagi, M. M.; Wang, Q. Predicting flammability-leading properties for liquid aerosol safety via machine learning. *Process Saf. Environ. Prot.* **2021**, *148*, 1357.

(13) Williams, R. T.; Williams, T. R. Environmental Science; Guiding Green Chemistry, Manufacturing, and Product Innovations. In *Green Techniques for Organic Synthesis and Medicinal Chemistry*; Zhang, W., Cue, B. W., Jr., Eds.; Wiley Online, 2021; pp 33−66. DOI: 10.1002/9780470711828.ch3.

(14) Campos, K. R.; Coleman, P. J.; Alvarez, J. C.; Dreher, S. D.; Garbaccio, R. M.; Terrett, N. K.; Tillyer, R. D.; Truppo, M. D.; Parmee, E. R. The importance of synthetic chemistry in the pharmaceutical industry. *Science* **2019**, *363*, eaat0805.

(15) Segler, M. H.; Preuss, M.; Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555*, 604−610.

(16) Kerr, W. J. Accelerated Discovery and Development of New Medicines: Prosperity Partnership for a Healthier Nation. *Engineering and Physical Sciences Research Council*. https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/S035990/1 (accessed 04/11/2021).

(17) Icke, J. Prestigious Award for New Research into Machine Learning for Sustainable Chemistry. *University of Nottingham*. https://www.nottingham.ac.uk/news/prestigious-award-for-new-research-into-machine-learning-for-sustainable-chemistry (accessed 04/15/2021).

(18) McElroy, C. R.; Constantinou, A.; Jones, L. C.; Summerton, L.; Clark, J. H. Towards a holistic approach to metrics for the 21st century pharmaceutical industry. *Green Chem.* **2015**, *17*, 3111−3121.

https://doi.org/10.1021/acssuschemeng.1c02741
ACS Sustainable Chem. Eng. 2021, 9, 6126−6129