

Artificial Intelligence in Chemistry: Current Trends and Future Directions

Zachary J. Baum,* Xiang Yu, Philippe Y. Ayala, Yanan Zhao, Steven P. Watkins, and Qiongqiong Zhou



Cite This: *J. Chem. Inf. Model.* 2021, 61, 3197–3212



Read Online

ACCESS |

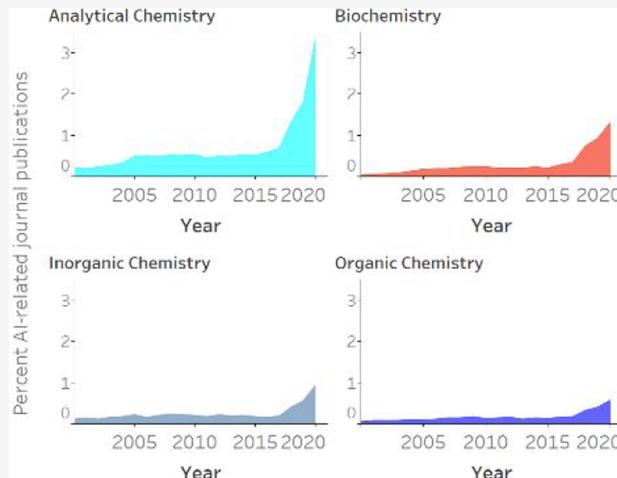
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The application of artificial intelligence (AI) to chemistry has grown tremendously in recent years. In this Review, we studied the growth and distribution of AI-related chemistry publications in the last two decades using the CAS Content Collection. The volume of both journal and patent publications have increased dramatically, especially since 2015. Study of the distribution of publications over various chemistry research areas revealed that analytical chemistry and biochemistry are integrating AI to the greatest extent and with the highest growth rates. We also investigated trends in interdisciplinary research and identified frequently occurring combinations of research areas in publications. Furthermore, topic analyses were conducted for journal and patent publications to illustrate emerging associations of AI with certain chemistry research topics. Notable publications in various chemistry disciplines were then evaluated and presented to highlight emerging use cases. Finally, the occurrence of different classes of substances and their roles in AI-related chemistry research were quantified, further detailing the popularity of AI adoption in the life sciences and analytical chemistry. In summary, this Review offers a broad overview of how AI has progressed in various fields of chemistry and aims to provide an understanding of its future directions.

KEYWORDS: *artificial intelligence, CAS Content Collection, analytical chemistry, biochemistry*



INTRODUCTION

Artificial intelligence (AI) refers to the ability of machines to act in seemingly intelligent ways, making decisions in response to new inputs without being explicitly programmed to do so. Whereas typical computer programs generate outputs according to explicit sets of instructions, AI systems are designed to use data-driven models to make predictions. These AI models are generally first trained on representative data sets with known output values, thereby “learning” input–output relationships. The resulting trained models can then be used to predict output values of data similar to the training set or to generate new data. Many problems involving data with complex input–output relationships are difficult or impractical to model procedurally, thus creating an opportunity for AI.

AI can feasibly be applied to various tasks in the field of chemistry, where complex relationships are often present in data sets. For example, the solubility of a new compound may be predicted either through equations based on empirical data or by using theoretical calculations. Alternatively, prediction of solubility may also be accomplished by an AI program that has developed structure–solubility relationships after being trained on numerous compounds with known solubilities. The use of AI for tasks, such as property prediction have proliferated in

recent years due to explosive growth in computing power, open-source machine-learning frameworks, and increasing data literacy among chemists.^{1–9} AI implementations have proven to dramatically reduce design and experimental effort by enabling laboratory automation,¹⁰ predicting bioactivities of new drugs,^{11–13} optimizing reaction conditions,¹⁴ and suggesting synthetic routes to complex target molecules.¹⁵

Although significant publicity has been given to AI and its application in chemistry, perspective on its use and development in chemistry is not obvious from the massive volume of available information. This Review uses the CAS Content Collection to contextualize the current AI landscape, classifying and quantifying chemistry publications related to AI from the years 2000–2020. The CAS Content Collection covers publications in 50 000 scientific journals from around the world in a wide range of disciplines, 62 patent authorities,

Received: June 1, 2021

Published: July 15, 2021



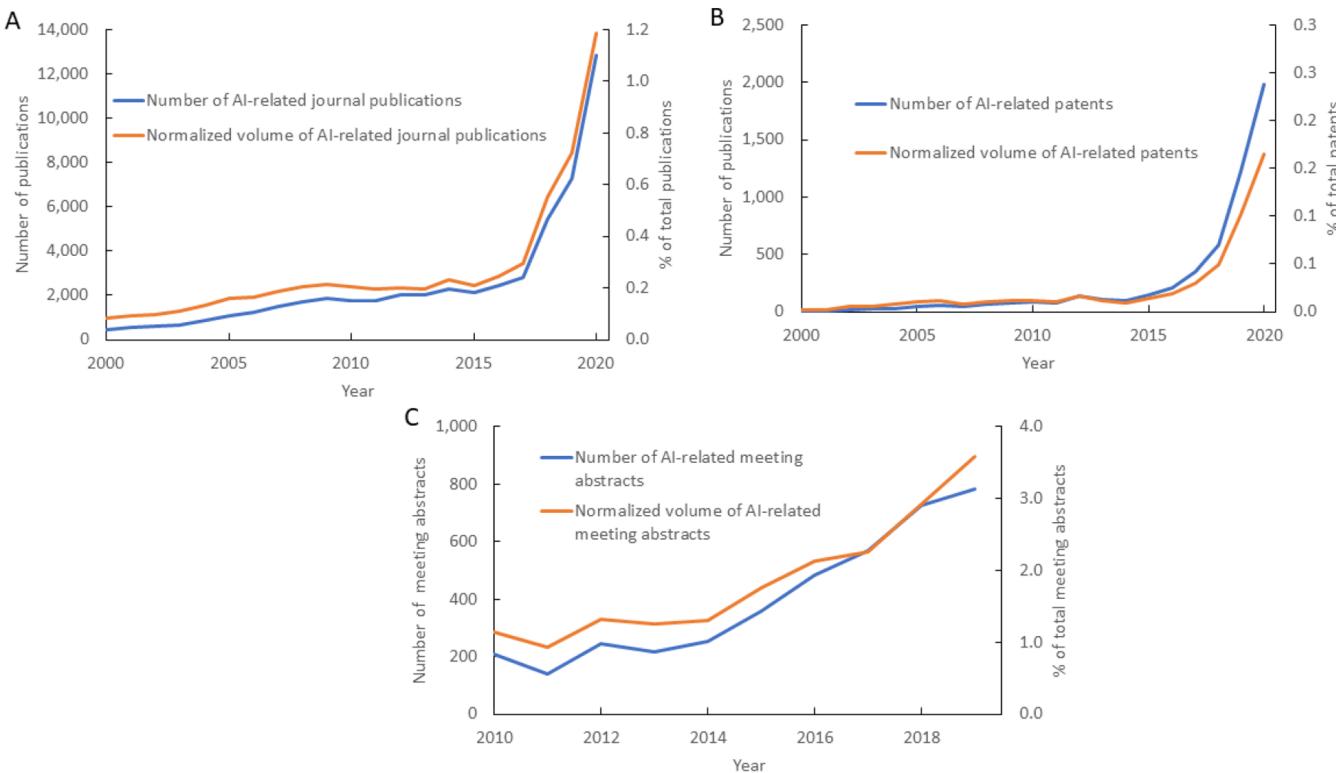


Figure 1. Annual publication volume in AI-related chemistry from 2000 to 2020: (A) Journal publications, (B) patent publications, and (C) ACS National Meeting abstracts.

and 2 defensive publications (Research Disclosures and IP.com).¹⁶ There are more than 1000 global scientists specialized in various scientific domains curating, analyzing, and connecting data from published sources at CAS. The CAS Content Collection, as one of the largest collections of scientific databases in the world, has many unique features and annotations added during data curation. Expert-curated CAS content is suitable for quantitative analysis of publications against variables, such as time, country, research area, and substance details. We first examine the growth and distribution of AI-related publications in chemistry, which includes the annual growth of publication volume and the distribution of publications among countries, organizations, and research areas, followed by a topic analysis revealing the evolution of frequently used concepts related to AI in chemistry. We then provide lists of notable AI-related journal and patent publications in a variety of research areas. Finally, we look at the types of chemical substances most frequently involved in the AI-related literature, highlighting the distribution of AI-related publications among various classes of substances and their roles. We hope that this Review can serve as a useful resource for those who would like to understand global trends in AI-oriented research efforts in chemistry.

GROWTH AND DISTRIBUTION OF PUBLICATION VOLUME IN AI-RELATED CHEMISTRY

Volume of Publications by Year. With the rapid growth in global research activity, scientific publication volume has steadily increased over the past 20 years. A quantitative analysis helps to understand just how fast chemistry publications using artificial intelligence are increasing relative to the increase in total chemistry publications. To this end, the

CAS Content Collection was searched to identify AI-related publications from 2000 to 2020 based on various AI terms in their title, keywords, abstract text, and CAS expert-curated concepts. The search query required screening of each term to minimize false positives due to polysemy; a maximum of a 2% false positive rate was allowed for each OR-delimited phrase, as determined by random screenings of 50–100 documents performed by CAS experts. In addition, matches on particularly problematic phrases, such as “brain” and “nerve” were excluded from consideration. The resulting search string is provided in the Supporting Information. From this search, roughly 70 000 journal publications and 17 500 patents from the CAS Content Collection were identified to be related to AI. Figure 1A and 1B shows the volume of these publications and their volume normalized by the overall number of journal publications or patents by year, respectively. Indeed, the numbers of both journal and patent publications increased with time, showing similar rapidly growing trends after 2015. This growth stems in part from the high-profile successes of deep learning projects in public data challenges starting around 2012, such as the Merck Molecular Activity Challenge¹⁷ and the ImageNet competition,¹⁸ which increasingly drew research interest from the scientific community. Additionally, the introduction of open-source machine learning frameworks, such as TensorFlow (2015) and PyTorch (2016), and the availability of increasingly powerful computing hardware sparked a global explosion in AI research, enabling further applications of AI to chemistry. In fact, as of 2020, over 50% of the documents on AI in chemistry were published during the past 4 years. Another way to measure recent scientific research trends is by examining scientific meeting abstracts. For this purpose, the abstracts from ACS National Meetings were analyzed for the

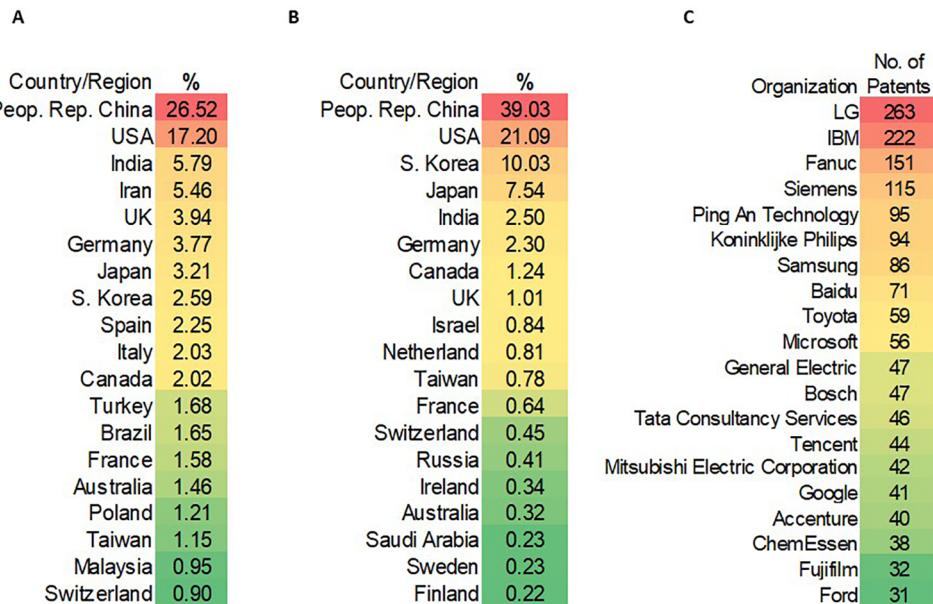


Figure 2. Distribution of AI-related publications by country/region and company from 2000 to 2020. (A) Top 20 countries/regions in number of journal publications. (B) Top 20 countries/regions in number of patent publications. (C) Top 20 companies in number of patent publications.

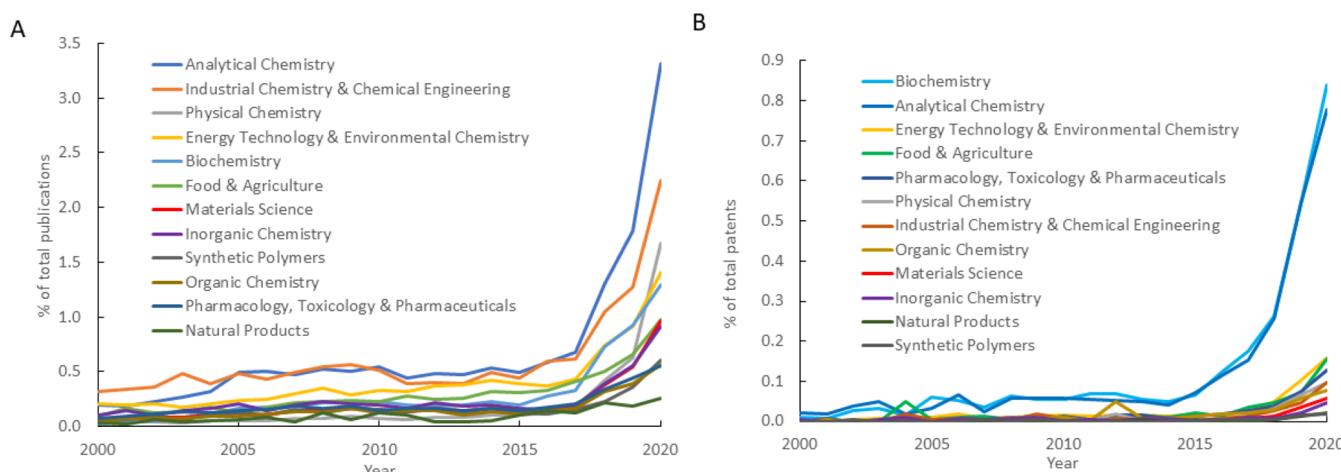


Figure 3. Publication trends of AI in specific research areas from 2000 to 2020: (A) journal publications and (B) patent publications.

presence of AI topics, and the number of AI-related abstracts per year and its amount relative to the total number of yearly abstracts are shown in Figure 1C. The abstract publications show similar behavior to the trends in journal and patent publication. These analyses suggest that not only has there been an absolute increase of research effort toward AI in chemistry but also that the proportion of AI-related research is increasing.

Distribution of AI-Related Publications by Country/Region and Company. The countries/regions and organizations of origin for AI-related chemistry documents were then extracted to determine their distributions. Figure 2A and 2B shows the percentages of AI-related journal articles and patents produced in selected countries/regions and by selected organizations in the years 2000–2020, respectively, with the top commercial patent assignees listed in Figure 2C. China and the United States contributed the largest numbers of publications for both journal articles and patents. Medical diagnostic developers and technology companies make up a

large portion of the commercial patent assignees for AI chemical research. These companies rely on AI for automation, control, and optimization of a variety of processes, such as semiconductor device fabrication and biomarker screening, which will be explored in more detail in the following sections.

DISTRIBUTION OF AI-RELATED CHEMISTRY PUBLICATIONS BY RESEARCH AREA

Trends of Publications in Specific Research Areas. To have a closer look at how AI is involved in different chemistry-related research areas, the roughly 70 000 journal and 17 500 patent publications were further classified into the following 12 categories by CAS experts: Analytical Chemistry, Biochemistry, Energy Technology and Environmental Chemistry, Food and Agriculture, Industrial Chemistry and Chemical Engineering, Inorganic Chemistry, Materials Science, Natural Products, Organic Chemistry, Physical Chemistry, Synthetic Polymers, and Pharmacology, Toxicology and Pharmaceuticals. The numbers of AI-related publications in each area are normalized

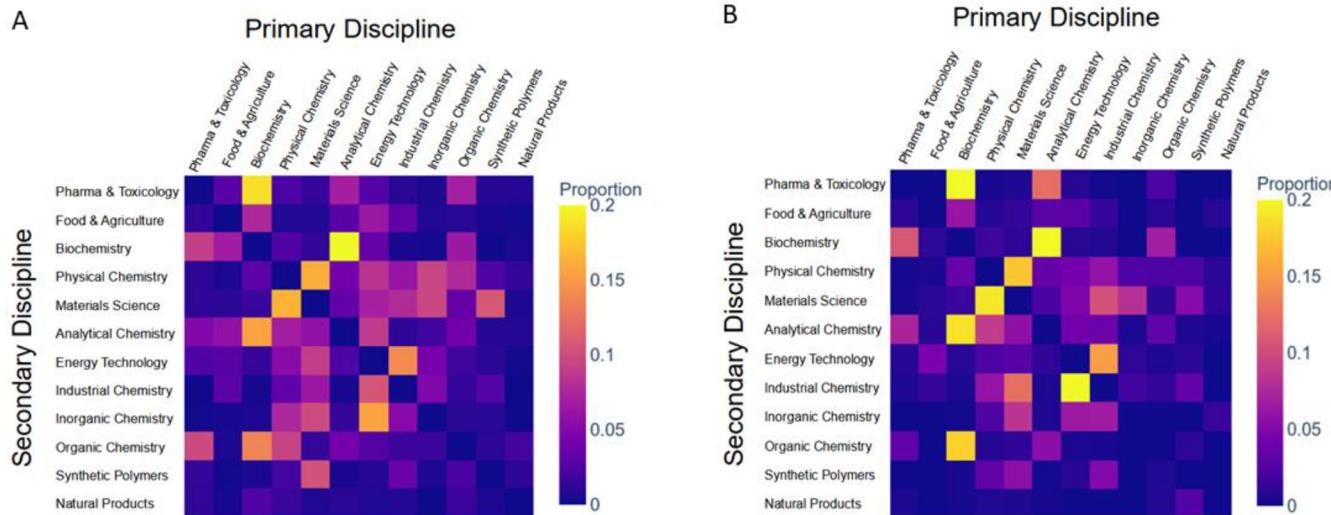


Figure 4. Relative prevalence of interdisciplinary studies in AI-related scientific publications: (A) journal publications and (B) patent publications. Columns denote primary research areas, rows denote secondary research areas, and each square denotes an interdisciplinary pair of primary and secondary research areas.

to that area's respective total yearly publication volume and shown in Figure 3A (journal publications) and Figure 3B (patents). The absolute numbers of journal publications in each area are shown in Figure S1. Among all these specific chemistry-related areas, documents in Analytical Chemistry (both journal and patent publications) have the highest normalized volume in the most recent 10 years; it has also risen steeply in the last 5 years. Energy Technology and Environmental Chemistry and Industrial Chemistry and Chemical Engineering are the two research areas ranked in the second tier in terms of proportion of research volume and momentum in journal publications (Figure 3A). Interestingly, while Biochemistry is among the fields most represented in AI-related patent publications, its proportion in journal publications is relatively moderate when compared to other research areas. This indicates a strong desire or incentive to patent AI technologies in biochemistry, possibly because of its use in drug research and development.

Relative Prevalence of Interdisciplinary Research in Specific Areas. Innovations in science and technology are often made by finding connections between multiple research areas to derive novel insights, methods, and products. A theoretical method developed for molecular dynamics, for example, may be applied to study the interaction of a ligand with a protein, which can in turn be used to predict the activity of a drug. Conversely, data collected from experimental measurements can be used to optimize parameters of theoretical simulations. With such continuous conversations, fields not traditionally associated with each other can be mutually informative. Interdisciplinary effects such as these are also present in the AI-related chemical literature, which we explore here in detail.

From the set of AI-related journal and patent publications, we identified approximately 15 000 and 3000 interdisciplinary journal articles and patents, respectively. CAS analysts determined the primary disciplines for each document and any secondary disciplines, which also contributed to the work. The resulting combinations of primary and secondary disciplines are summarized in Figure 4, which are both

normalized to the total number of interdisciplinary documents containing each respective primary and secondary disciplines.

In Figure 4, several relationships are apparent among chemical disciplines. In journal articles, the strongest correlations are observed between primary and secondary research areas in Analytical Chemistry and Biochemistry, in Materials Science and Physical Chemistry, and in Biochemistry with applications to Pharmacology, Toxicology and Pharmaceuticals (Figure 4A). In patents (Figure 4B), the trend is similar, but inventions in Energy Technology and Environmental Chemistry related to Industrial Chemistry and Chemical Engineering also show prominently. For example, journal documents using analytical chemistry techniques such as mass spectrometry and nuclear magnetic resonance, infrared, and Raman spectroscopies are augmented with machine learning for use in medical diagnostics,^{19–28} studies of metabolomics,^{29–36} and microbial identification,^{37–40} while biochemistry-related analytical chemistry patents concentrate on the development of analytical devices and methods for use in similar studies.^{41–53} AI-related journal documents with interest in Materials Science and Physical Chemistry discuss topics, such as the evaluation of structure–property relationships in materials by augmenting first-principles calculations with machine learning models,^{54–60} using data from high-throughput experimentation to optimize the properties of functional materials,^{61–68} and the use of published data to enable the discovery of new materials.^{69–74} In patents, the combination of AI, Materials Science, and Physical Chemistry is used in methods for improving semiconductor device fabrication^{75–86} and polymer performance.^{87–90} Additionally, AI is being used in Biochemistry–Pharmacology Toxicology and Pharmaceuticals research to understand drug–biomolecule interactions,^{91–100} apply biomarker data to the prediction of drug activities,^{101–106} and model toxicity.^{107–111} Finally, patents in Energy Technology and Environmental Chemistry related to Industrial Chemistry and Chemical Engineering are often using AI in control systems for fuel production^{112–119} and engines.^{120–125} These examples demonstrate how AI can be applied in research areas where the relationships between

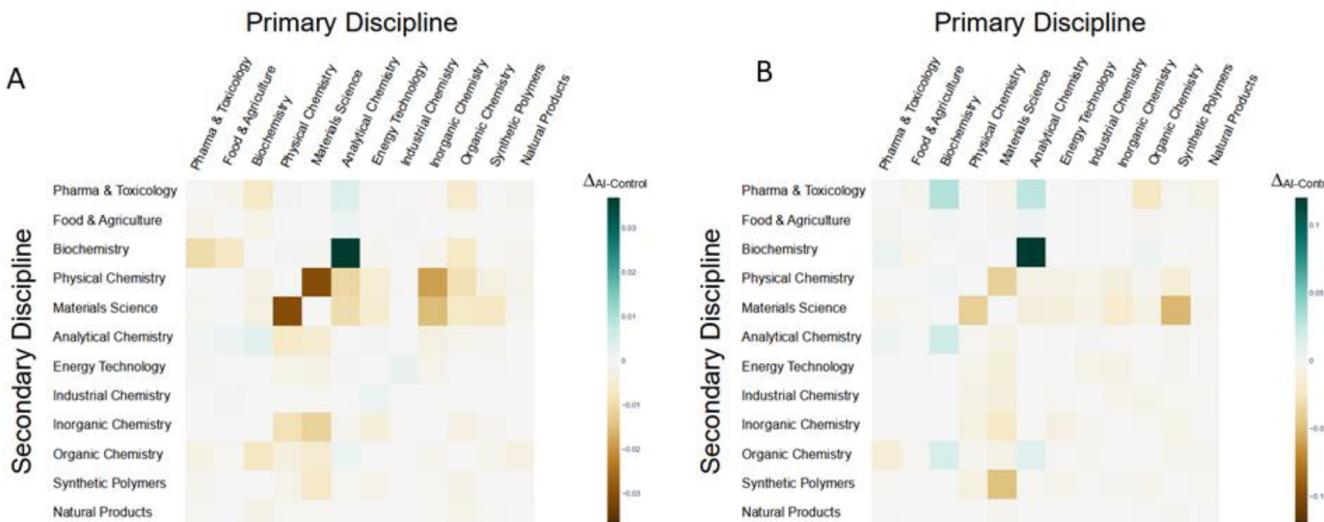


Figure 5. Difference in proportion of total AI-related publications and control group (non-AI related) by interdisciplinary pair: (A) journal publications and (B) patent publications.

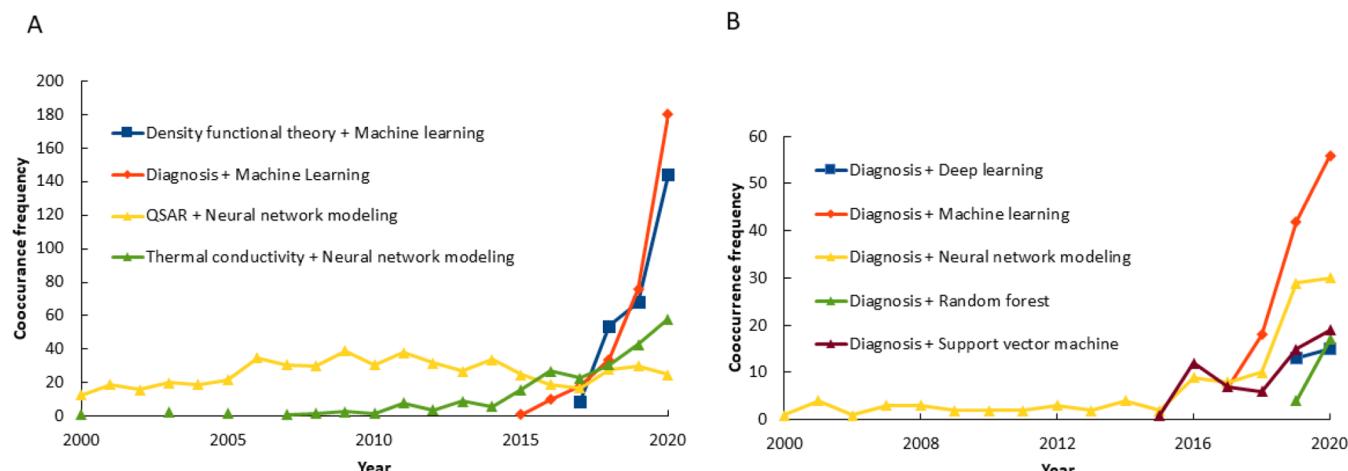


Figure 6. Trends of co-occurrence in scientific publications for selected research topics and AI algorithms: (A) journal publications and (B) patent publications.

available data in separate domains are not obvious to researchers.

While interdisciplinary relationships do appear in AI-related chemistry research, it is natural to question the extent to which AI is indeed facilitating connections between fields. To answer this, we first selected random control groups of journal ($n = 81\,601$) and patent ($n = 12\,181$) publications and identified sets of interdisciplinary documents ($n = 32\,097$ and $n = 4426$, respectively) using the same 12 research areas. In both the AI and control groups, we then calculated the proportions of documents belonging to each primary–secondary discipline pair. By comparing the corresponding proportions in these two groups, the resulting difference maps (Figure 5) reveal how AI is bringing disciplines together (positive values) and areas in which the use of AI is lagging (negative values). Notably, these maps show that interdisciplinary biochemical–analytical research is greatly facilitated by AI, and that despite recent advances, Physical Chemistry and Materials Science have used AI techniques less than other chemistry fields in the period 2000–2020. This lag is observed despite the relatively high proportion of interdisciplinary Physical Chemistry–Materials

Science documents seen in Figure 4. This may be attributable to the reliance of Materials Science on Physical Chemistry principles and techniques—the incidence of publications at this intersection is high even in the absence of AI. However, the use of AI in interdisciplinary research is still maturing; a similar analysis was done with a more recent time window (2016–2020), where AI’s capability in bringing disciplines together seems to be increasing (Figure S2).

■ EVOLUTION OF RESEARCH TOPICS IN AI-RELATED CHEMISTRY PUBLICATIONS

Topic Analysis in Journal Article Publications. By analyzing the connections of CAS-indexed concepts over time, one can see when a research topic became potentially addressable using AI techniques. Figure S3 shows the most frequently co-occurring concepts and the number of documents in which they co-occur (presented at a 97.5th percentile cutoff for co-occurrence). For the years 2000–2004 (Figure S3A), we see only a few concepts connected to the concepts Neural network modeling and Algorithms. Several biochemistry-related concepts that appear in conjunction with AI

Table 1. Notable AI-Related Journal Publications in the Areas of Biochemistry and Pharmacology, Toxicology and Pharmaceuticals

journal	title	organization	year of publication	highlight
<i>Nature</i>	Improved Protein Structure Prediction Using Potentials from Deep Learning ¹²⁵	DeepMind Technologies Ltd., UK	2020	neural network, AlphaFold, distance between amino acid residues, potential of mean force, protein structure prediction
<i>Cell</i>	A Deep Learning Approach to Antibiotic Discovery ¹²⁷	Massachusetts Institute of Technology, US	2020	deep neural network, property-structure correlation, molecule prediction, compound screening, empirical test, halicin
<i>Nature Biotechnology</i>	SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks ¹²⁸	Technical University of Denmark, Denmark	2019	deep recurrent neural network combined with conditional random field classification and transfer learning, signal peptide prediction across all organisms
<i>Nature Biotechnology</i>	Determining Cell Type Abundance and Expression from Bulk Tissues with Digital Cytometry ¹²⁹	Stanford University, US	2019	CIBERSORTX, single-cell RNA-seqencing
<i>Cell</i>	Predicting Splicing from Primary Sequence with Deep Learning ¹³⁰	Illumina Inc., US	2019	prediction of splice junctions from pre-mRNA transcript sequence, splice-altering consequence, pathogenic mutations
<i>Cell</i>	Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity ¹³¹	Massachusetts General Hospital and Harvard Medical School, US	2019	single-cell RNA sequencing and genotyping, acute myeloid leukemia
<i>Nature</i>	DNA Methylation-Based Classification of Central Nervous System Tumors ¹³²	Hopp Children's Cancer Center at the NCT Heidelberg (KITZ), Germany	2018	machine learning, DNA methylation-based tumor classification
<i>ACS Central Science</i>	Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks	Westfälische Wilhelms-Universität Münster, Germany	2018	recurrent neural network, drug molecule generation, drug discovery
<i>Nucleic Acids Research</i>	BepiPred-2.0: Improving Sequence-Based B-Cell Epitope Prediction Using Conformational Epitopes ¹³⁴	Technical University of Denmark, Denmark	2017	random forest algorithm, BepiPred-2.0 web server, sequence-based B-cell epitope prediction
<i>Cell</i>	A Landscape of Pharmacogenomic Interactions in Cancer ¹³⁵	Wellcome Trust Sanger Institute, UK	2016	machine learning, correlating oncogenic alterations (somatic mutations, copy number alterations, and hypermethylation) with drug sensitivity in human cancer cell lines
<i>Cell</i>	Personalized Nutrition by Prediction of Glycemic Responses ¹³⁶	Weizmann Institute of Science, Israel	2015	machine learning, personalized postprandial blood glucose level prediction, therapeutic food intervention
<i>Nature Genetics</i>	A General Framework for Estimating the Relative Pathogenicity of Human Genetic Variants ¹³⁷	University of Washington, Seattle, US	2014	combined annotation-dependent depletion, integrated annotation method of human genetic variants
<i>Nature</i>	Persistent Gut Microbiota Immaturity in Malnourished Bangladeshi Children ¹³⁸	Washington University in St. Louis, US	2014	machine learning, 16S rRNA, correlation of microbiota immaturity index with human malnourished state

concepts are proteins, protein sequences, and protein conformation.

In the years 2005–2009 (Figure S3B), *Homo sapiens* become a more popular topic because of the increasing AI-related effort in disease diagnosis and prognosis, and related concepts, such as biomarkers, tumor markers, prognosis, and diagnosis start to appear during this period. Protein-associated concepts, such as protein motifs, protein–protein interactions, secondary structure, and amino acids, became more prevalent in AI-related documents, likely because of the use of AI in solving high-resolution protein structures. Genetics-associated concepts, such as sequence annotation and gene expression profile, were also indexed more frequently. Finally, high-throughput screening and proteomics were frequently used concepts during this period.

In the years 2010–2014 (Figure S3C), genome-related concepts, such as genome and single nucleotide polymorphism, were more often studied using AI methods. The application of AI to pharmaceutical and biomedical fields became more common, as the concepts drug discovery, drug design, blood analysis, neoplasm, and microRNA were frequently used. The use of AI techniques for environmental remediation is evidenced by the occurrence of concepts, such as absorptive wastewater treatment and Chemical oxygen demand in this period.

In the years 2015–2019 (Figure S3D), the use of AI becomes more prominent in research topics, such as DNA methylation, mutation, nanofluids, heat transfer, and biodiesel fuel to solve problems in those research areas. AI also appeared frequently in publications related to cancer and Alzheimer's disease. Since the beginning of 2020, when the critical need for research into COVID-19 became apparent, AI has been used frequently in the areas of drug discovery, disease diagnosis, and disease tracking (Figure S4).

Quantifying the co-occurrence of use case-specific concepts with AI-related concepts over time further reveals the progression of AI adoption. As Figure 6A shows, studies of QSAR (quantitative structure–activity relationships), a perennial topic in drug discovery research, have employed Neural network models consistently for some time. On the other hand, in Materials Science-related topics, such as thermal conductivity, the use of neural network modeling has grown more slowly, with its use in publications not increasing rapidly until the second half of the 2010s. The use of machine learning in topics such as medical diagnosis and Density functional theory has only recently begun to increase significantly.

Topic Analysis in Patent Publications. Frequently co-occurring concepts were also identified in the patent literature in 5-year time windows (Figure S5, presented at a 95th percentile cutoff for co-occurrence). Similar patterns in the evolution of associated concepts were observed in the patent literature as those observed in the journal literature. Previously unseen research topics, such as Diagnosis, Prognosis, Peptides, and Transcription factors, were introduced in the years 2005–2009 (Figure S5B). The use of AI in the study of organic compounds and hydrocarbons and in the development of QSPR (quantitative structure–property relationships) becomes more prominent in the years 2010–2014 (Figure S5C), and connections between these topics and AI-related concepts have increased further since 2015 (Figure S5D). However, unlike in journal publications, few COVID-related concepts co-occurred with AI-related concepts in the patent literature in 2020 (Figure S6). This may be due in part to the

longer turnover time in the patent application process compared to scientific journal publication.

It is telling to examine the progression of the concept diagnosis with various AI concepts. While the growth of documents associating diagnosis with the concept Neural network modeling is unsubstantial between 2000 and 2015, the number of documents associating diagnosis with various AI concepts increases rapidly after 2015, with the concepts of deep learning, random forest, and support vector machine seeing significant usage (Figure 6B). This pattern is consistent across a variety of topics in chemistry, in which the increase in usage of AI after 2015 is general rather than being limited to a single AI methodology.

■ NOTABLE AI-RELATED JOURNAL AND PATENT PUBLICATIONS

To highlight the most influential journal publications using AI in chemistry, a bibliometric analysis was performed in the primary literature from our search query since 2014. Publications with over 100 citations were selected and further classified into groups of related research areas; then, they were reviewed and selected based on apparent novelty: Biochemistry and Pharmacology, Toxicology and Pharmaceuticals (Table 1), Materials Science (Table 2), and Analytical Chemistry, Synthetic Chemistry, and Physical Chemistry (Table 3). The US is the leading country of origin: 15 of the 34 papers in Tables 1–3 are affiliated with US organizations. Other countries with significant numbers of important AI documents are Germany (6) and Switzerland (5). Among organizations, the Massachusetts Institute of Technology (US) and the University of Basel (Switzerland) were the two biggest contributors. Three commercial organizations, DeepMind Technologies, Ltd. (UK), Illumina Inc. (US), and Intel Corp. (US) contributed significantly.

Among these 34 journal papers, the most frequently indexed concepts are Machine learning, Neural network, Deep learning, Density functional theory, and Random forest. In Biochemistry and Pharmacology and Toxicology and Pharmaceuticals (Table 1), many of the articles apply AI technology to research topics involving high-throughput drug screening, nucleic acid sequence analysis, and protein structure prediction. Publications in Materials Science research (Table 2) reported AI-driven structure–property relationship predictions enabling the discovery of new functional materials as well as memristors with applications in neuromorphic computing. In Analytical Chemistry, Synthetic Chemistry and Physical Chemistry (Table 3), new methods were developed with AI to complement analytical data, automate flow chemistry, improve retrosynthetic planning, and predict reaction outcomes. In addition, user-friendly computational tools were developed, and methods combining AI with physics-based approaches such as density functional theory were reported to improve the accuracy of calculations.

To identify notable AI-related patents, results from the search query were first sorted by size of patent family for each year. A patent family is a collection of patents filed in multiple countries covering the same or similar content¹⁶⁰ and, thus, represents high-priority intellectual property for organizations, which we use here as a proxy to estimate importance. Documents were then selected from the top 50 largest patent families per year on apparent novelty and relevance to overall research trends of AI in chemistry and presented in Table S1. The US has made the largest contribution to these patent

Table 2. Notable AI-Related Journal Publications in Materials Science

journal	title	organization	year	highlight
<i>Nature</i>	Accelerated Discovery of CO ₂ Electrocatalysts Using Active Machine Learning ¹³⁹	University of Toronto, Canada	2020	machine learning, density functional theory, CO adsorption energy prediction, surface and adsorption site screening in Cu-containing intermetallic crystals, new electrocatalysts
<i>Nature Materials</i>	Scalable Energy-Efficient Magnetoelectric Spin–Orbit Logic ¹⁴⁰	Intel Corp, US	2019	scalable logic device based on quantum materials, electron's angular-linear momentum transduction with magnetoelectric switching
<i>Journal of the American Chemical Society</i>	Ionic Modulation and Ionic Coupling Effects in MoS ₂ Devices for Neuromorphic Computing ¹⁴¹	University of Michigan, US	2019	logic device based on MoS ₂ film, local phase transition controlled by ion migration
<i>Physical Review Letters</i>	Accelerated Discovery of Organic Polymer Photocatalysts for Hydrogen Evolution from Water through the Integration of Experiment and Theory ¹⁴²	University of Liverpool, UK	2019	machine learning, correlation of sacrificial hydrogen evolution rate with four predicted molecular properties in conjugated polymers, new photocatalyst discovery
<i>Nature Materials</i>	Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties ¹⁴³	Massachusetts Institute of Technology, US	2018	convolutional neural network, density functional theory calculation, property-crystal graph correlation, crystalline material design
<i>Nature Materials</i>	A Non-Volatile Organic Electrochemical Device as a Low-Voltage Artificial Synapse for Neuromorphic Computing ¹⁴⁴	Stanford University, US	2017	organic memristor, polymer material
<i>Nature Materials</i>	Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach ¹⁴⁵	Harvard University, US	2016	combining time-dependent density functional theory with machine learning to predict electroluminescent molecules with high quantum efficiency
<i>Nature</i>	Machine-Learning-Assisted Materials Discovery Using Failed Experiments ¹⁴⁶	Haverford College, US	2016	cheminformatics utilized for hydrothermal synthesis of organic–inorganic hybrid materials via support vector machine-derived interpretable decision trees

inventions, with 13 of the 15 patents selected granted to companies based in the United States. Interestingly, most of the patent assignees are startup companies founded in the past 10 years. This is consistent with the rapid growth of AI-related chemistry inventions since 2015 and indicates how the emerging paradigm of AI provides opportunities for innovative enterprises.

The adoption of AI in the life sciences is prominent, comprising 8 of the 15 patents covering biomarker development, gene expression profiling, and biosequence analysis (**Table S1**). These patents also reflect a strong interest in applying machine learning to medical diagnostics, consistent with the topic analysis in **Figure SSD**. The remaining 7 patents cover research areas, including Analytical Chemistry, Environmental Chemistry, Materials Science, Industrial Chemistry & Chemical Engineering, and Synthetic Chemistry.

DISTRIBUTION OF SUBSTANCE INFORMATION IN AI-RELATED CHEMICAL LITERATURE

Journal Publications by Substance Class. The distribution of AI-related research activity can also be probed by studying the numbers of documents involving different types of substances. Because the barriers to AI implementation in chemistry include challenges in substance representation⁶ and data availability,⁷ enumeration of the most common substance types studied in the literature will point to areas in which researchers have, in some instances, been able to overcome such challenges. The substances indexed by CAS are categorized into multiple classes. The numbers of AI-related journal publications for some frequently occurring substance classes, namely Alloy, Coordination Compound, Element, Manual Registration, Ring Parent, Small Molecule, Polymer, Salt, and Inorganic Compound, are shown in **Figure 7A**. Substances in the Manual Registration class are predominantly biomolecules, such as enzymes, hormones, vaccines, and antibodies. Biosequences are not included in the analysis of journal documents. Ring Parents represent scaffolds defining the composition and connectivity of molecular ring systems.

As **Figure 7A** shows, publications containing Small Molecule substances are the highest in number, followed by those containing Element and Manual Registration substances, far outnumbering publications containing substances in the remaining classes. The high volume of research and invention in AI involving these classes is likely facilitated by their relative simplicity and ease of modeling compared to substances in other classes, such as Coordination Compound and Polymer. The large number of documents containing Manual Registration substances in **Figure 7A** is consistent with the relatively high publication volume in Biochemistry (**Figure S1**). Also shown in **Figure 7A** are the total numbers of substances contained in AI-related journal publications for each substance class. The data show similar trends as those for document count, albeit skewed by the larger number of small molecule substances per document.

Figure 7B shows the change in the number of AI-related journal publications by substance class for the years 2000–2020. While the document count of each substance type increased during this period (and particularly after 2017), those containing Small Molecule, Element and Manual Registration substances displayed the largest increases, consistent with the data shown in **Figure 7A**.

Patent Publications by Substance Class. The patent literature was analyzed using the same methods as for the

Table 3. Notable AI-Related Journal Publications in the Areas of Analytical Chemistry, Synthetic Chemistry, and Physical Chemistry

journal	title	organization	year	highlight
Science	Global Threat of Arsenic in Groundwater ¹⁴⁷	Swiss Federal Institute of Aquatic Science and Technology, Switzerland	2020	random forest model classifying high-risk population areas
Nature Methods	Deep Learning Enables Cross-Modality Super-Resolution in Fluorescence Microscopy ¹⁴⁸	University of California, Los Angeles, US	2019	deep learning, fluorescence microscopy
Bioinformatics	Trainable Weka Segmentation: A Machine Learning Tool for Microscopy Pixel Classification ¹⁴⁹	Ikerbasque-Basque Foundation for Science, Spain	2017	unsupervised machine learning, segmentation, clustering
Nature Methods	The Perseus Computational Platform for Comprehensive Analysis of (Prote)omics Data ¹⁵⁰	Max Planck Institute of Biochemistry, Germany	2016	connecting proteomics researchers with bioinformatics tools leveraging machine learning
Science	A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning ¹⁵¹	Massachusetts Institute of Technology, US	2019	automated synthesis enabled by AI
Nature	Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI ¹⁵²	Westfälische Wilhelms-Universität Münster, Germany	2018	deep neural network, Monte Carlo tree search, expansion policy network, filter network, retrosynthesis
Science	Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning ¹⁵³	Princeton University, US	2018	random forest, high-throughput experimentation, reaction outcome prediction
Journal of Chemical Theory and Computation	PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges ¹⁵⁴	University of Basel, Switzerland	2019	neural network, property prediction, using models trained on small peptide fragments to predict protein properties
Science Advances	Machine Learning of Accurate Energy-Conserving Molecular Force Fields ¹⁵⁵	Technische Universität Berlin, Germany	2017	gradient-domain machine learning, ab initio molecular dynamics
Science	Solving the Quantum Many-Body Problem with Artificial Neural Networks ¹⁵⁶	ETH Zurich, Switzerland	2017	machine learning of wave functions to solve many-body problem in quantum physics
Nature Communications	Quantum-Chemical Insights from Deep Tensor Neural Networks ¹⁵⁷	Technische Universität Berlin, Germany	2017	deep tensor neural network, quantum chemistry, intermediate size molecule property prediction
Journal of Chemical Theory and Computation	Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error ¹⁵⁸	University of Basel, Switzerland	2017	machine learning, molecule property prediction more accurate than DFT
Journal of Chemical Theory and Computation	Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach ¹⁵⁹	University of Basel, Switzerland	2015	machine learning, organic molecule thermochemical properties prediction

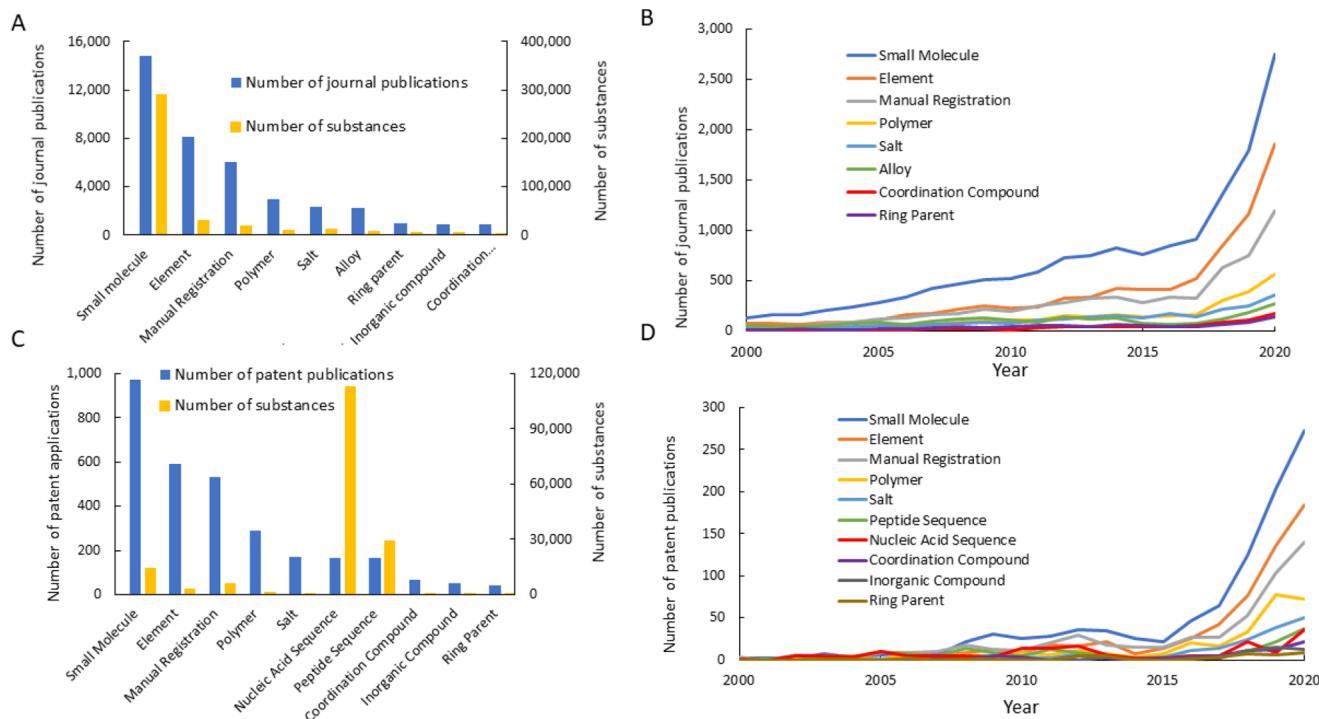


Figure 7. Publications in AI-related chemistry associated with substance class from 2000 to 2020. (A) Number of AI-related journal publications and number of substances associated with each class. (B) Trends of AI-related journal publications associated with each substance class. (C) Number of AI-related patent publications and number of substances associated with each class. (D) Trends of AI-related patent publications associated with each substance class.

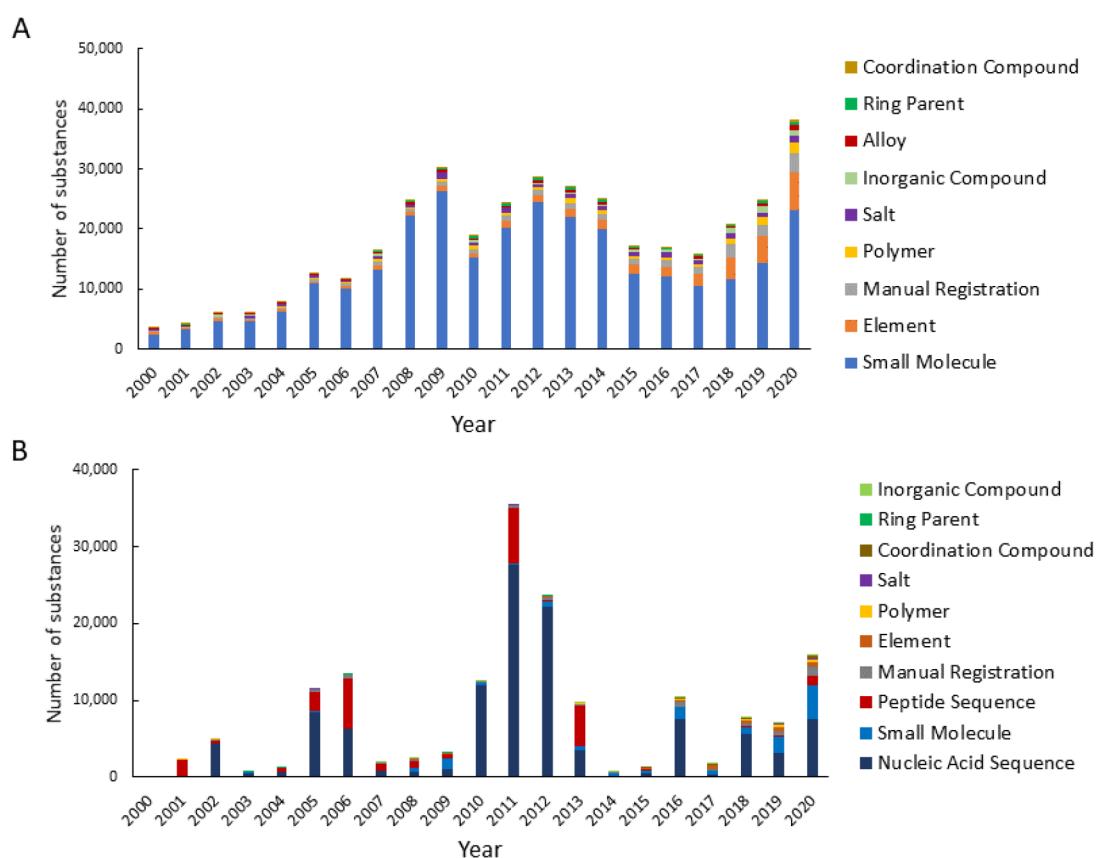


Figure 8. Trends of substance classes in AI-related chemistry publications from 2000 to 2020: (A) journal publications and (B) patent publications.

journal literature. Figure 7C shows the numbers of AI-related patent publications and substances associated with different substance classes. Nucleic Acid Sequences and Peptide Sequences are highest in number, whereas the remaining relative document and substance counts are similar to those found in Figure 7A. Patents containing Peptide Sequences or Nucleic Acid Sequences often contain large numbers of sequences per document, often far greater than other substances per patent. The change in the number of AI-related patent publications containing various substance classes over time is shown in Figure 7D, again showing trends consistent with those in Figure 7B.

Analysis of Substances Contained in AI-Related Chemical Literature. A substance-level perspective of chemical research over time is also useful for understanding the utilization of AI. It is interesting to see that in both journal and patent publications (Figure 8A and 8B, respectively), the number of substances present do not follow a monotonic increase over time, as was the case in the progression of total research volume. Rather, a lull in substance count can be seen in the first half of the 2010s before catching up with the massive increase in publications. This may be partially due to a small number of documents between the years of 2008–2014 containing large amounts of Small Molecule substances or biosequences on the order of 10^3 – 10^4 , which sometimes can be seen in the literature. We have also studied the distribution of these substances across a variety of role indicators, which are controlled vocabulary terms that describe the use of a substance within the context of a specific document (Table S2, Figure S7).

CONCLUSIONS AND OUTLOOK

Applications of AI in chemistry have become increasingly popular in recent years, as evidenced by the strong growth in publication volume. Yet, it is striking that growth has not been uniform. For some fields of chemistry, AI is much further along the proverbial Hype Cycle of Emerging Technologies¹⁶¹ than others. In life-sciences and Analytical Chemistry, for example, AI-adoption is likely already past the so-called “peak of inflated expectations” and “trough of disillusionment”. The utility of AI in a given domain is intrinsically linked to the quantity and quality of its data, as well as opportunities to gain insights from its analysis. AI can help gain insights that would not otherwise follow from established knowledge. AI is also useful for extracting insights from large intractable data sets, as well as aiding in the automation of repetitive tasks. With this in mind, it is not a surprise to see a surge in AI deployment within analytical chemistry, where large training sets are readily obtained, or in biochemistry, which contains a wealth of data for macromolecules whose structure–property relationships are not obvious to researchers. Successes in these more traditionally data-intensive fields are now being emulated in other areas of chemistry.

The large numbers and rapid growth of AI-related chemistry publications involving small molecules reflect the popularity of AI applications in drug discovery. Analyses of total substance numbers for each class in AI-related publications revealed large numbers of Nucleic Acid Sequences and Peptide Sequences in patents, consistent with the prevalence of AI applications in biochemistry. The distribution of the role indicators assigned to substances in AI-related publications contextualizes how AI is being used in recent biochemical and pharmaceutical research.

Multiple factors likely explain the significantly increased use of AI in chemistry after 2015. The greater availability of software and hardware tools to implement AI decreased the barriers to using it in chemical research, while research area-specific data sets amenable to AI methods have proliferated. In addition, many researchers have learned techniques in generating and handling data for use in AI methods.

Between the years 2000–2020, the co-occurrence of AI- and research area-specific concepts in publications shows how AI has been incorporated into a variety of research areas. Many AI methods have been adapted for chemistry research and are being further introduced to new areas of chemical study.

In conclusion, thanks to an increasingly interdisciplinary research landscape, many AI methods have been successfully adapted to chemistry research. Use of AI has even become routine in some fields. There are still areas of Chemistry like organic synthetic chemistry where AI is yet to make an impact. Perhaps, it is a matter of time before improvements in AI itself, lessons from successful applications of AI, and interdisciplinary research combine to help lift these areas out of the “trough of disillusionment” and onto the “plateau of productivity”.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00619>.

Total AI-related journal publications by discipline, difference in proportion of interdisciplinary journal publications from 2016 to 2020, evolution of co-occurring concepts, search string used for retrieval of all publications, table of notable AI-related patents, and common substance role indicators ([PDF](#))

AUTHOR INFORMATION

Corresponding Author

Zachary J. Baum — *Chemical Abstracts Service, Columbus, Ohio 43210, United States; [orcid.org/0000-0002-0585-8503](#); Email: ZBaum@cas.org*

Authors

Xiang Yu — *Chemical Abstracts Service, Columbus, Ohio 43210, United States*

Philippe Y. Ayala — *Chemical Abstracts Service, Columbus, Ohio 43210, United States*

Yanan Zhao — *Chemical Abstracts Service, Columbus, Ohio 43210, United States*

Steven P. Watkins — *Chemical Abstracts Service, Columbus, Ohio 43210, United States*

Qiongqiong Zhou — *Chemical Abstracts Service, Columbus, Ohio 43210, United States; [orcid.org/0000-0001-6711-369X](#)*

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.1c00619>

Notes

The authors declare no competing financial interest.

Publications using artificial intelligence were identified by optimizing a search of relevant terms on the CAS Content Collection using CAS STN. While the full data set is considered proprietary by CAS, the search string used for retrieval is included in the Supporting Information. Substance information, primary and secondary disciplines, concepts, and

institutional information were extracted directly from the CAS Content Collection.

ACKNOWLEDGMENTS

We sincerely appreciate Rumiana Tenchov's assistance curating references, Joshua Blair for obtaining ACS National Meeting abstracts, Laura Czuba for project coordination, Peter Jap and Cristina Tomeo for insightful discussion, and Susan Jersey and Robert Bird for proofreading. We are also grateful to Manuel Guzman, Gilles Georges, Michael Dennis, Carmin Gade, Dawn George, and Cynthia Casebolt for executive sponsorship.

REFERENCES

- (1) Dutta, S. Data Modeling: A Fundamental Pillar of Your Future Ai Technology. *CAS Blog*. <https://www.cas.org/resource/blog/data-modeling-fundamental-pillar-your-future-ai-technology> (accessed May 13, 2021).
- (2) Villalba, M.; Wollenhaupt, M.; Ravitz, O. Predicting New Chemistry: Impact of High-Quality Training Data on Prediction of Reaction Outcomes. *CAS Whitepapers*. <https://www.cas.org/resources/whitepapers/predicting-new-chemistry> (accessed May 13, 2021).
- (3) Sharma, Y. Data Quality: The Not-So Secret Sauce for Ai and Machine Learning. *CAS Blog*. <https://www.cas.org/resource/blog/data-quality-not-so-secret-sauce-ai-and-machine-learning> (accessed May 13, 2021).
- (4) Griffen, E. J.; Dossetter, A. G.; Leach, A. G. Chemists: Ai Is Here; Unite to Get the Benefits. *J. Med. Chem.* **2020**, *63*, 8695–8704.
- (5) Mater, A. C.; Coote, M. L. Deep Learning in Chemistry. *J. Chem. Inf. Model.* **2019**, *59*, 2545–2559.
- (6) Wills, T. J.; Polshakov, D. A.; Robinson, M. C.; Lee, A. A. Impact of Chemist-in-the-Loop Molecular Representations on Machine Learning Outcomes. *J. Chem. Inf. Model.* **2020**, *60*, 4449–4456.
- (7) Tkatchenko, A. Machine Learning for Chemical Discovery. *Nat. Commun.* **2020**, *11*, 4125.
- (8) Lo, Y.-C.; Rensi, S. E.; Torng, W.; Altman, R. B. Machine Learning in Chemoinformatics and Drug Discovery. *Drug Discovery Today* **2018**, *23*, 1538–1546.
- (9) *Machine Learning in Chemistry: The Impact of Artificial Intelligence*; The Royal Society of Chemistry, 2020.
- (10) Mullin, R. The Lab of the Future Is Now. *Chem. Eng. News* **2021**, *28*.
- (11) Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep Learning for Molecular Design—a Review of the State of the Art. *Molecular Systems Design & Engineering* **2019**, *4*, 828–849.
- (12) Bender, A.; Cortés-Ciriano, I. Artificial Intelligence in Drug Discovery: What Is Realistic, What Are Illusions? Part 1: Ways to Make an Impact, and Why We Are Not There Yet. *Drug Discovery Today* **2021**, *26*, 511–524.
- (13) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. Qsar without Borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564.
- (14) Strieth-Kalthoff, F.; Sandfort, F.; Segler, M. H. S.; Glorius, F. Machine Learning the Ropes: Principles, Applications and Directions in Synthetic Chemistry. *Chem. Soc. Rev.* **2020**, *49*, 6154–6168.
- (15) Struble, T. J.; Alvarez, J. C.; Brown, S. P.; Chytil, M.; Cisar, J.; DesJarlais, R. L.; Engkvist, O.; Frank, S. A.; Greve, D. R.; Griffin, D. J.; Hou, X.; Johannes, J. W.; Kreatsoulas, C.; Lahue, B.; Mathea, M.; Mogk, G.; Nicolaou, C. A.; Palmer, A. D.; Price, D. J.; Robinson, R. I.; Salentin, S.; Xing, L.; Jaakkola, T.; Green, W. H.; Barzilay, R.; Coley, C. W.; Jensen, K. F. Current and Future Roles of Artificial Intelligence in Medicinal Chemistry Synthesis. *J. Med. Chem.* **2020**, *63*, 8667–8682.
- (16) CAS Content. <https://www.cas.org/about/cas-content> (accessed June 15, 2021).
- (17) Dahl, G. E.; Jaitly, N.; Salakhutdinov, R. Multi-Task Neural Networks for QSAR Predictions. *arXiv*, 2014, 1406.1231. <https://arxiv.org/abs/1406.1231>.
- (18) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. Imagenet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90.
- (19) Cauchi, M.; Weber, C. M.; Bolt, B. J.; Spratt, P. B.; Bessant, C.; Turner, D. C.; Willis, C. M.; Britton, L. E.; Turner, C.; Morgan, G. Evaluation of Gas Chromatography Mass Spectrometry and Pattern Recognition for the Identification of Bladder Cancer from Urine Headspace. *Anal. Methods* **2016**, *8*, 4037–4046.
- (20) Kim, J. Y.; Oh, D.; Sung, K.; Choi, H.; Paeng, J. C.; Cheon, G. J.; Kang, K. W.; Lee, D. Y.; Lee, D. S. Visual Interpretation of [(18)F]Florbetaben PET Supported by Deep Learning-Based Estimation of Amyloid Burden. *Eur. J. Nucl. Med. Mol. Imaging* **2021**, *48*, 1116–1123.
- (21) Jaiswal, A.; Gianchandani, N.; Singh, D.; Kumar, V.; Kaur, M. Classification of the Covid-19 Infected Patients Using DenseNet201 Based Deep Transfer Learning. *J. Biomol. Struct. Dyn.* **2020**, *1*–8.
- (22) Song, C. L.; Vardaki, M. Z.; Goldin, R. D.; Kazarian, S. G. Fourier Transform Infrared Spectroscopic Imaging of Colon Tissues: Evaluating the Significance of Amide I and C-H Stretching Bands in Diagnostic Applications with Machine Learning. *Anal. Bioanal. Chem.* **2019**, *411*, 6969–6981.
- (23) Ren, X.; Ghassemi, P.; Kanaan, Y. M.; Naab, T.; Copeland, R. L.; Dewitt, R. L.; Kim, I.; Strobl, J. S.; Agah, M. Kernel-Based Microfluidic Constriction Assay for Tumor Sample Identification. *ACS Sens* **2018**, *3*, 1510–1521.
- (24) Shen, H.; Zhang, W.; Chen, P.; Zhang, J.; Fang, A.; Wang, B. In A Feature Selection Scheme for Accurate Identification of Alzheimer's Disease. *Bioinformatics and Biomedical Engineering*; Ortúñoz, F., Rojas, I., Eds.; Springer International Publishing, 2016; pp 71–81.
- (25) Ramos, A. G.; Antón, A. P.; Sánchez, M. D. N.; Pavón, J. L. P.; Cordero, B. M. Urinary Volatile Fingerprint Based on Mass Spectrometry for the Discrimination of Patients with Lung Cancer and Controls. *Talanta* **2017**, *174*, 158–164.
- (26) Zhai, M. Y.; Zhao, Y.; Gao, H.; Shang, L. W.; Yin, J. H. Quantitative Study on Articular Cartilage by Fourier Transform Infrared Spectroscopic Imaging and Support Vector Machine. *Chinese Journal of Analytical Chemistry* **2018**, *46*, 896–901.
- (27) Zheng, Q.; Li, J.; Yang, L.; Zheng, B.; Wang, J.; Lv, N.; Luo, J.; Martin, F. L.; Liu, D.; He, J. Raman Spectroscopy as a Potential Diagnostic Tool to Analyse Biochemical Alterations in Lung Cancer. *Analyst* **2020**, *145*, 385–392.
- (28) Li, X.; Yang, T.; Li, S.; Yao, J.; Song, Y.; Wang, D.; Ding, J. Study on Spectral Parameters and the Support Vector Machine in Surface Enhanced Raman Spectroscopy of Serum for the Detection of Colon Cancer. *Laser Phys. Lett.* **2015**, *12*, 115603.
- (29) Zhang, L.; Ma, F.; Qi, A.; Liu, L.; Zhang, J.; Xu, S.; Zhong, Q.; Chen, Y.; Zhang, C.-y.; Cai, C. Integration of Ultra-High-Pressure Liquid Chromatography-Tandem Mass Spectrometry with Machine Learning for Identifying Fatty Acid Metabolite Biomarkers of Ischemic Stroke. *Chem. Commun.* **2020**, *56*, 6656–6659.
- (30) Alakwa, F. M.; Chaudhary, K.; Garmire, L. X. Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. *J. Proteome Res.* **2018**, *17*, 337–347.
- (31) Grissa, D.; Pétéra, M.; Brandolini, M.; Napoli, A.; Comte, B.; Pujo-Guillot, E. Feature Selection Methods for Early Predictive Biomarker Discovery Using Untargeted Metabolomic Data. *Front Mol. Biosci* **2016**, *3*, 30.
- (32) O'Shea, K.; Cameron, S. J.; Lewis, K. E.; Lu, C.; Mur, L. A. Metabolomic-Based Biomarker Discovery for Non-Invasive Lung Cancer Screening: A Case Study. *Biochim. Biophys. Acta, Gen. Subj.* **2016**, *1860*, 2682–7.
- (33) Paul, A.; Srivastava, S.; Roy, R.; Anand, A.; Gaurav, K.; Husain, N.; Jain, S.; Sonkar, A. A. Malignancy Prediction among Tissues from

- Oral SCC Patients Including Neck Invasions: A ¹H NMR Based Metabolomic Study. *Metabolomics* **2020**, *16*, 38.
- (34) Pan, H.; Yao, C.; Yao, S.; Yang, W.; Wu, W.; Guo, D. A Metabolomics Strategy for Authentication of Plant Medicines with Multiple Botanical Origins, a Case Study of Uncariae Rammulus Cum Uncis. *J. Sep. Sci.* **2020**, *43*, 1043–1050.
- (35) Gaul, D. A.; Mezencev, R.; Long, T. Q.; Jones, C. M.; Benigno, B. B.; Gray, A.; Fernández, F. M.; McDonald, J. F. Highly-Accurate Metabolomic Detection of Early-Stage Ovarian Cancer. *Sci. Rep.* **2015**, *5*, 16351.
- (36) Hall, L. M.; Hill, D. W.; Bugden, K.; Cawley, S.; Hall, L. H.; Chen, M.-H.; Grant, D. F. Development of a Reverse Phase HPLC Retention Index Model for Nontargeted Metabolomics Using Synthetic Compounds. *J. Chem. Inf. Model.* **2018**, *58*, 591–604.
- (37) Moawad, A. A.; Silge, A.; Bocklitz, T.; Fischer, K.; Rösch, P.; Roesler, U.; Elschner, M. C.; Popp, J.; Neubauer, H. A Machine Learning-Based Raman Spectroscopic Assay for the Identification of Burkholderia Mallei and Related Species. *Molecules* **2019**, *24*, 4516.
- (38) Kusić, D.; Rösch, P.; Popp, J. Fast Label-Free Detection of Legionella spp. In Biofilms by Applying Immunomagnetic Beads and Raman Spectroscopy. *Syst. Appl. Microbiol.* **2016**, *39*, 132–40.
- (39) Yu, F. L.; Zhao, N.; Wu, Z. S.; Huang, M.; Wang, D.; Zhang, Y. B.; Hu, X.; Chen, X. L.; Huang, L. Q.; Pang, Y. X. Nir Rapid Assessments of Blumea Balsamifera (Ai-Na-Xiang) in China. *Molecules* **2017**, *22*, 1730.
- (40) Lund, J. A.; Brown, P. N.; Shipley, P. R. Differentiation of Crataegus spp. Guided by Nuclear Magnetic Resonance Spectrometry with Chemometric Analyses. *Phytochemistry* **2017**, *141*, 11–19.
- (41) Lu, J.; Chen, C.; Wang, H.; Wen, Y. Establishing a Machine Learning Model for Cancer Anticipation and a Method of Detecting Cancer by Using Multiple Tumor Markers in the Machine Learning Model for Cancer Anticipation. US2018173847A1.
- (42) Bazemore, K. Machine Learning Algorithms and Applied Diagnostic Methods for Predicting Diseases. US2019353639A1.
- (43) Shi, T.; Ding, W.; Chen, G. Methylation Markers, and Application Thereof in Cancer Diagnosis and Classification. CN109680060A.
- (44) Barnes, M.; Bifulco, C.; Chen, T.; Tubbs, A. Imaging Processing System Using Convolutional Neural Network for Processing Imaging of Biological Staining of Animal Tissue and Cells. WO 2015181371A1; AU2015265811-A1; CA2944831-A1; EP3149700-A1; US2017154420-A1; JP2017529513-W; EP3149700-B1; US10275880-B2; CA2944831-C; AU2015265811-B2; JP6763781-B2.
- (45) Horimoto, K.; Fukui, K. Toxicity Learning Apparatus, Toxicity Learning Method, Learned Model, Toxicity Prediction Apparatus and Program. JP2020025471A.
- (46) Sogawa, I.; Kimura, A.; Koh, Y.; Kimura, A.; Ko, Y.; Sogawa, I.; Koh, Y. Method and Device for Detecting Tumor Cells by Analyzing Spectroscopic Data by Statistical Method. WO2017195772A1; JP2017203637-A; IN201847044553-A; CN109073547-A; US2019072484-A1; EP3457116-A1; EP3457116-A4.
- (47) Park, J. Y.; Oh, J. Y.; Kim, J. J.; Lee, B. S.; Yang, H. S.; Sik, Y. H. Clinical Diagnostic Data Processing System for Predicting Mortality Risk Levels. WO2020130238A1; KR2020075477-A.
- (48) Park, H. S.; Hyoung, R.; Chung, S. Method and Apparatus for Identifying Strains Based on Mass Spectrometry and Mass Spectra Peak Database. KR 2020050434 A.
- (49) Shimizu, Y.; Takashima, N. Event Estimation Method, Event Estimation Program and Server Device, Biological Information Measurement Device, Event Estimation System. JP 2020031701 A.
- (50) Sik, H. W.; Park, J. S. Computerized Diagnostic Information System Method Using Biomarker Genes and Proteins for Risk Assessment of Urogenital System Cancers and Drug Screening. KR 2020074555 A; KR2164052-B1.
- (51) Segal, E.; Bar, N.; Korem, T. Predicting Blood Metabolites. WO 2020157762 A1.
- (52) Apte, Z.; Richman, J.; Almonacid, D.; Pedroso, I.; Dumas, V.; Marquez, V.; Araya, I.; Castro, R.; Saavedra, M.; Alegria, M. Method and System for Characterization of Metabolism-Associated Conditions, Including Diagnostics and Therapies, Based on Bioinformatics Approach. WO 2019178610 A1; SG11202002522-A1; AU2019233926-A1; CN111373481-A; KR2020132954-A; EP3766073-A1; US2021074384-A1.
- (53) Dutta, A.; Kashefaghghi, D.; Kia, A.; Jaganathan, K.; Gobbel, J. R. Artificial Intelligence-Based Sequencing. WO 2020191391 A2; WO2020191391-A3; AU2020240141-A1; CA3104951-A1.
- (54) Lim, J. S.; Vandermause, J.; van Spronsen, M. A.; Musaelian, A.; Xie, Y.; Sun, L.; O'Connor, C. R.; Egle, T.; Molinari, N.; Florian, J.; Duanmu, K.; Madix, R. J.; Sautet, P.; Friend, C. M.; Kozinsky, B. Evolution of Metastable Structures at Bimetallic Surfaces from Microscopy and Machine-Learning Molecular Dynamics. *J. Am. Chem. Soc.* **2020**, *142*, 15907–15916.
- (55) Takahashi, K.; Takahashi, L.; Miyazato, I.; Tanaka, Y. Searching for Hidden Perovskite Materials for Photovoltaic Systems by Combining Data Science and First Principle Calculations. *ACS Photonics* **2018**, *5*, 771–775.
- (56) Liu, C.; Li, Y.; Takao, M.; Toyao, T.; Maeno, Z.; Kamachi, T.; Hinuma, Y.; Takigawa, I.; Shimizu, K.-i. Frontier Molecular Orbital Based Analysis of Solid-Adsorbate Interactions over Group 13 Metal Oxide Surfaces. *J. Phys. Chem. C* **2020**, *124*, 15355–15365.
- (57) Li, X.; Xie, Y.; Hu, D.; Lan, Z. Analysis of the Geometrical Evolution in on-the-Fly Surface-Hopping Nonadiabatic Dynamics with Machine Learning Dimensionality Reduction Approaches: Classical Multidimensional Scaling and Isometric Feature Mapping. *J. Chem. Theory Comput.* **2017**, *13*, 4611–4623.
- (58) Deng, C.; Su, Y.; Li, F.; Shen, W.; Chen, Z.; Tang, Q. Understanding Activity Origin for the Oxygen Reduction Reaction on Bi-Atom Catalysts by DFT Studies and Machine-Learning. *J. Mater. Chem. A* **2020**, *8*, 24563–24571.
- (59) Quaranta, V.; Behler, J.; Hellström, M. Structure and Dynamics of the Liquid-Water/Zinc-Oxide Interface from Machine Learning Potential Simulations. *J. Phys. Chem. C* **2019**, *123*, 1293–1304.
- (60) Lansford, J. L.; Vlachos, D. G. Spectroscopic Probe Molecule Selection Using Quantum Theory, First-Principles Calculations, and Machine Learning. *ACS Nano* **2020**, *14*, 17295–17307.
- (61) Datar, A.; Chung, Y. G.; Lin, L.-C. Beyond the BET Analysis: The Surface Area Prediction of Nanoporous Materials Using a Machine Learning Method. *J. Phys. Chem. Lett.* **2020**, *11*, 5412–5417.
- (62) Qi, X.; Ma, W.; Dang, Y.; Su, W.; Liu, L. Optimization of the Melt/Crystal Interface Shape and Oxygen Concentration During the Czochralski Silicon Crystal Growth Process Using an Artificial Neural Network and a Genetic Algorithm. *J. Cryst. Growth* **2020**, *548*, 125828.
- (63) Karim, M. R.; Ferrandon, M.; Medina, S.; Sture, E.; Kariuki, N.; Myers, D. J.; Holby, E. F.; Zelenay, P.; Ahmed, T. Coupling High-Throughput Experiments and Regression Algorithms to Optimize Pgm-Free Orr Electrocatalyst Synthesis. *ACS Applied Energy Materials* **2020**, *3*, 9083–9088.
- (64) Khmaissia, F.; Frigui, H.; Andriotis, A. N.; Menon, M. Data Driven Modeling of Magnetism in Dilute Magnetic Semiconductors: Correlation between the Magnetic Features of Diluted Magnetic Semiconductors and Electronic Properties of the Constituent Atoms. *J. Phys.: Condens. Matter* **2019**, *31*, 445901.
- (65) Rong, K.; Wei, J.; Huang, L.; Fang, Y.; Dong, S. Synthesis of Low Dimensional Hierarchical Transition Metal Oxides Via a Direct Deep Eutectic Solvent Calcining Method for Enhanced Oxygen Evolution Catalysis. *Nanoscale* **2020**, *12*, 20719–20725.
- (66) Wu, T.; Wang, J. Deep Mining Stable and Nontoxic Hybrid Organic-Inorganic Perovskites for Photovoltaics Via Progressive Machine Learning. *ACS Appl. Mater. Interfaces* **2020**, *12*, 57821–57831.
- (67) Suresh, T.; Sivarajasekar, N.; Balasubramani, K. Enhanced Ultrasonic Assisted Biodiesel Production from Meat Industry Waste (Pig Tallow) Using Green Copper Oxide Nanocatalyst: Comparison of Response Surface and Neural Network Modelling. *Renewable Energy* **2021**, *164*, 897–907.

- (68) Wu, L.; Guo, T.; Li, T. Rational Design of Transition Metal Single-Atom Electrocatalysts: A Simulation-Based, Machine Learning Accelerated Study. *J. Mater. Chem. A* **2020**, *8*, 19290–19299.
- (69) Ulissi, Z. W.; Singh, A. R.; Tsai, C.; Nørskov, J. K. Automated Discovery and Construction of Surface Phase Diagrams Using Machine Learning. *J. Phys. Chem. Lett.* **2016**, *7*, 3931–3935.
- (70) Graziosi, P.; Kumarasinghe, C.; Neophytou, N. Material Descriptors for the Discovery of Efficient Thermoelectrics. *ACS Applied Energy Materials* **2020**, *3*, 5913–5926.
- (71) Wang, Z.; Zhang, H.; Li, J. Accelerated Discovery of Stable Spinels in Energy Systems Via Machine Learning. *Nano Energy* **2021**, *81*, 105665.
- (72) Masood, H.; Toe, C. Y.; Teoh, W. Y.; Sethu, V.; Amal, R. Machine Learning for Accelerated Discovery of Solar Photocatalysts. *ACS Catal.* **2019**, *9*, 11774–11787.
- (73) Davies, D. W.; Butler, K. T.; Walsh, A. Data-Driven Discovery of Photoactive Quaternary Oxides Using First-Principles Machine Learning. *Chem. Mater.* **2019**, *31*, 7221–7230.
- (74) Li, Z.; Achenie, L. E. K.; Xin, H. An Adaptive Machine Learning Strategy for Accelerating Discovery of Perovskite Electro-catalysts. *ACS Catal.* **2020**, *10*, 4377–4384.
- (75) Gheith, M. E. M.; Stobert, I.; Hamouda, A. Curvilinear Mask Models in Semiconductor Structure Manufacture by Machine Learning. US 10831977 B1; US2020380089-A1.
- (76) Chang, B. Y.; Jang, B. Y.; Zhang, F. Substrate Treating Apparatus and Substrate Treating Method. US 2020192308 A1; KR2020072060-A; CN111312613-A.
- (77) Van Den Brink, M.; Cao, Y.; Zou, Y.; Van Den Brink, M. A. Machine Learning Based Inverse Optical Proximity Correction and Process Model Calibration. WO 2019238372 A1; TW202001447-A; KR2021010897-A; CN112384860-A.
- (78) Yati, A. Defect Discovery Using Electron Beam Inspection and Deep Learning with Real-Time Intelligence to Reduce Nuisance. US 2019213733 A1; WO2019136190-A1; TW201939634-A; KR2020096993-A; CN11542915-A; US10970834-B2.
- (79) Chang, R. Photolithography Mask Design-Rule Check Assistance Using Computer for Semiconductor Wafer Manufacture. US 10713411 B1.
- (80) Wang, D. Y.; Salcin, E.; Friedmann, M.; Shaughnessy, D.; Shchegrov, A. V.; Madsen, J. M.; Kuznetsov, A. Methods and Systems for Co-Located Metrology of Semiconductor Structures. US 2020243400 A1; WO2020154152-A1; US10804167-B2.
- (81) Honda, T.; Kekatpure, R. D.; David, J. D. Temporal Dependencies of Process Targets for Different Machine Learning Models for Robust Machine Learning Predictions for Semiconductor Manufacturing Processes. US 2018356807 A1.
- (82) David, J. D. Process Control Techniques for Semiconductor Manufacturing Processes. US 2016148850 A1; WO2016086138-A1; KR2017086585-A; CN107004060-A; JP2017536584-W; US2018358271-A1; US10734293-B2; JP6751871-B2.
- (83) Lauber, J.; Vajaria, H.; Zhang, Y.; Yong, Z. Multi-Step Image Alignment Method for Large Offset Die-Die Inspection for Defects in Semiconductor Device Manufacture. US 2019122913 A1; WO2019079658-A1; TW201928541-A; US10522376-B2; CN111164646-A; KR2020060519-A; EP3698322-A1; JP2021500740-W.
- (84) Bhosale, P.; Rizzolo, M.; Yang, C. Automated Method for Integrated Analysis of Back End of the Line Yield, Line Resistance/Capacitance and Semiconductor Device Fabrication Process Performance. US2018349535A1; US10303829-B2.
- (85) Sriraman, H. P.; Pathangi, H. S. Defect Detection on Semiconductor Wafers, Classification, and Process Window Control by Sem. US2019287238A1; WO2019177800-A1; TW201941162-A; US10679333-B2; KR2020122401-A; CN111837225-A.
- (86) Kwon, N.; Kang, H.; Kim, Y.; Quan, N. Semiconductor Defect Classification Device, Method for Classifying Defect of Semiconductor, and Semiconductor Defect Classification System. US2019188840 A1; KR2019073756-A; CN110060228-A; US10713778-B2.
- (87) Lai, W.; Song, W.; Li, X.; Yan, Y.; Huang, W. Intrinsic Stretchable Electroluminescent Block Copolymer Elastomer and Preparation Method Thereof. CN111635504A.
- (88) Blaier, O.; Schiller, E. Optimization for 3d Printing. US2019366644A1; EP3584723-A2; EP3584723-A3; US10926475-B2.
- (89) Deetz, J. D.; Wood, C. E.; Truong, R. A. Resin Viscosity Detection in Additive Manufacturing. US2020338830A1.
- (90) Chen, A.; Zheng, X. Repairable Multi-Response Deformable Liquid Crystal Elastomer Film and Its Preparation Method and Application in Artificial Intelligence. CN108727544A; CN108727544-B.
- (91) Chhabra, S.; Xie, J.; Frank, A. T. Rnappers: Machine Learning Classifiers for Ribonucleic Acid-Ligand Poses. *J. Phys. Chem. B* **2020**, *124*, 4436–4445.
- (92) Shamsara, J.; Schüürmann, G. A Machine Learning Approach to Discriminate Mr1 Binders: The Importance of the Phenol and Carbonyl Fragments. *J. Mol. Struct.* **2020**, *1217*, 128459.
- (93) Ji, B.-Y.; You, Z.-H.; Jiang, H.-J.; Guo, Z.-H.; Zheng, K. Prediction of Drug-Target Interactions from Multi-Molecular Network Work Based on Line network Representation Method. *J. Transl. Med.* **2020**, *18*, 347.
- (94) Yuan, Y.; Chang, S.; Zhang, Z.; Li, Z.; Li, S.; Xie, P.; Yau, W.-P.; Lin, H.; Cai, W.; Zhang, Y.; Xiang, X. A Novel Strategy for Prediction of Human Plasma Protein Binding Using Machine Learning Techniques. *Chemom. Intell. Lab. Syst.* **2020**, *199*, 103962.
- (95) Aniceto, N.; Freitas, A. A.; Bender, A.; Ghafourian, T. Simultaneous Prediction of Four Atp-Binding Cassette Transporters' Substrates Using Multi-Label Qsar. *Mol. Inf.* **2016**, *35*, 514–528.
- (96) Zhao, Y.; Zheng, K.; Guan, B.; Guo, M.; Song, L.; Gao, J.; Qu, H.; Wang, Y.; Shi, D.; Zhang, Y. Dldti: A Learning-Based Framework for Drug-Target Interaction Identification Using Neural Networks and Network Representation. *J. Transl. Med.* **2020**, *18*, 434.
- (97) Hochuli, J.; Helbling, A.; Skaist, T.; Ragoza, M.; Koes, D. R. Visualizing Convolutional Neural Network Protein-Ligand Scoring. *J. Mol. Graphics Modell.* **2018**, *84*, 96–108.
- (98) Erdas, O.; Andac, C. A.; Gurkan-Alp, A. S.; Alpaslan, F. N.; Buyukbingol, E. Compressed Images for Affinity Prediction-2 (Cifap-2): An Improved Machine Learning Methodology on Protein-Ligand Interactions Based on a Study on Caspase 3 Inhibitors. *J. Enzyme Inhib. Med. Chem.* **2015**, *30*, 809–15.
- (99) Fan, J.; Liu, K.; Xiangyan, S. Computational Method for Classifying and Predicting Ligand Docking Conformations. WO2018213767A1; US2018341754-A1; EP3427170-A1; EP3427170-A4.
- (100) Feinberg, E. N.; Pande, V. S. Machine Learning and Molecular Simulation Based Methods for Enhancing Binding and Activity Prediction. US2019272887A1; WO2019173407-A1; CA3093260-A1; AU2019231261-A1; KR2020128710-A; EP3762730-A1; CN112204402-A.
- (101) Mamoshina, P.; Volosnikova, M.; Ozerov, I. V.; Putin, E.; Skibina, E.; Cortese, F.; Zhavoronkov, A. Machine Learning on Human Muscle Transcriptomic Data for Biomarker Discovery and Tissue-Specific Drug Target Identification. *Front. Genet.* **2018**, DOI: 10.3389/fgene.2018.00242.
- (102) Han, M.; Liu, Q.; Yu, J.; Zheng, S. Identification of Candidate Molecular Markers Predicting Chemotherapy Resistance in Non-Small Cell Lung Cancer. *Clin. Chem. Lab. Med.* **2010**, *48*, 863–867.
- (103) Thishya, K.; Vattam, K. K.; Naushad, S. M.; Raju, S. B.; Kutala, V. K. Artificial Neural Network Model for Predicting the Bioavailability of Tacrolimus in Patients with Renal Transplantation. *PLoS One* **2018**, *13*, No. e0191921.
- (104) Liu, Q.; Muglia, L. J.; Huang, L. F. Network as a Biomarker: A Novel Network-Based Sparse Bayesian Machine for Pathway-Driven Drug Response Prediction. *Genes* **2019**, *10*, 602.
- (105) Miyoshi, F.; Honne, K.; Minota, S.; Okada, M.; Ogawa, N.; Mimura, T. A Novel Method Predicting Clinical Response Using Only Background Clinical Data in Ra Patients before Treatment with Infliximab. *Mod. Rheumatol.* **2016**, *26*, 813–816.

- (106) Khojasteh, M.; Martin, J.; Pestic-Dragovich, L.; Tang, L.; Wang, X.; Zhang, W.; Anders, R.; Diaz, L. Methods and Systems for Predicting Response to Pd-1 Axis Directed Therapeutics of Tumors. WO 2020072348 A1.
- (107) Yang, H.-Y. Prediction of Pneumoconiosis by Serum and Urinary Biomarkers in Workers Exposed to Asbestos-Contaminated Minerals. *PLoS One* 2019, 14, No. e0214808.
- (108) Schrey, A. K.; Nickel-Seeber, J.; Drwal, M. N.; Zwicker, P.; Schultze, N.; Haertel, B.; Preissner, R. Computational Prediction of Immune Cell Cytotoxicity. *Food Chem. Toxicol.* 2017, 107, 150–166.
- (109) Lee, J. J.; Miller, J. A.; Basu, S.; Kee, T. V.; Loo, L. H. Building Predictive in Vitro Pulmonary Toxicity Assays Using High-Throughput Imaging and Artificial Intelligence. *Arch. Toxicol.* 2018, 92, 2055–2075.
- (110) Hamadache, M.; Hanini, S.; Benkortbi, O.; Amrane, A.; Khaouane, L.; Moussa, C. S. Artificial Neural Network-Based Equation to Predict the Toxicity of Herbicides on Rats. *Chemom. Intell. Lab. Syst.* 2016, 154, 7–15.
- (111) Noskov, S.; Wacker, S.; Duff, H.; Guo, J. Systems and Methods for Predicting Cardiotoxicity of Molecular Parameters of a Compound Based on Machine Learning Algorithms. WO2016201575A1; US2018172667-A1.
- (112) Lee, F. K.; Friesth, K. L. Quintuple-Effect Generation Multi-Cycle Hybrid Renewable Energy System with Integrated Energy Provisioning, Storage Facilities and Amalgamated Control System Cross-Reference to Related Applications. US 2015143806 A1; AU2015203118-A1; EP2955372-A2; JP2016000995-A; CA2891435-A1; CN105257425-A; EP2955372-A3; HK1218148-A0; US10060296-B2; BR102015013592-A2.
- (113) Rangarajan, K.; Winston, J. B.; Jain, A.; Wang, X.; Jian, A.; Rangarajan, K. P. Integrated Surveillance and Control of Oilfield Activity. FR3070178A1; WO2019040125-A1; AU2018319552-A1; CA3065094-A1; NO201901443-A; US2020182036-A1; GB2579739-A.
- (114) Tang, J.; Hou, J. Intelligent Ammonia Injection Control Method and Intelligent Ammonia Injection Controller. CN111804146A.
- (115) Shen, D.; Liu, G.; Wang, Q.; Luo, K. Accurate Ammonia Injection Control Method for Scr System with Strong Self-Adaptive Ability. CN109046021A; CN109046021-B.
- (116) Liu, X.; Lin, P.; Hu, G.; Wu, K. A Kind of Scr Downstream Nox Closed-Loop Process Control Method and System. CN109339916A; WO2020062865-A1; CN109339916-B.
- (117) Guo, L.; Zou, S.; Liu, W.; Wu, Q.; Zhou, W.; Zhao, X.; Yao, T.; Xu, Q. Oil and Gas Gathering and Transportation Riser System Harmful Flow Type Warning Method and System, Control Method and System. CN109458561A; WO2020082749-A1.
- (118) Li, Y.; Dong, S.; Wu, Q.; Zhang, S. Flue Gas Desulfurization Device with Flue Gas Monitoring and Control Function. CN110756039 A; CN211753913-U.
- (119) Meng, L.; Gu, X.; Ma, W.; Ning, X.; Jiang, C.; Li, Y.; Jia, Y. Scr Denitrification Ammonia-Spraying Optimization Method and System Based on Advanced Measuring Meter and Advanced Control Algorithm. CN108837698A.
- (120) Kim, K. S. Control System of Dual-Fuel Engine. US2019093572A1; US10260432-B1.
- (121) Brummel, H.; Pfeifer, U.; Sterzing, V. Method and Assembly for Controlling a Combustion Engine with Multiple Burners. EP3726139A1; WO2020212067-A1.
- (122) Eun, L. Device and Method for Efficiently Controlling Fuel Additive Injector of Engine. KR2180985B1.
- (123) Chen, S. K.; Mandal, A.; Chien, L.; Ortiz-Soto, E. Machine Learning for Misfire Detection in a Dynamic Firing Level Modulation Controlled Engine of a Vehicle. US2019145859A1; WO2019099228-A1; US10816438-B2.
- (124) Denys, F.; Leroy, T.; Ngo, C.; Rudloff, J. Device and Method for Control of a Vehicle Thermal Engine. FR3085442A1.
- (125) Streib, M.; Luo, L.; Klinkhammer, T.; Leuz, M.; Kluth, C.; Polach, S.; Pollach, S. Method for Controlling Engine Knock of an Internal Combustion Engine. DE102015208359A1; WO2016177531-A1; DE102015208359-B4; CN107624144-A; US2018112631-A1; US10451009-B2; CN107624144-B.
- (126) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C. L.; Zidek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* 2020, 577, 706–710.
- (127) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W. G.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackermann, Z.; Tran, V. M.; Chiappino-Pepe, A.; Badran, A. H.; Andrews, I. W.; Chory, E. J.; Church, G. M.; Brown, E. D.; Jaakkola, T. S.; Barzilay, R.; Collins, J. J. A Deep Learning Approach to Antibiotic Discovery. *Cell* 2020, 180, 688–702.
- (128) Almagro Armenteros, J. J.; Tsirigos, K. D.; Sonderby, C. K.; Petersen, T. N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat. Biotechnol.* 2019, 37, 420–423.
- (129) Newman, A. M.; Steen, C. B.; Liu, C. L.; Gentles, A. J.; Chaudhuri, A. A.; Scherer, F.; Khodadoust, M. S.; Esfahani, M. S.; Luca, B. A.; Steiner, D.; Diehn, M.; Alizadeh, A. A. Determining Cell Type Abundance and Expression from Bulk Tissues with Digital Cytometry. *Nat. Biotechnol.* 2019, 37, 773–782.
- (130) Jaganathan, K.; Kyriazopoulou Panagiotopoulou, S.; McRae, J. F.; Darbandi, S. F.; Knowles, D.; Li, Y. I.; Kosmicki, J. A.; Arbelaez, J.; Cui, W. W.; Schwartz, G. B.; Chow, E. D.; Kanterakis, E.; Gao, H.; Kia, A.; Batzoglou, S.; Sanders, S. J.; Farh, K. K. H. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 2019, 176, 535–548.
- (131) van Galen, P.; Hovestadt, V.; Wadsworth, M. H.; Hughes, T. K.; Griffin, G. K.; Battaglia, S.; Verga, J. A.; Stephansky, J.; Pastika, T. J.; Lombardi Story, J.; Pinkus, G. S.; Pozdnyakova, O.; Galinsky, I.; Stone, R. M.; Graubert, T. A.; Shalek, A. K.; Aster, J. C.; Lane, A. A.; Bernstein, B. E. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* 2019, 176, 1265–1281.
- (132) Capper, D.; Jones, D. T. W.; Sill, M.; Hovestadt, V.; Schrimpf, D.; Sturm, D.; Koelsche, C.; Sahm, F.; Chavez, L.; Reuss, D. E.; Kratz, A.; Wefers, A. K.; Huang, K.; Pajtler, K. W.; Schweizer, L.; Stichel, D.; Olar, A.; Engel, N. W.; Lindenberg, K.; Harter, P. N.; Braczynski, A. K.; Plate, K. H.; Dohmen, H.; Garvalov, B. K.; Coras, R.; Holsken, A.; Hewer, E.; Bewerunge-Hudler, M.; Schick, M.; Fischer, R.; Beschorner, R.; Schittenhelm, J.; Staszewski, O.; Wani, K.; Varlet, P.; Pages, M.; Temming, P.; Lohmann, D.; Selt, F.; Witt, H.; Milde, T.; Witt, O.; Aronica, E.; Giangaspero, F.; Rushing, E.; Scheurlen, W.; Geisenberger, C.; Rodriguez, F. J.; Becker, A.; Preusser, M.; Haberler, C.; Bjerkvig, R.; Cryan, J.; Farrell, M.; Deckert, M.; Hench, J.; Frank, S.; Serrano, J.; Kannan, K.; Tsirigos, A.; Bruck, W.; Hofer, S.; Brehmer, S.; Seiz-Rosenthal, M.; Hanggi, D.; Hans, V.; Rozsnoki, S.; Hansford, J. R.; Kohlhof, P.; Kristensen, B. W.; Lechner, M.; Lopes, B.; Mawrin, C.; Ketter, R.; Kulozik, A.; Khatib, Z.; Heppner, F.; Koch, A.; Jouvet, A.; Keohane, C.; Muhleisen, H.; Mueller, W.; Pohl, U.; Prinz, M.; Benner, A.; Zapata, M.; Gottardo, N. G.; Drieber, P. H.; Kramm, C. M.; Muller, H. L.; Rutkowski, S.; von Hoff, K.; Fruhwald, M. C.; Gnekow, A.; Fleischhacker, G.; Tippelt, S.; Calaminus, G.; Monoranu, C.-M.; Perry, A.; Jones, C.; Jacques, T. S.; Radlwimmer, B.; Gessi, M.; Pietsch, T.; Schramm, J.; Schackert, G.; Westphal, M.; Reifenberger, G.; Wesseling, P.; Weller, M.; Collins, V. P.; Blumcke, I.; Bendszus, M.; Debus, J.; Huang, A.; Jabado, N.; Northcott, P. A.; Paulus, W.; Gajjar, A.; Robinson, G. W.; Taylor, M. D.; Jaunmuktane, Z.; Ryzhova, M.; Platten, M.; Unterberg, A.; Wick, W.; Karajannis, M. A.; Mittelbronn, M.; Acker, T.; Hartmann, C.; Aldape, K.; Schuller, U.; Buslei, R.; Lichter, P.; Kool, M.; Herold-Mende, C.; Ellison, D. W.; Hasselblatt, M.; Snuderl, M.; Brandner, S.; Korshunov, A.; von Deimling, A.; Pfister, S. M. DNA Methylation-Based Classification of Central Nervous System Tumours. *Nature* 2018, 555, 469–474.

- (133) Segler, M. H. S.; Kogej, T.; Tyrchan, C.; Waller, M. P. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Cent. Sci.* **2018**, *4*, 120–131.
- (134) Jespersen, M. C.; Peters, B.; Nielsen, M.; Marcatili, P. Bepipred-2.0: Improving Sequence-Based B-Cell Epitope Prediction Using Conformational Epitopes. *Nucleic Acids Res.* **2017**, *45*, W24–W29.
- (135) Iorio, F.; Knijnenburg, T. A.; Vis, D. J.; Bignell, G. R.; Menden, M. P.; Schubert, M.; Aben, N.; Goncalves, E.; Barthorpe, S.; Lightfoot, H.; Cokelaer, T.; Greninger, P.; van Dyk, E.; Chang, H.; de Silva, H.; Heyn, H.; Deng, X. M.; Egan, R. K.; Liu, Q. S.; Mironenko, T.; Mitropoulos, X.; Richardson, L.; Wang, J. H.; Zhang, T. H.; Moran, S.; Sayols, S.; Soleimani, M.; Tamborero, D.; Lopez-Bigas, N.; Ross-Macdonald, P.; Esteller, M.; Gray, N. S.; Haber, D. A.; Stratton, M. R.; Benes, C. H.; Wessels, L. F. A.; Saez-Rodriguez, J.; McDermott, U.; Garnett, M. J. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **2016**, *166*, 740–754.
- (136) Zeevi, D.; Korem, T.; Zmora, N.; Israeli, D.; Rothschild, D.; Weinberger, A.; Ben-Yacob, O.; Lador, D.; Avnit-Sagi, T.; Lotan-Pompan, M.; Suez, J.; Mahdi, J. A.; Matot, E.; Malka, G.; Kosower, N.; Rein, M.; Zilberman-Schapira, G.; Dohnalova, L.; Pevsner-Fischer, M.; Bikovsky, R.; Halpern, Z.; Elinav, E.; Segal, E. Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **2015**, *163*, 1079–1094.
- (137) Kircher, M.; Witten, D. M.; Jain, P.; O’Roak, B. J.; Cooper, G. M.; Shendure, J. A General Framework for Estimating the Relative Pathogenicity of Human Genetic Variants. *Nat. Genet.* **2014**, *46*, 310–315.
- (138) Subramanian, S.; Huq, S.; Yatsunenko, T.; Haque, R.; Mahfuz, M.; Alam, M. A.; Benezra, A.; DeStefano, J.; Meier, M. F.; Muegge, B. D.; Barratt, M. J.; VanArendonk, L. G.; Zhang, Q. Y.; Province, M. A.; Petri, W. A.; Ahmed, T.; Gordon, J. I. Persistent Gut Microbiota Immaturity in Malnourished Bangladeshi Children. *Nature* **2014**, *510*, 417–421.
- (139) Zhong, M.; Tran, K.; Min, Y. M.; Wang, C. H.; Wang, Z. Y.; Dinh, C. T.; De Luna, P.; Yu, Z. Q.; Rasouli, A. S.; Brodersen, P.; Sun, S.; Voznyy, O.; Tan, C. S.; Askerka, M.; Che, F. L.; Liu, M.; Seifitokaldani, A.; Pang, Y. J.; Lo, S. C.; Ip, A.; Ulissi, Z.; Sargent, E. H. Accelerated Discovery of Co₂ Electrocatalysts Using Active Machine Learning. *Nature* **2020**, *581*, 178–183.
- (140) Manipatruni, S.; Nikonorov, D. E.; Lin, C. C.; Gosavi, T. A.; Liu, H. C.; Prasad, B.; Huang, Y. L.; Bonturim, E.; Ramesh, R.; Young, I. A. Scalable Energy-Efficient Magnetoelectric Spin-Orbit Logic. *Nature* **2019**, *565*, 35–42.
- (141) Zhu, X. J.; Li, D.; Liang, X. G.; Lu, W. D. Ionic Modulation and Ionic Coupling Effects in MoS₂ Devices for Neuromorphic Computing. *Nat. Mater.* **2019**, *18*, 141–148.
- (142) Bai, Y.; Wilbraham, L.; Slater, B. J.; Zwijnenburg, M. A.; Sprick, R. S.; Cooper, A. I. Accelerated Discovery of Organic Polymer Photocatalysts for Hydrogen Evolution from Water through the Integration of Experiment and Theory. *J. Am. Chem. Soc.* **2019**, *141*, 9063–9071.
- (143) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, DOI: 10.1103/PhysRevLett.120.145301.
- (144) van de Burgt, Y.; Lubberman, E.; Fuller, E. J.; Keene, S. T.; Faria, G. C.; Agarwal, S.; Marinella, M. J.; Talin, A. A.; Salleo, A. A Non-Volatile Organic Electrochemical Device as a Low-Voltage Artificial Synapse for Neuromorphic Computing. *Nat. Mater.* **2017**, *16*, 414–418.
- (145) Gomez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D. G.; Wu, T.; Markopoulos, G.; Jeon, S.; Kang, H.; Miyazaki, H.; Numata, M.; Kim, S.; Huang, W. L.; Hong, S. I.; Baldo, M.; Adams, R. P.; Aspuru-Guzik, A. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15*, 1120–1127.
- (146) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533*, 73–76.
- (147) Podgorski, J.; Berg, M. Global Threat of Arsenic in Groundwater. *Science* **2020**, *368*, 845–850.
- (148) Wang, H. D.; Rivenson, Y.; Jin, Y. Y.; Wei, Z. S.; Gao, R.; Gunaydin, H.; Bentolila, L. A.; Kural, C.; Ozcan, A. Deep Learning Enables Cross-Modality Super-Resolution in Fluorescence Microscopy. *Nat. Methods* **2019**, *16*, 103–110.
- (149) Arganda-Carreras, I.; Kaynig, V.; Rueden, C.; Eliceiri, K. W.; Schindelin, J.; Cardona, A.; Seung, H. S. Trainable Weka Segmentation: A Machine Learning Tool for Microscopy Pixel Classification. *Bioinformatics* **2017**, *33*, 2424–2426.
- (150) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform for Comprehensive Analysis of (Proteo)Omics Data. *Nat. Methods* **2016**, *13*, 731–740.
- (151) Coley, C. W.; Thomas, D. A.; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H. Y.; Hicklin, R. W.; Plehiers, P. P.; Byington, J.; Piotti, J. S.; Green, W. H.; Hart, A. J.; Jamison, T. F.; Jensen, K. F. A Robotic Platform for Flow Synthesis of Organic Compounds Informed by Ai Planning. *Science* **2019**, *365*, eaax1566.
- (152) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic Ai. *Nature* **2018**, *555*, 604–610.
- (153) Ahneman, D. T.; Estrada, J. G.; Lin, S. S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C-N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186–190.
- (154) Unke, O. T.; Meuwly, M. Physnet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693.
- (155) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schutt, K. T.; Muller, K. R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Science Advances* **2017**, *3*, e1603015.
- (156) Carleo, G.; Troyer, M. Solving the Quantum Many-Body Problem with Artificial Neural Networks. *Science* **2017**, *355*, 602–605.
- (157) Schutt, K. T.; Arbabzadah, F.; Chmiela, S.; Muller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, DOI: 10.1038/ncomms13890.
- (158) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower Than Hybrid Dft Error. *J. Chem. Theory Comput.* **2017**, *13*, 5255–5264.
- (159) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Delta-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.
- (160) European Patent Office. *Patent Families*. <https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/patent-families.html> (accessed June 15, 2021).
- (161) Gartner Glossary. <https://www.gartner.com/en/information-technology/glossary/hype-cycle> (accessed June 15, 2021).