ACS OMEGA

http://pubs.acs.org/journal/acsodf

Editorial

Diving into the Deep End: Machine Learning for the Chemist

Cite This: *ACS Omega* 2022, 7, 25906−25908

Read Online

ACCESS | Metrics & More | Article Recommendations

In the past year, *ACS Omega* has seen a dramatic increase in the number of articles published with an Artificial Intelligence (AI) or Machine Learning, Deep Learning, Neural Networks theme. In 2021, 105 articles were published vs 45 articles published in the previous year, an increase of 133%. This exceptional growth for a fully open access broad scope journal pairs with the growth seen at many other journals in the ACS portfolio. Interestingly, the growth registered in the past 2−3 years is not confined to the journals that specialize in chemical informatics: the *Journal of Chemical Information and Modeling*, the *Journal of Chemical Theory and Computation*, and, to some extent, the *Journal of Physical Chemistry C*. It also encompasses journals in the materials sciences, the physical sciences, measurement science, chemical engineering, and environmental science in the broader ACS portfolio (Figure 1).

Looking at the wider publication landscape, the Directory of Open Access Journal lists 65 journal entries for scientific publications that pertain to the topic of AI.[1] Eleven of these were added in the last year alone, and this includes only those journals queried in the computational science category. In addition to these, numerous other open access, broader scope journals also publish work without any perceived evaluation of immediate impact and where existing AI tools have been successfully applied to a variety of chemistry questions.

Over the past 10 to 15 years, AI, especially deep learning, has effected dramatic technological progress and proven success in areas such as computer vision,[2] speech recognition,[3] natural language processing,[4,5] common sense knowledge,[6,7] strategic reasoning,[8] and robotics.[9] Exceptional results have also been reported in the medical sciences; for example, deep neural networks facilitated accurate diagnosis of skin cancer,[10] and deep learning enabled extraction of new knowledge from old data, enabling accurate prediction of age, gender, smoking status, blood pressure, and heart attack propensity of individuals just by analyzing previously acquired retinal images.[11]

But, what about the status of AI and its perception in chemistry? In the ensuing text, as an entrée to the associated Virtual Issue, I will present a brief overview of the perceived usefulness of AI at this time in some fields of the chemical sciences and related areas, based on recently published reviews and perspectives by experts in the area, as well as other resources.

A review by Baum et al.[12] charted the growth and distribution of AI-related chemistry publications in the last two decades using the CAS Content Collection, which includes patents as well as research articles. In their paper, they refer to the "Hype Cycle of Emerging Technologies",[13] and from the data gathered, they determine that AI adoption in *life sciences* and *analytical chemistry* has navigated the so-called "peak of inflated expectations" and "trough of disillusionment" and successfully progressed to the "plateau of productivity".

It is common to overestimate the effect of a technology in the short run and underestimate it in the long run (anecdotally known as Amara's law). For example, despite commercial interests and consequent investments being enormous, to this day, no new drug has yet been synthesized using AI.[14,15] As another (counter) example, Peiretti and Brunel[16], in their Perspective published in 2018, ponder whether organic chemistry and in particular retrosynthesis might be the ideal next application of AI techniques. After all, it "*fits perfectly*" the definition of AI as a problem with "*complex input−output relationships [that] are difficult or impractical to model procedurally*". However, Baum et al.[12] firmly assess that "*there are still areas of Chemistry like organic synthetic chemistry where AI is yet to make an impact*". Still, work is in progress, as standardized formats for reporting a chemical synthesis procedure are being developed and even classified (an essential step for scientific fields to exist) in the new taxonomy of *digital chemistry or chemputation*.[17]

At this stage in the discussion, it may be interesting to explore what the drivers are to greater or lesser success in applying AI methods to scientific problems. This point is examined in the Editorial by Jones et al.[18], which accompanies an excellent *JACS Au* special collection on "Emerging Chemistry & Machine Learning". In their overview, three main reasons are indicated. Two of them are quite intrinsic to the method (algorithm development and theoretical derivation of descriptors), while one of them is, notably, not specific but resides in the availability of a collection of standardized, high-quality data. A case in point, and unanimously defined as the most spectacular recent success of AI, is the recent prediction of a protein's three-dimensional structure from its amino acid sequence via AlphaFold.[19,20] The tool was trained on large publicly available databases, such as the RSCB Protein Data Bank,[21] as well as protein sequences of unknown structure. In return, it is somewhat expected, although not guaranteed, that the new information generated is also made available to the public.[22] The AlphaFold code is now publicly available. It can

Published: July 20, 2022

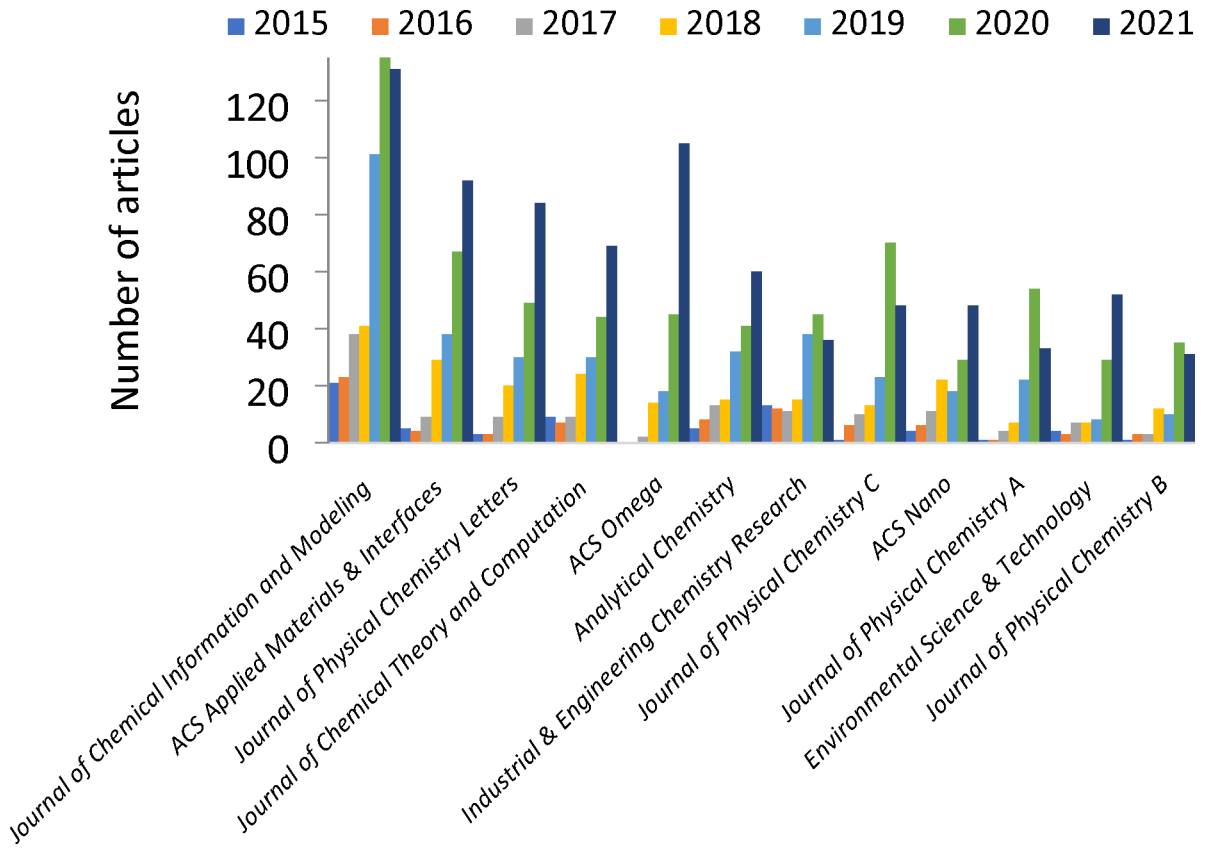

Figure 1. Top 12 journals at ACS Publications for publishing AI research. Bar graphs represent Web of Science results for the named ACS journals using "artificial intelligence", "machine learning", "deep learning", and "neural networks" as search topics. *ACS Omega* was launched in 2016. These 12 journals constitute approximately two-thirds of the publications of the entire portfolio in this area.

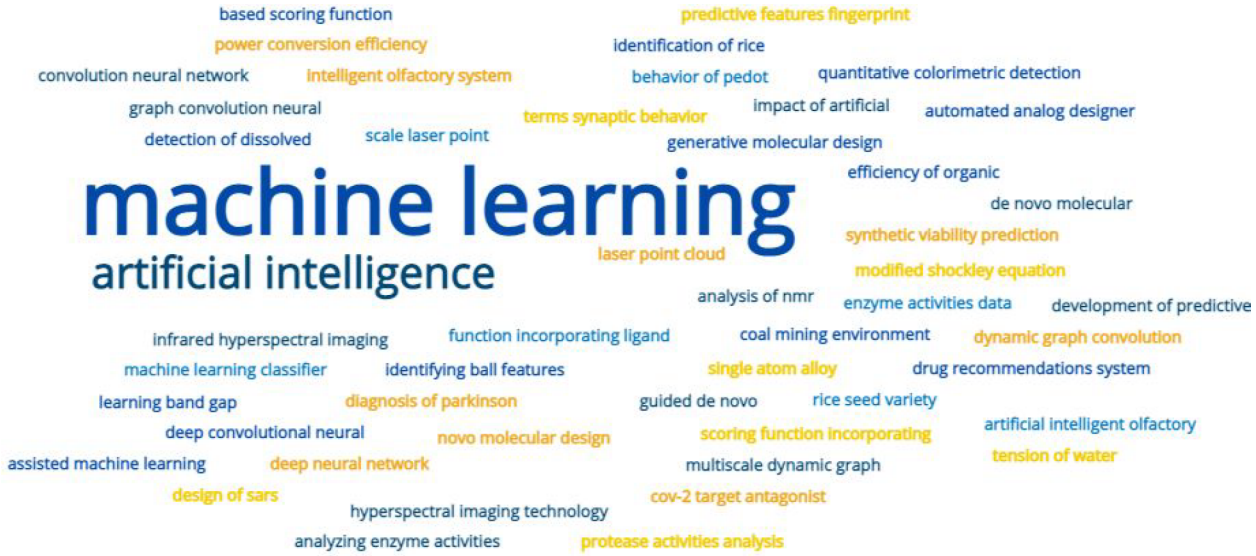

Figure 2. This AI-generated word cloud is obtained from the selected article titles in this Virtual Issue via MonkeyLearn.

be run with moderate computational resources,[23] and the "most complete" database of predicted 3D structures of human proteins has been shared with the world.[24,25] In the words of EMBL-EBI Director Ewan Birney, this is *the most important data set since the mapping of the Human Genome*.[26]

The success of this AI project seems to rely, among other things, on the availability of an extensive, high-quality, and structured database as well as on its scientific and engineering advancements.[27] This is not a trivial endeavor and one that requires not only great effort and coordination but also *ad hoc* funding. This is argued for in a recently published perspective by Campo et al. "Now Is the Time to Build a National Data Ecosystem for Materials Science and Chemistry Research Data":[28] *access to data and associated analysis software—cyber infrastructure—is critical to efficiency and productivity in an information-driven economy*. In short, the fast development of

https://doi.org/10.1021/acsomega.2c04373
ACS Omega 2022, 7, 25906−25908

AI seems to be going hand-in-hand with the availability of open-well structured data in high-quality databases, i.e., where metadata standards are adopted to enable interoperability, according to the "FAIR" data principles.[29]

In this collection (https://pubs.acs.org/page/vi/ACSomega_AI), we are proud to present a selection of articles that provide a quick pulse of the growing spectrum of topics (Figure 2) that are "experimenting" with AI tools. The topics vary from bioinformatics, chemometrics, and small-molecule design and synthesis to diagnostics in medicine as well as in physical and chemical processes, such as energy materials, metallurgy, wastewater treatment, and, finally, image analysis applied to agriculture. The ever-increasing volume of submissions and the diversity of the themes have encouraged us to start looking at Artificial Intelligence research, not just as an attractive future development but as a technique that is reaching a steady state of application and productivity in several fields.

Silvia Imberti, Publishing Editor, *ACS Omega* orcid.org/0000-0001-7037-6829

AUTHOR INFORMATION

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c04373

Notes

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

REFERENCES

(1) Directory of Open Access Journals. https://doaj.org. https://bit.ly/3LNlNpo [Accessed May 25, 2022].
(2) Athanasios, V.; Nikolaos, D.; Anastasios, D.; Eftychios, P. Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience* **2018**, No. 7068349, DOI: 10.1155/2018/7068349.
(3) Nassif, A. B.; Shahin, I.; Attili, I.; Azzeh, M.; Shaalan, K. Speech Recognition Using Deep Neural Networks: A Systematic Review. *IEEE Access* **2019**, *7*, 19143−19165.
(4) Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.; Le, Q. V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. DOI: 10.48550/arXiv.1906.08237.
(5) Chowdhery, A. et al.PaLM: Scaling Language Modeling with Pathways. DOI: 10.48550/arXiv.2204.02311.
(6) Sap, M.; LeBras, R.; Allaway, E.; Bhagavatula, C.; Lourie, N.; Rashkin, H.; Roof, B.; Smith, N. A.; Choi, Y. ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning. DOI: 10.48550/arXiv.1811.00146.
(7) The Allen Institute for Artificial Intelligence. https://mosaic.allenai.org/research [Accessed June 29, 2022].
(8) Brown, N.; Sandholm, T. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science* **2018**, *359* (6374), 418−424.
(9) Rao, Q.; Frtunikj, J. Deep learning for self-driving cars: chances and challenges. *Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems (SEFAIS '18)*; Association for Computing Machinery, New York, NY, USA; pp 35−38. DOI: 10.1145/3194085.3194087.
(10) Esteva, A.; Kuprel, B.; Novoa, R.; et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115−118.
(11) Poplin, R.; Varadarajan, A. V.; Blumer, K.; et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2018**, *2*, 158−164.
(12) Baum, Z. J.; Yu, X.; Ayala, P. Y.; Zhao, Y.; Watkins, S. P.; Zhou, Q. Artificial Intelligence in Chemistry: Current Trends and Future Directions. *J. Chem. Inf. Model.* **2021**, *61* (7), 3197−3212.
(13) Gartner, Inc. Gartner Identifies Key Emerging Technologies Spurring Innovation Through Trust, Growth and Change. https://www.gartner.com/en/newsroom/press-releases/2021-08-23-gartner-identifies-key-emerging-technologies-spurring-innovation-through-trust-growth-and-change [Accessed May 15, 2022].
(14) Wills, T. Assessing structural novelty of the first AI-designed drug candidates to go into human clinical trials. https://www.cas.org/resources/blog/ai-drug-candidates. [Accessed May 15, 2022]. "(···) medicinal chemists more than likely would have identified these molecules as potential drug candidates using traditional approaches based on existing scientific literature".
(15) Lowe, D. AI-Generated Clinical Candidates, So Far. https://www.science.org/content/blog-post/ai-generated-clinical-candidates-so-far [Accessed May 15, 2022].
(16) Peiretti, F.; Brunel, J. M. Artificial Intelligence: The Future for Organic Chemistry? *ACS Omega* **2018**, *3* (10), 13263−13266.
(17) Hammer, A. J. S.; Leonov, A. I.; Bell, N. L.; Cronin, L. Chemputation and the Standardization of Chemical Informatics. *JACS Au* **2021**, *1* (10), 1572−1587.
(18) Jones, C. W.; Lawal, W.; Xu, X. Emerging Chemistry & Machine Learning. *JACS Au* **2022**, *2* (3), 541−542.
(19) Jumper, J.; Evans, R.; Pritzel, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.
(20) AlphaFold: a solution to a 50-year-old grand challenge in biology. https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology [Accessed May 15, 2022].
(21) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235−242.
(22) Callaway, E. What's next for AlphaFold and the AI protein-folding revolution. DOI: 10.1038/d41586-022-00997-5 [Accessed May 15, 2022].
(23) Outeiral, C. Oxford Protein Informatics Group. AlphaFold 2 is here: what's behind the structure prediction miracle. https://www.blopig.com/blog/2021/07/alphafold-2-is-here-whats-behind-the-structure-prediction-miracle/ [Accessed May 15, 2022].
(24) Hatch, V. DeepMind and EMBL release the most complete database of predicted 3D structures of human proteins. https://www.ebi.ac.uk/about/news/announcements/alphafold-database-launch/ [Accessed May 15, 2022].
(25) Varadi, M.; et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50* (D1), D439−D444.
(26) Toews, R. AlphaFold Is The Most Important Achievement In AI—Ever. https://www.forbes.com/sites/robtoews/2021/10/03/alphafold-is-the-most-important-achievement-in-ai-ever/?sh=26cdd2606e0a [Accessed May 15, 2022].
(27) Lowe, D. Computational Happiness, Delayed. https://www.science.org/content/blog-post/computational-happiness-delayed [Accessed June 8, 2022].
(28) Campo, E. M.; Shankar, S.; Szalay, A. S.; Hanisch, R. J. Now Is the Time to Build a National Data Ecosystem for Materials Science and Chemistry Research Data. *ACS Omega* **2022**, *7* (16), 13398−13402.
(29) Wilkinson, M. D.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **2016**, *3*, 160018.