



Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering

Dennis M. Dimiduk^{1,2} · Elizabeth A. Holm³ · Stephen R. Niezgoda^{2,4}

Received: 28 June 2018 / Accepted: 7 August 2018 / Published online: 27 August 2018
© The Minerals, Metals & Materials Society 2018

Abstract

The fields of machine learning and artificial intelligence are rapidly expanding, impacting nearly every technological aspect of society. Many thousands of published manuscripts report advances over the last 5 years or less. Yet materials and structures engineering practitioners are slow to engage with these advancements. Perhaps the recent advances that are driving other technical fields are not sufficiently distinguished from long-known informatics methods for materials, thereby masking their likely impact to the materials, processes, and structures engineering (MPSE). Alternatively, the diverse nature and limited availability of relevant materials data pose obstacles to machine-learning implementation. The glimpse captured in this overview is intended to draw focus to selected distinguishing advances, and to show that there are opportunities for these new technologies to have transformational impacts on MPSE. Further, there are opportunities for the MPSE fields to contribute understanding to the emerging machine-learning tools from a physics basis. We suggest that there is an immediate need to expand the use of these new tools throughout MPSE, and to begin the transformation of engineering education that is necessary for ongoing adoption of the methods.

Keywords ICME · MGI · Machine learning · Deep learning · Artificial intelligence · Multiscale modeling · Digital engineering

✉ Dennis M. Dimiduk
dennis.dimiduk@bluequartz.net

Elizabeth A. Holm
eaholm@andrew.cmu.edu

Stephen R. Niezgoda
niezgoda.6@osu.edu

¹ BlueQuartz Software, LLC, 400 S. Pioneer Blvd, Springboro, OH 45066, USA

² Department of Materials Science and Engineering, The Ohio State University, 2041 College Road, Columbus, OH 43210, USA

³ Department of Materials Science and Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

⁴ Department of Mechanical and Aerospace Engineering, The Ohio State University, 2041 College Road, Columbus, OH 43210, USA

Introduction and Motivation

Since 2012, society has seen drastic improvements in the fields of automated/autonomous data analysis, informatics, and deep learning (defined later). The advancements stem from gains in widespread digital data, computing power, and algorithms applied to machine-learning (ML) and artificial intelligence (AI) systems. Here, we distinguish the term ML as obtaining a computed model of complex non-linear relationships or complex patterns within data (usually beyond human capability or established physics to define), and AI as the framework for making machine-based decisions and actions using ML tools and analyses. Both of these are necessary but not sufficient steps for attaining autonomous systems. Autonomy requires at least three concurrently operating technologies: (i) perception or sensing a field of information and making analyses (i.e., ML); (ii) predicting or forecasting how the sensed field will evolve or change over time; and (iii) establishing a policy or decision basis for a machine (robot) to take *unsupervised* action based on (i) and (ii). We note that item (ii) in the

aforementioned list is not often discussed with respect to ML since the technical essence of item (ii) resides within the realm of control theory/control system engineering. Nonetheless, these control systems are increasingly using both models and ML/AI for learning the trajectories of the sensed field evolution and generating the navigation policy, going beyond ML for interpretation of the sensed field [1–3] In this context, we also note that making predictions or forecasts about engineered systems is a core strength of the materials, processes, and structures engineering (MPSE) fields of practice. As we discuss later, that core strength will be essential to leverage for both bringing some aspects of ML/AI tools into MPSE and for aiding with understanding the tools themselves. Thus, a natural basis exists for a marriage between ML or data science and MPSE for attaining autonomous materials discovery systems.

From another perspective, in engineering and materials, “Big Data” often refers to data itself and repositories for it. However, more vexing issues are tied to myriad sources of data and the often sparse nature of materials data. Within current MPSE practices, the scale and velocity of acquiring data, the veracity of data, and even the volatility of the data are additional challenges for practitioners. These raise the question of how to analyze and use MPSE data in a practical manner that supports decisions for developers and designers. That challenge looms large since the data sources and their attributes have defied development within a structured overall ontology, thus leaving MPSE data “semi-structured” at best. Here too ML/AI technologies are likely transformational for advancing new solutions to the long-standing data structure challenge. By embracing ML/AI tools for dealing with data, one naturally evolves data structures associated with the use of ML tools, related both to the input form and the output. Further, when the tools are employed, one gains insights in the sufficiency of data for attaining a given level of analysis. Finally, since the tools for ML and AI are primarily being developed to treat unstructured data, there may be gains in understanding the broad MPSE data ontology by employing them within MPSE.

Materials data have wide-ranging scope and often relatively little depth. In this context, depth can be interpreted as the number of independent observations of the state of a system. The lack of data depth stems from not only the historically high costs and difficulty of acquiring materials data, especially experimentally, but also from the nature of the data itself (i.e., small numbers (<100) of mechanical tests, micrographs or images, chemical spectra, etc.). Yet utilizing data to its fullest is a key aspect of advanced engineering design systems. Consequently, the emerging ML/AI technologies that support mining and extracting knowledge from data may form an important aspect of future data, informatics, and visualization aspects

of engineering design systems, provided that the ML/AI tools can be evolved for use within more limited data sets. That evolution must include modeling the means/systems for acquiring data itself. That is, because the data are so expensive and typically difficult to acquire, the data must exist within model frameworks such that models permit synthesizing data that is related to that which is actually acquired, or fills gaps in the data to facilitate further analysis and modeling. Having such structures would permit ML/AI tools to form rigorous relationships between these types of data, measured and synthesized. Most likely, MPSE practitioners will need to evolve methods such that they are purposefully designed to provide the levels of data needed for ML/AI within this data–model construct.

The role of ML/AI in the broader context of integrated computational materials engineering (ICME) is still evolving. Although materials data has been a topic of interest in MPSE for some time [4, 5] ML/AI was not called out in earlier ICME reports and roadmaps [6, 7] or in the Materials Genome Initiative (MGI) that incorporates ICME in the MPSE workflow [8, 9]. However, it is an obvious component of a holistic ICME approach, supporting MGI goals in data analytics and experimental design as well as materials discovery through integrated research and development [9]. As detailed in the discussion below, ML/AI is rapidly being integrated into ICME and MGI efforts, supporting accelerated materials development, autonomous and high-throughput experiments, novel simulation methodologies, advanced data analytics, among others.

ML and AI technologies already impact our every-day lives. However, as practitioners of the physical sciences, we may ask what has changed, or why should a scientist be concerned now with ML and AI technologies for MPSE? Aren’t these technologies simply sophisticated curve fits or “black box” tools? Is there any physics there? Less skeptically and more objectively, one might also ask what are the important achievements from these tools, and how are those achievements related to familiar physics? Or, how can one best apply the newest advances in ML and AI to improve MPSE results? Speculating still further, why are there no emerging AI-based engineering design systems that recognize component features, attributes, or intended performance to make recommendations about directions for final design, manufacturing processes, and materials selections or developments? Such systems are possible over the next 20 years. Indeed, Jordon and Mitchell suggest that “. . . machine learning is likely to be one of the most transformative technologies of the 21st century. . .” [10] and therefore cannot be neglected in any long-range development of engineering practices.

The present overview is intended to serve as a selective introduction to ML and AI methods and applications, as well as to give perspective on their use in the MPSE fields,

especially for modeling and simulation. The computer science and related research communities are producing in excess of 2000 papers per year over the last 3 years (more than 15,000 in the last decade) on new algorithms and applications of ML technologies.¹ One cannot hope to offer a comprehensive review and discussion of these in a readable introductory review. As such, we examined perhaps 10% of recent literature and chose to highlight a small fraction of the papers examined. These reveal selected aspects of the field (perhaps some of which are lesser known) that we believe should capture the attention of MPSE practitioners, knowing that the review will be outdated upon publication.

Selected Context from Outside of MPSE

Readers may already be familiar with applications of ML- and AI-based commercial technologies, e.g., music identification via real-time signal processing on commodity smartphone hardware; cameras having automatic facial recognition; and recommendation systems for consumers that inform users about movies, news stories, or products [11, 12]. Further, AI technologies are used to monitor agricultural fields for insect types and populations, to manage power usage in computer server centers exceeding human performance, and are now being deployed in driver-assisted and driverless vehicles [13–16].

Just since 2016, a data-driven, real-time, computer vision and AI system has been deployed to identify weeds individually in agricultural fields and to locally apply herbicides, as a substitute for broadcast spraying [17]. Google switched its old “rules-based” language translation system to a deep-learning neural network-based system, realizing step-function improvements in the quality of translations, and they continue to grow that effort and many others around deep learning, abandoning rules-based systems [18, 19]. The games of “Go,” “Chess,” and “Poker” have been mastered by machines to a level that exceeds the play of the best human players [20–23]. Perhaps more important to MPSE, the new power of deep-learning networks was vividly shown in 2012, when researchers not only made step-function improvements in image recognition and classification but also surprisingly discovered that deep networks could teach themselves in an *unsupervised* fashion [24, 25]. Most recently, a self-taught *unsupervised* gaming machine exceeded the playing capability of the prior “Go” champion, also a machine that was developed with human supervised learning [26]. For selected instances, the machines can now even self-teach tasks better than the best-skilled human experts! The

powers and applications of ML/AI tools are expanding so rapidly that it is hard to envisage any aspect of MPSE or multiscale modeling and simulation, or engineering overall, that will not be impacted over the next decade. Our primary challenge is to discern how such capabilities can be best integrated into MPSE practices as standard methods, and for implementing them in appropriate ways as soon as possible.

Background and Selected Terms

To better understand aspects of the current ML/AI revolution, it is useful to consider selected background and terms from literature about the field. AI as a field of study has been around since the middle 1950s; however, it is the recent growth in data availability, algorithms, and computing power that have brought a resurgence to the field, especially for ML based on deep-learning neural networks (DLNN) [25, 27]. In practice, it has become important to distinguish the term “AI,” that is now most commonly associated with having machines achieve specific tasks within a narrow domain or discipline, from the term “artificial general intelligence” (AGI) that embodies the original and futuristic goal of having machines behave as humans do. The former is in the present while the latter is likely beyond foreseeable horizons.

ML has long been used for non-linear regression, to find patterns in data, and served as one approach for achieving AI goals [28, 29]. Three types of learning are commonly recognized as “supervised” where the system learns from known data; “unsupervised” where the unassisted system finds patterns in data; and, “reinforcement” learning where the system is programmed to make guesses at solutions and is “rewarded” in some way for correct answers, but is offered no guidance about incorrect answers. All three modes are used at today’s frontiers.

For the purposes of this overview, “data science” is a general term that implies *systematic acquisition and analysis, hypothesis testing, and prediction* around data. The field thereby encompasses wide-ranging aspects of the information technologies employed in data acquisition, fusion, mining, forecasting, and decision-making [30]. For example, all aspects of data science would be employed for autonomous systems. Alternatively, materials “informatics” is focused on analysis of materials data to modify its form and to find the most effective use of the information; i.e. materials informatics is a subset of materials data science. Aspects of these concepts are shown schematically in Fig. 1. Data sciences, informatics, and some ML technologies are related to each other, and selectively were used in research and engineering for over half a century. However, until the last decade, their impact was minimal on materials and processes development, structures engineering, or

¹As gleaned from Google Scholar Internet searches by year.

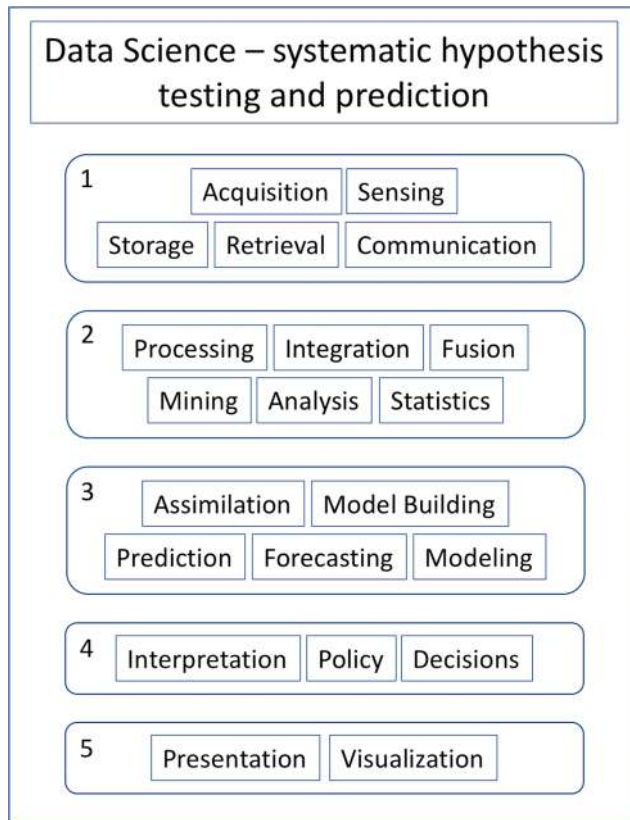


Fig. 1 Data science may be considered as the technologies associated with acquiring data, forming and testing hypotheses about it, and making predictions by learning from the data. Five domains of activity are evident: (1) data acquisition technologies; (2) processing the data and making analyses of it; (3) building models and making forecasts from the data; (4) decision-making and policies driven from the data; and (5) visualizing and presenting the data and results. “Informatics” has primarily been involved with items 2 and 3 and has expanded slightly into items 1 and 5. ML principally encompasses items 1–3 and 5. AI usually encompasses items 1–4, while placing emphasis on item 4

the experimental methods used for parameterizing and verifying models. The challenges in MPSE are simply too complex and data was too limited and expensive to obtain. Now, studies do show that the ML technologies can find relationships, occasionally discover physical laws, and suggest functional forms that may otherwise be hidden to ordinary scientific study, but these are few [29, 31].

Historical efforts in ML attempted uses of “artificial neural networks” (ANN or NN) to mimic the neural connections and information processing understood to take place in human brains (biological neural networks or BNN). In a fashion that loosely mimics the human brain, these networks consist of mathematical frameworks that define complex, non-linear relationships between input information and outputs. Generally, for all network learning methods, the ANN contains layers of nodes (matrices) that hold processed data that was transformed by the functional

relationship that connects the nodes. A given node receives weighted inputs from a previous layer, performs an operation to adjust a net weight, and passes the result to the next layer. This is done by forming large matrices of repeatedly applied mathematical functions/transforms connecting nodes, and expansion of features at each node. To use the network, one employs “training data” of known relationship to the desired outputs, to “teach” the networks about the relationships between known inputs and favorable outputs (the weights). By iteration of the training data, the networks “learn” to assign appropriate weighting factors to the mathematical operations (linear, sigmoidal, etc.) that make the connections, and to find both strong and weak relationships within data.

Importantly, the early networks typically had only one-to-three hidden layers between the input and output layers, and a limited number of connections between “neurons;” thus, they were not so useful for AI-based decision-making. Until recently, computers did not have the capacity and algorithms were underdeveloped to permit any deeper networks or significant progress on large-scale challenges [28, 32–34]. The techniques fell short of today’s deep-learning tools connected to AI decision-making. Consequently, with few exceptions, the historical technical approaches for achieving AI, even within specific applications, have been arduously tied to “rules,” requiring human experts to delineate and update the rules for ever-expanding use cases and learned instances—that is until now.

Generally speaking, *today’s ANN have changed completely*. The availability of vast amounts of digital data for training; improvements to algorithms that permit new network architectures, ready training, and even self-teaching; and parallel processing and growth in computing power including graphics processor unit (GPU) and tensor processing unit (TPU) architectures have all led to deep-learning neural networks DLNN or “deep learning” (DL). Such DLNN often contain *tens-to-thousands* of hidden layers, rather than the historical one-to-three layers (thus the term “deep learning”). These advanced networks can contain a billion nodes or more, and many more connections between nodes [19, 25]. Placing this into perspective, the human brain is estimated to contain on the order of 100-to-1000 trillion connections (synapses) between less than 100 billion neurons. By comparison, today’s best deep networks are still 4–5 orders of magnitude smaller than a human brain. However, BNN still serve as models for the architectures being explored, and being only 4–5 orders of magnitude smaller than a human BNN still provides tremendous, unprecedented capabilities.

Within DL technologies, there are several use-case-dependent architectures and implementations that provide powerful approaches to different AI domains. Those

based on DLNN typically require extensive data sets for training (tens of thousands to millions of annotated instances for training). As mentioned previously, this presents a major challenge for their use in MPSE that most likely will have to be mitigated using simulated data in symbiosis with experimentally acquired data. “Convolutional and de-convolutional or more appropriately transposed convolutional neural networks” (CNN and TCNN, respectively) and their variations have three important network architecture attributes including 3D volumes of node arrays and deep layers of these arrays, local connectivity such that only a few 10s of nodes communicate with each other at a time, and shared weights for each unit of connected nodes. These attributes radically speed up training, permitting the all-important greater depths. During use, the mathematical convolution (transposed convolution) operation allows concurrent learning and use of information from all of the locally narrow but deep array elements. Architecturally, the networks roughly mimic the BNN of the human eye, and have proven their effectiveness in image recognition and classification tasks, now routinely beating human performance in several tasks [34–37].

Several even more advanced DLNN architectures emerged recently including “Recurrent” (RNN) that have taken on renewed utility in their use for unsupervised language translation [38], “Regional” (R-CNN) used for image object detection, [39] and “Generative Adversarial Networks” (GAN) [40] for unsupervised learning and training-data reduction, to name but a few (for overviews and reviews, see work by Li and by Schmidhuber [41–43]). Each of these architectures adapts DL to different task domains. For example, language translation and speech recognition benefit by adding a form of memory for time series analysis (RNN). GAN include *simulated-plus-unsupervised* training, or $S + U$ learning, for which simulated data is “corrected” using *unlabeled* real data, as shown in Fig. 2. Reinforcement learning technology, of which GAN are a subset, was used for the self-taught machine that mastered “Go” and has been used for the most recent language translation methods [26, 44]. Further, in a task that has similarities to aspects of MPSE, $S + U$ training was used to correct facial recognition systems for the effects of pose changes, purely from simulated data [45, 46]. Given the widely expanding applications of DL, there is a high likelihood that architectures, algorithms, and methods for training will continue to evolve rapidly over the next 3–5 years.

Perhaps the most challenging goal for ML/AI methods is to continue the expansion of autonomous systems, especially for MPSE research and development [47–52]. Slowly, these systems are making their way into life sciences, drug discovery, and the search for new functional materials [47, 52, 53]. ML enabled progress in materials composition discovery does not of-and-to-itself imply mastery of the

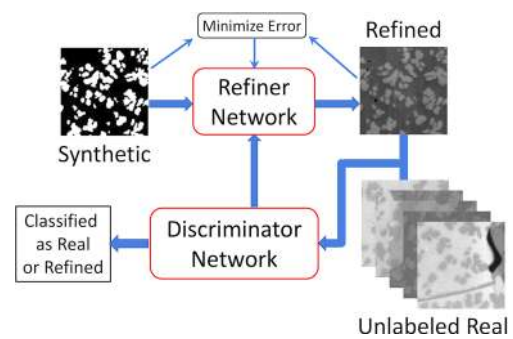


Fig. 2 One may envisage producing materials microstructure models using $S + U$ learning. In the schematic, numerous *unlabeled* real images are fed into a “discriminator” CNN that learns from both real and refined synthetic images, and classifies images from the refiner as real or fake. The “refiner” is a CNN that operates on simulated microstructure images, enhancing them toward the realism of measured micrographs. The simulations may be used to sample more microstructure spaces, or nuances of microstructure, that are difficult or expensive to measure experimentally, while the measured micrographs enhance the simulated images adding realism. Adapted from [45]

processing and microstructure design space. For these latter design challenges, new autonomous tools are needed, largely based on imaging sciences being better coupled to high-throughput experimentation. Most recently, Zhang et al. took steps toward autonomy for materials characterization by using ML for dynamic sampling of microstructure, while Kraus et al. demonstrated the power of automatic classifiers for biological images [54, 55]. Further, given the advances in deep learning and crowd sourcing used for annotating image and video data [56–58], perhaps the seeds have been sewn for long-range development of systems to autonomously map ontologies for materials data, while keeping them continuously updated. One may envisage that when combined with DL for computer vision, long-range developments should permit autonomous materials characterization, and ultimately to the mastery of materials hierarchical microstructure for new materials design through autonomous microstructure search.

Selected Applications and Achievements in Materials and Structures

For the case of multiscale materials and structures, we consider applications of ML/AI techniques in two main areas. First, selected examples illuminate accomplishments for materials discovery and design. While not necessarily noted in the works, these tie directly to MGI and IMCE goals. These are followed by some examples applying ML/AI methods in structures analysis. Here again, the MGI goal of accelerating materials development, deployment,

and life-cycle sustainment directly ties to the structures analysis aspect of ML/AI.

Materials Discovery

For about the last two decades, ML for materials structure-property relationships has used comparatively mature informatics methods. For example, principal component analysis (PCA) operating upon human-based materials descriptors can lend insights into data. For PCA, the descriptor space is transformed using mathematics to maximize data variance in the descriptor dependencies, yielding a new representation for finding relationships. The new representation usually involves a dimensionality reduction to the data resulting in a loss of more nuanced aspects of the data. Past efforts used microstructure descriptors (in a mean-field sense), such as average grain size, constituent phase fractions or dimensions, or material texture, and sought to relate these to mean-field properties, such as elastic modulus or yield stress [59, 60]. In the absence of high-throughput computational tools for obtaining materials kinetics information, structure-property relationships, and extreme-value microstructure influences, other studies resorted to experimental data to establish or to narrow the search domains for new materials [61–70].

Our expectation is that these approaches will also become more efficient, reliable, and prevalent in the coming decade or more, particularly since open data, open-source computing methods, and technology businesses are becoming available to support the methods and approaches [71–74]. Further, advancements in materials characterization capabilities, process monitoring and sensing methods, and software tools that have taken place over the previous 20 years [75] are giving unprecedented access to 3- and 4D materials microstructure data, and huge data sets pertaining to factory-floor materials processing. Such advancements suggest that the time is ripe for bringing ML/DL/AI tools into the materials and processes domain.

Mechanics, Mechanical Properties, and Structures Analysis

Historically, multiscale modeling, structures analysis, and structures engineering have all benefited from ML/AI tools. Largely because of their general ability to represent non-linear behaviors, different forms of ANN architectures have been used since the 1990s to model materials constitutive equations of various types [76, 77], optimize composites architectures [78], and to represent hysteresis curves or non-linear behavior in various applications (such as fatigue) [79–81]. The closely related field of non-destructive evaluation also benefited from standard ANN techniques [82], though this field is not treated herein. Further, these methods were

used for more than two decades in applications such as active structures control [83–85], and even for present-day flight control of drones [86]. Today, DL is bringing entirely new capabilities to structures and mechanics analysis.

In more recent work, ML methods are being used to address challenging problems in non-linear materials and dynamical systems and to evolve established ANN and informatics methods [87–90]. Further, newer deep learning and other powerful data methods are beginning to be employed. For example, Versino et. al. showed that symbolic regression ML is effective for constructing constitutive flow models that span over 16 orders of magnitude in strain rate [91]. Symbolic regression methods involve fitting arbitrarily complex functional forms to data, but doing so under constraints that penalize total function complexity, thus resulting in the simplest sufficient function to adequately fit the data [92]. Integrated frameworks are also beginning to appear [93]. These suggest a promising future that we consider more fully in what follows.

A Perspective on the Unfolding Future

Looking forward, it is appropriate to consider the question, what has changed in ML/AI technologies, and what has fostered the explosive growth of this field? Also, how might these advancements impact MPSE? This section considers these questions and provides selected insights into the prospects for ML/DL/AI and their associated technologies. The perspective focuses on examples of using these tools for materials characterization, model development, and materials discovery, rather than a complete assessment of ML, DL, AI, and data science or informatics. Further, the emphasis is on achievements from 2015 to the present, with many examples from the last year or so.

Imaging and Quantitative Understanding

Most recently, computer vision tools, specifically CNN/DL methods, were applied to microstructure classification, thus forming initial building blocks for objective microstructure methods and opening a pathway to advanced AI-based materials discovery [63, 94–98]. By adopting CNN tools developed for other ML applications outside of engineering, these researchers were able to objectively define microstructure classes and automate micrograph classification [95, 97, 98]. Figure 3 shows an example of the methods being applied to correlate visual appearance to processing conditions for ultrahigh carbon steel microstructures. Today, even while they remain in their infancy, such methods are demonstrated to have about a 94% accuracy in classifying types of microstructure, and they rival human capabilities for these challenges [94, 97].

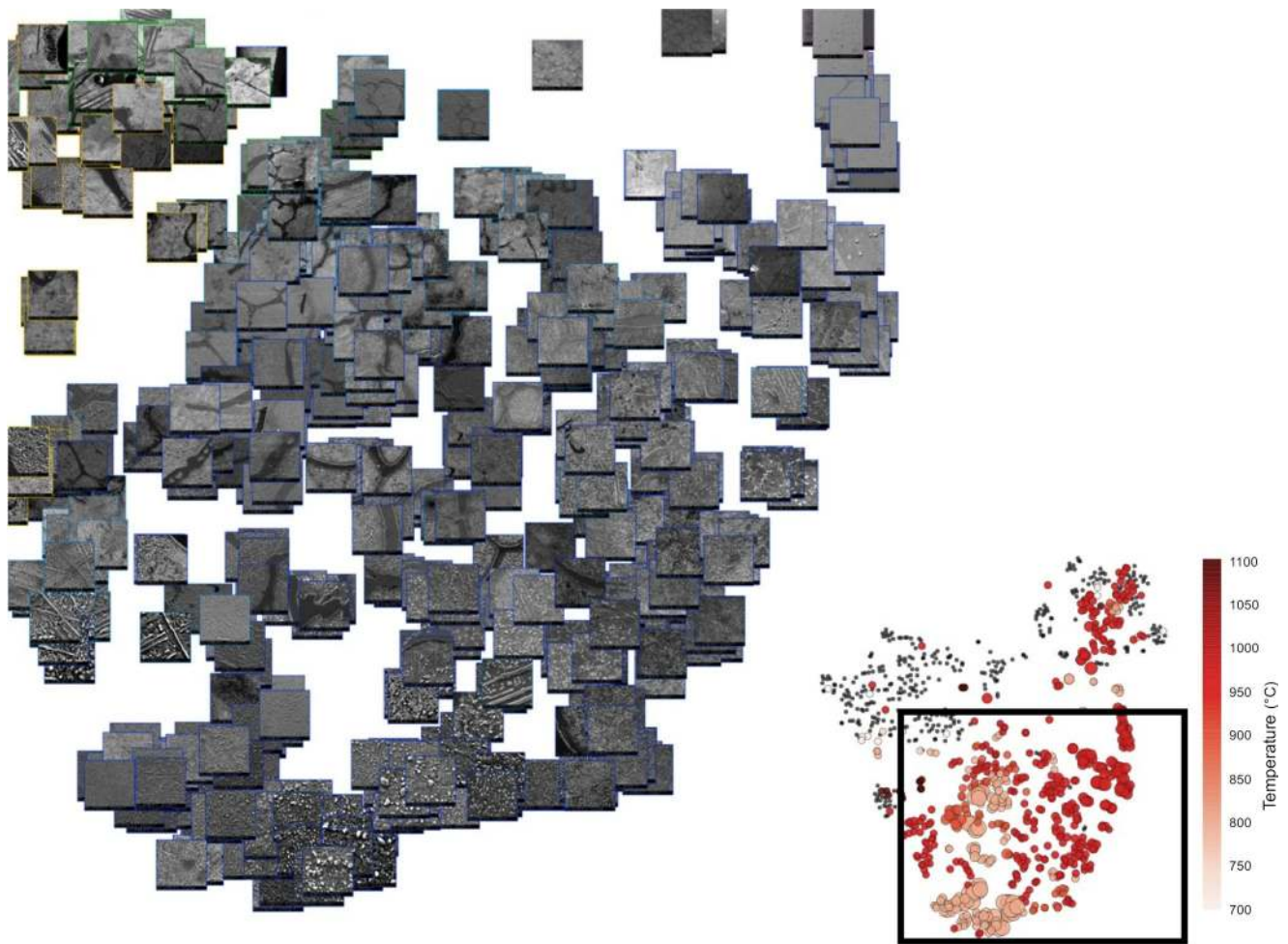


Fig. 3 A t-SNE map (see L. van der Maaten and G. Hinton, Visualizing data using t-SNE, *Jrnl. Mach. Learn. Res.*, 9 (2008), p. 2579.) of 900 ultrahigh carbon steel microstructures in the database by Hecht [99] showing a reduced-dimensionality representation of multi-scale CNN representation of these microstructures [94]. Images are grouped by visual similarity. The inset at the bottom right shows the annealing conditions for each image: annealing temperature is indicated by the color map and annealing time is indicated by the relative marker size. The map is computed in an unsupervised fashion from the structural

information obtained from the CNN; microstructures having similar structural features tend to have similar processing conditions. This is especially evident tracing the high-temperature micrographs from the bottom of the figure to the top right: as the annealing time increases, the pearlite particles also tend to coarsen. Note—the Widmanstatten structures at the left resulting from similar annealing conditions were formed during a slow in-furnace cooling process, as opposed to the quench cooling for most of the other samples

These early materials image classifiers are also showing promise for improved monitoring of manufacturing processes, such as powder feed material selection for additive manufacturing processes [100, 101]. Over the next 20 years, autonomous image classification will be common, with the classifiers themselves being trained in an unsupervised fashion, choosing the image classes without human intervention, thereby opening entire new dimensions to the MGI/ICME paradigms [34]. This means that materials and process engineers are likely to have machine companions monitoring all visual- and image-based aspects of their discipline, in order to provide guidance to their decision-making, if not making the decisions autonomously. In the coming decades, machine-based methods may have aggregated

sufficient knowledge to autonomously inform engineers without having any prior knowledge of the image data collection context. They will likely operate autonomously to identify outliers in production systems or other data. Thus, one should expect radical changes to materials engineering practices, especially those based upon image data.

Further, current work by DeCost, Holm, and others is beginning to address the challenge of materials image segmentation. While the use of DL and CNN methods has recently made great strides for segmenting and classifying pathologies in biological and medical imaging [35, 102, 103], the methods are completely new in their application to materials and structures analysis. Figure 4 shows an example metal alloy microstructure image segmentation

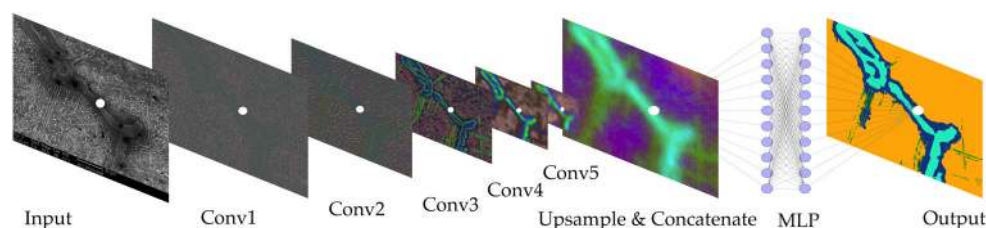


Fig. 4 CNN (PixelNet architecture, A. Bansal et. al., Pixelnet: Representation of the Pixels, By the Pixels, and for the Pixels, CoRR (2017). arXiv:1702.06506 [cs.CV]). trained to segment ultrahigh carbon steel micrographs. This schematic diagram shows the intermediate representations of an ultrahigh carbon steel micrograph [99] being segmented by the CNN into high-level regions: carbide network (cyan), particle matrix (yellow), denuded zone (blue), and Widmanstten cementite (green). Such a CNN can support novel automated workflows in microstructure analysis, such as high-throughput

quantitative measurements of the denuded zone width distribution. The PixelNet architecture uses the standard convolution and pooling architecture to compute multiscale image features (Conv1–5), which are up-sampled and concatenated to obtain multiscale high-dimensional features for each pixel in the input image. A multilayer perceptron (MLP) classifier operating on these features produces the final predicted label for each pixel; the entire architecture is trained end-to-end with no post processing of the prediction maps. Source for figure: [107]

using a CNN tool. Note how the CNN learns features with increasing depth (layers) of the network, going from left-to-right in the image. Given that image segmentation and quantification (materials analytics) is among the major obstacles to bringing 3D (and 4D) materials science tools into materials engineering, the ML methods represent nascent capabilities that will result in dramatic advances in 5 years and beyond. Further, current computer science and methods research is focusing on understanding the transference capabilities of CNN/DL tools [104]. Transference refers to understanding and building network architectures that are trained for one type of image class or data set, and then using the same trained network to classify completely different types of images/features on separate data, without re-training.

As the methods mature, there is a high likelihood that the definitional descriptors for materials hierarchical structure (microstructure) will also evolve and be defined by the computational machines, more so than by humans. Those in turn will need to be integrated with modeling and simulation methodologies to have the most meaningful outcomes. That is, since the ML/AI methods are devised to operate on high-dimensional, multi-modal data, they also bring new, unfamiliar parameter sets to the MPSE modeling and simulation communities that define the output from the analyses. These may bring challenges for engineering design systems as they seek to establish meaningful data and informatics frameworks for futuristic designs. For addressing this nascent challenge, ICME paradigms must evolve to be better coupled to engineering design.

The DL-based image analysis tools are already being made available to users via web-based application environments and open-source repositories that help to lower the barrier to entry into this new and dynamic field [105, 106]. Recently, at least one company, Citrine Informatics [71],

has formed with the intended purpose of using informatics and data science tools, together with modern ML tools, to enhance materials discovery. One may also expect that the high driving force for having such tools available for medical imaging analysis, and their use for other aspects of computer vision, will keep these types of tools emerging at a rapid pace. This implies that materials and structures practitioners might be well advised to keep abreast of the advancements taking place outside of the materials and structures community, and to assure that the progress is transferred into the MPSE domain.

Materials and Processes Discovery

Computational materials discovery and design, as well as high-throughput experimental search and data mining, have been a visible domain of MGI-related research and development. These practices too are seeing significant benefit from current ML/AI tools. Some of these advancements were recently summarized [108–117]. However, the possibilities for materials compositions, microstructure, and architectures are vast—beyond human capacity alone to comprehensively search, discover, or design. Thus, machine-assisted and autonomous capabilities are needed to perform comprehensive search. More recently, much attention has been given to ML for discovering functional compounds [49, 53, 108, 118–121]. Some research efforts computed ground states, selected ground-state phase diagrams, and physical properties for comparatively simple (up to quaternary) classes of inorganic compounds, while other efforts computed chemical reactivity and functional response for organic materials [108, 118, 122–124]. Notable studies demonstrate that machine learning applied to appropriate experimental data is more reliable or convenient than DFT-based simulations [125–128]. Nevertheless, exploring the

complexity of materials and processes for finite temperatures, extending into kinetics-driven materials states or realizing hierarchical materials structures and responses, requires so far unachieved search capability over many more spatiotemporal parameters. One needs to be able to efficiently acquire information and then to perform search and classification, over vast portions of multiscale materials chemistry and kinetics (transport), structure (crystallography and morphology), response (properties) space. The next essential building block for widespread materials discovery is linking the tools for composition search to synthesis, processing, and materials response [50]. In these respects, the MGI is only in its infancy.

What is needed for achieving those linkages in an objective fashion is to build them upon spatiotemporal hierarchical microstructure (from electrons and atoms to material zones/features and engineering designs). Most likely, ML/DL/AI tools will play a pivotal role in establishing these complex relationships. However, the MPSE community remains limited by the relatively small databases of microstructure (spatiotemporal) information in comparison to the requirements that appear to be necessary for an ML/AI-driven approach. One clear pathway for circumventing that formidable barrier is to take advantage of the considerable capabilities for materials modeling and simulation that are now well established within MPSE. Methods such as the $S + U$ GAN technique discussed previously and shown in Fig. 2 must be generalized to make full use of both simulation and experimental data, beyond microstructure data. This implies that a long-range theme in MPSE practices (also within the MGI) needs to be centered around building *models for the methods by which data are produced*, thus allowing for the symbiosis between real and synthetic data that is so powerful in an ML/AI environment. Having these tools will be a major advantage for completing the reciprocity relationships between microstructure–properties–models that are a foundation for MPSE design. Further, there is a high likelihood that over the next 20 years, the growth in computer vision and decision-making systems will make great strides in achieving larger amounts of data through computational, high-throughput, and autonomous AI-based systems [51, 53, 129, 130].

As more curated and public databases for materials information lead to increasing data availability, the methods and benefits of ML/AI are likely to grow rapidly [131–138]. Notable are two recent actions that make relevant materials data more openly available. First is a private sector entry into the domain of publicly accessible large-scale materials databases, including an effort to simplify use of informatics/ML tools. Citrine Informatics has adopted a business model that supports open-access use of the ML tools they have deployed, provided that the user data being analyzed is contributed to the Citrine database. (Citrine's

tools are also available for proprietary use on a fee basis.) Second, the Air Force Research Laboratory has posted data pertaining to additive manufacturing, along with a data-use challenge, analogous to the Kaggle competitions established more than a decade ago for data science practitioners [139, 140].

Recent progress using large-scale accessible databases also shows success in searching for new functional materials [130, 141–143]. These searches involve computing compound or molecular structures and screening them for selected functional properties. Given these successes, it is hard to imagine functional material development continuing to be performed in a heuristic manner after the passing of the next decade.

Far more challenging are searches for (i) synthesis and process conditions; (ii) materials transformation kinetics; and (iii) microstructures with responses that satisfy design requirements. In these areas too, progress is rapid using data-driven methods and ML techniques. Machine-learning tools are being applied today to guide chemical synthesis/processing tasks (see next section). However, analogous frameworks for metals or composites processing are barely emerging. Ling et. al. demonstrated that process pathways can be optimally sought using real-time ML methods to guide experiments [144]. The models not only indicate what experiment is the next-most-useful one but they also permit bounding error on the model to indicate how useful an experiment will be. Similar methods were already developed for optimally sampling microstructure when collecting time-consuming and expensive data, such as electron backscatter diffraction (EBSD) scans in 3D [54, 145, 146]. From these and other developments, it is neither too difficult to imagine ways to implement such tools for enhanced high-throughput data acquisition and learning nor too difficult to conjecture that the ML-based, high-throughput methods will markedly expand over the next 10 years. One may expect that the advancements will lead to both new materials and materials concepts and to more robust bounding of manufacturing processes for existing materials.

One additional area that is ripe for development is the connection of ML/DL/AI tools to applications in ICME and the larger MGI that involve “inverse design.” A central theme in ICME is the replacement of expensive (both in terms of time and resources) experimentation with simulation, especially for materials development. In effect, ICME seeks to replace the composition and process search, or statistical confidence obtained through repeated physical testing, with those developed through simulation. This requires that (i) model calibration, (ii) model verification and validation, and (iii) model uncertainty quantification (UQ) be carried out in a more complete and systematic manner than is common within the materials community. Additionally, ICME applications extend into inverse design

problems in that the objective is to establish a material and processing route that optimize a set of properties and performance criteria (including cost), while most material models are *material*→*properties*. The design problem of interest requires inverting the model typically through numerical optimization. The additional uncertainty quantification requirements and design optimization mean that the models will be exercised for a large number of times. For even simple models, this can be very expensive in terms of computational resources and rapidly becomes computationally intractable for 3D spatio/temporally resolved simulations. ML can relieve this computational bottleneck by serving as “reduced order” or “fast acting” models. Once trained, ML models can be exercised very quickly and multiple instantiations can be exercised in parallel on typical computer systems. More importantly, the inverse model *properties*→*materials* can be trained in parallel with the forward model speeding up the design process. The critical open-research question for the community becomes “How do we train ML models for ICME applications with limited experimental data and how do we ensure proper UQ?”

Computational Chemistry Methods

The early impacts of ML/DL/AI methods are being realized today in the fields of computational chemistry, chemical synthesis, and drug discovery [147–150]. One compelling demonstration is the power of using machine-based pre-planning for chemical synthesis [149, 151]. These applications use ML in a data-mining-like mode to learn the complex relationships involved in molecular synthesis from known past experience. Analogous applications are well underway to search for inorganic materials using computed large-scale databases and applying ML/AI for learning complex non-linear relationships between variables [152, 153].

In yet a different mode, the ML methods are also having an impact on computational quantum chemistry calculations themselves that are used to predict molecular stability, reactivity, and other properties. Historically, both quantum chemistry and density functional theory (DFT) codes are widely known to be limited by poor computational scaling ($O(N^6)$ and $O(N^3)$, respectively) that constrains accessible system sizes [154]. However, recent work is revealing that ML/AI methods can learn the many-body wave functions and force fields, thereby mitigating the need for some computationally intensive investigations [155–160]. These methods have just been demonstrated in the past year, and are some of the many promising frontiers in ML. Clearly, as these methods are brought in to widespread practice, the landscape for multiscale materials and structures simulation will be drastically improved.

Multiscale Mechanics and Properties of Materials and Structures

Recently, Geers and Yvonnet offered perspectives on the future of multiscale materials and structures modeling [161], and McDowell and LeSar did the same for materials informatics for microstructure-property relationships [162]. Both perspectives pointed out the considerable challenges remaining in the field. Note however that these authors did not address the possible role of ML/AI in pushing the frontier forward, in part because there appears to be much slower progress in applying ML/AI in these fields. At the smallest scales, one major emerging application is the invention of “machine-learning potentials” for atomistic simulations [163, 164]. These have good prospects for speeding the development of interatomic potential functions while improving their reliability and accuracy, especially for systems that include covalent and ionic bonding.

At coarser scales, there are limited advances emerging for developing constitutive models [87, 88, 91], modeling hysteretic response [89], improving reduced order models [165, 166], and even for optimizing numerical methods [167]. At still coarser scales, there is research to understand and model complex dynamical systems and to use ML methods for dimensionality reduction [168, 169].

Multiscale materials and structures modeling also includes advanced experimental methods that will benefit from ML/AI implementations, but the field is in its infancy. For example, the digital image correlation (DIC) method has been developed over the last decade to the point that it is now successfully used to learn constitutive parameters for materials [91, 170, 171]. However, using discrete dislocations dynamics simulations as a test bed, Papanikolaou et. al. recently demonstrated that ML tools can extract much more information from DIC measurements, suggesting new ways for using the experimental DIC data, especially in lock-step with simulation [172, 173]. As the ML/AI methods are better understood by the MPSE community, one may expect cross-the-board advances in experimental methods, especially for model development, validation, and uncertainty quantification.

Noteworthy Limitations: Relationships to Physics, Software, and Education

As this overview suggests, the prospects for ML/AI methods to bring significant advances to the domain of materials and structures modeling and simulation are exceptionally high. Indeed, the field is advancing so rapidly that it is difficult to estimate how radical the advances may be, and over what time-frame. More generally, some even suggest that the prospect of true AGI is on the horizon within the next 10

years [174], implying that the technology space for MPSE will simply be unrecognizable by today's measures more quickly than anticipated. Another assessment was provided by Grace et. al. suggesting a longer time horizon for AGI but nonetheless predicting significant general impacts over the coming few decades [175].

In the face of such a radical change and technical revolution, it is prudent to maintain caution and to be wary of shortcomings. For example, there remains considerable debate in the AI community regarding interpretability of the DL methods and models [176–179]. To many, the DL methods appear as “black boxes” and in some respects function as such. While black boxes may be valid solutions for some applications (e.g., performing repetitive tasks), they may be unacceptable for others, particularly where the cost of a wrong answer is high (e.g., flight qualifying an aerospace component).

Other work points to the questions of reproducibility, reusability, and robustness of ML/DL/AI methods, especially in the broad domain of reinforcement learning [180–182]. Fortunately, having these important issues raised is beginning to lead to recommendations of best practices for ML, and to software platforms to facilitate those practices [183–186]. Knowing the existence of such issues again suggests a need for caution when deploying ML/DL/AI methods in MPSE practices.

For high consequence applications, engineers must insist upon ML/DL/AI methods that make decisions based on underlying scientific principles. One research frontier in computer sciences is exactly this pursuit of an understanding of how such technologies work and their relationships to physical science [176–179, 187–189]. For this challenge, the MPSE community may be uniquely positioned in several respects. First, the complexity of the MPSE fields together with the high-value-added products and systems to which they lead provides strong driving forces and widespread application domains for advancements via ML/DL/AI tools. Second and perhaps more important, the physical sciences have long been engaged with not only retrospective modeling for explanation of the physical world but also “forward” or “system” modeling to provide a manifold for efficient data collection and constraints on predictive tools. Most recently, the methods are adding powerful capabilities in materials characterization, for example [54, 190]. The power of these modeling frameworks relative to the ML/DL/AI understanding challenges is that the models can provide an ever-expanding source of “phantom instances” for materials and processes that are completely known virtual test beds to use within the ML/DL/AI tools.

Finally, while there are laudable efforts to introduce the ML/DL/AI tools into widely accessible and somewhat user-friendly software libraries and codes [106, 191–194], it is

not clear that the educational systems, both formal and informal, are keeping pace with the developments or providing ML/DL/AI models and systems to MPSE practice. This suggests real risks of models and systems being developed, perhaps from outside of MPSE, without enough understanding of their limits, or consequences of their failures. Naturally, a strategy for educating MPSE practitioners in the use of these advanced tools is needed in the very near term. Perhaps much of this could be achieved through appropriately structured teaming around ML/DL/AI development for specific MPSE challenges.

Summary and Conclusions

The fields of machine learning, deep learning, and artificial intelligence are rapidly expanding and are likely to continue to do so for the foreseeable future. There are many driving forces for this, as briefly captured in this overview. In some cases, the progress has been obviously dramatic, opening new approaches to long-standing technology challenges, such as advances in computer vision and image analysis. Those capabilities alone are opening new pathways and applications in the ICME/MGI domain. In other instances, the tools have only provided evolutionary progress so far, such as in most aspects of computational mechanics and mechanical behavior of materials. Generally speaking, the fields of materials and processes science and engineering, as well as structural mechanics and design, are lagging other technical disciplines in embracing ML/DL/AI tools and exploring how they may benefit from them. Nor are these fields using their formidable foundations in physics and deep understanding of their data to contribute to the ML/DL/AI fields. Nonetheless, technology leaders and those associated with MPSE should expect unforeseeable and revolutionary impacts across nearly the entire domain of materials and structures, processes, and multiscale modeling and simulation over the next two decades. In this respect, the future is now, and it is appropriate to make immediate investments in bringing these tools into the MPSE fields and their educational processes.

Acknowledgements The authors wish to thank D. Furrer, X. Liu, and A. Boyne of Pratt & Whitney, S. Donegan and C. Ward of the Air Force Research Laboratory, M. Jackson of BlueQuartz Software, and B. DeCost of the National Institute of Standards and Technology, for valuable conversations and insights that helped inspire the effort behind this overview.

Funding Information DMD acknowledges NASA Glenn Research Center, Pratt & Whitney, and The Ohio State University for partial funding of the effort. SRN would like to acknowledge DARPA Young Faculty Award grant D15AP00103 and the US Department of Energy National Energy Technology Laboratory award DE-FE0027776 for partial funding of the effort. EAH is supported by the US National Science Foundation award numbers DMR-1507830 and DMR-1826218.

References

- Jardine P (2018) A reinforcement learning approach to predictive control design: autonomous vehicle applications. PhD thesis, Queens University, Kingston
- Li L, Ota K, Dong M (2018) Human-like driving: empirical decision-making system for autonomous vehicles. *IEEE Transactions on Vehicular Technology*
- Redding JD, Johnson LB, Levihn M, Meuleau NF, Brechtel S (2018) Decision making for autonomous vehicle motion control. US Patent App 15:713,326
- National Research Council Defense Materials Manufacturing and Infrastructure Standing Committee (2014) Big data in materials research and development: summary of a workshop. National Academies Press, Washington
- Warren J, Boisvert RF (2012) Building the materials innovation infrastructure: data and standards. US Department of Commerce, Washington. <https://doi.org/10.6028/NIST.IR.7898>
- National Research Council Committee on Integrated Computational Materials Engineering (2008) Integrated computational materials engineering: a transformational discipline for improved competitiveness and national security. National Academies Press, Washington
- The Minerals, Metals & Materials Society (2013) Integrated computational materials engineering (ICME): implementing ICME in the aerospace, automotive, and maritime industries TMS, Warrendale, PA
- National Science and Technology Council Committee on Technology Subcommittee on the Materials Genome Initiative (2011) Materials Genome Initiative for global competitiveness. National Science and Technology Council, Washington, DC. https://www.mgi.gov/sites/default/files/documents/materials_genome_initiative-final.pdf
- National Science and Technology Council Committee on Technology Subcommittee on the Materials Genome Initiative (2014) Materials Genome Initiative strategic plan. National Science and Technology Council, Washington, DC. http://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/mgi_strategic_plan_-_dec_2014.pdf
- Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. *Science* 349(6245):255–260
- Jones N (2014) The learning machines. *Nature* 505(7482):146
- Metz R (2015) Deep learning squeezed onto a phone. <https://www.technologyreview.com/s/534736/deep-learning-squeezed-onto-a-phone/>. Accessed 28 June 2018
- Silva DF, De Souza VM, Batista GE, Keogh E, Ellis DP (2013) Applying machine learning and audio analysis techniques to insect recognition in intelligent traps. In: 2013 12th international conference on machine learning and applications (ICMLA), vol 1. IEEE, pp 99–104
- Li K, JC Príncipe (2017) Automatic insect recognition using optical flight dynamics modeled by kernel adaptive arma network. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 2726–2730
- Vincent J (2017) Google uses deepmind AI to cut data center energy bills. <https://www.theverge.com/2016/7/21/12246258/google-deepmind-ai-data-center-cooling>. Accessed 28 June 2018
- Johnson BD (2017) Brave new road. *Mech Eng* 139(3):30
- Sowmya G, Srikanth J (2017) Automatic weed detection and smart herbicide spray robot for corn fields. *Int J Sci Eng Technol Res* 6(1):131–137
- Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, Krikun M, Cao Y, Gao Q, Macherey K et al (2016) Google's neural machine translation system: bridging the gap between human and machine translation. arXiv:160908144
- Lewis-Kraus G (2016) The great AI awakening. *The New York Times Magazine*, pp 1–37
- Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M et al (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529(7587):484–489
- Moravčík M, Schmid M, Burch N, Lisý V, Morrill D, Bard N, Davis T, Waugh K, Johanson M, Bowling M (2017) Deepstack: expert-level artificial intelligence in heads-up no-limit poker. *Science* 356(6337):508–513
- Brown N, Sandholm T (2017) Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*. <http://science.sciencemag.org/content/early/2017/12/15/science.aao1733>. <https://doi.org/10.1126/science.aao1733>
- Gibbs S (2017) Alphazero AI beats champion chess program after teaching itself in four hours. <https://www.theguardian.com/technology/2017/dec/07/alphazero-google-deepmind-ai-beats-champion-program-teaching-itself-to-play-four-hours>. Accessed 28 June 2018
- Le QV, Ranzato M, Monga R, Devin M, Chen K, Corrado GS, Dean J, Ng AY (2012) Building high-level features using large scale unsupervised learning. arXiv:11126209v5
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436
- Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A et al (2017) Mastering the game of go without human knowledge. *Nature* 550(7676):354
- Hutson M (2017) AI glossary: artificial intelligence, in so many words. *Science* 357(6346):19. <https://doi.org/10.1126/science.357.6346.19>
- MacKay DJ (2003) Information theory, inference and learning algorithms. Cambridge University Press, Cambridge
- Schmidt M, Lipson H (2009) Distilling free-form natural laws from experimental data. *Science* 324(5923):81–85
- Dhar V (2013) Data science and prediction. *Commun ACM* 56(12):64–73
- Rudy SH, Brunton SL, Proctor JL, Kutz JN (2017) Data-driven discovery of partial differential equations. *Science Advances* 3(4):e1602.614
- Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
- Kiser M (2017) Why deep learning matters and whats next for artificial intelligence. <https://www.linkedin.com/pulse/why-deep-learning-matters-whats-next-artificial-matt-kiser>. Accessed 28 June 2018
- Real E, Moore S, Selle A, Saxena S, Suematsu YL, Le Q, Kurakin A (2017) Large-scale evolution of image classifiers. arXiv:170301041
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542(7639):115
- Johnson R (2015) Microsoft, Google beat humans at image recognition. *EE Times*
- Smith G (2017) Google brain chief: AI tops humans in computer vision, and healthcare will never be the same. *Silicon Angle*
- Artetxe M, Labaka G, Agirre E, Cho K (2017) Unsupervised neural machine translation. arXiv:171011041
- He K, Gkioxari G, Dollár P, Girshick RB (2017) Mask R-CNN. arXiv:1703.06870
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: *Advances in neural information processing systems*, pp 2672–2680

41. Li Y (2017) Deep reinforcement learning: an overview. arXiv:170107274
42. Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
43. Baltrusaitis T, Ahuja C, Morency L (2017) Multimodal machine learning: a survey and taxonomy. arXiv:1705.09406
44. Lample G, Denoyer L, Ranzato M (2017) Unsupervised machine translation using monolingual corpora only. arXiv:171100043
45. Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W, Webb R (2016) Learning from simulated and unsupervised images through adversarial training. arXiv:161207828
46. Masi I, Tran AT, Hassner T, Leksut JT, Medioni G (2016) Do we really need to collect millions of faces for effective face recognition? In: *European conference on computer vision*. Springer, pp 579–596
47. King RD, Rowland J, Oliver SG, Young M, Aubrey W, Byrne E, Liakata M, Markham M, Pir P, Soldatova LN et al (2009) The automation of science. *Science* 324(5923):85–89
48. Sparkes A, Aubrey W, Byrne E, Clare A, Khan MN, Liakata M, Markham M, Rowland J, Soldatova LN, Whelan KE et al (2010) Towards robot scientists for autonomous scientific discovery. *Automated Experimentation* 2(1):1
49. Raccuglia P, Elbert KC, Adler PD, Falk C, Wenny MB, Mollo A, Zeller M, Friedler SA, Schrier J, Norquist AJ (2016) Machine-learning-assisted materials discovery using failed experiments. *Nature* 533(7601):73
50. Nosengo N et al (2016) The material code. *Nature* 533(7601):22–25
51. Nikolaev P, Hooper D, Perea-Lopez N, Terrones M, Maruyama B (2014) Discovery of wall-selective carbon nanotube growth conditions via automated experimentation. *ACS Nano* 8(10):10,214–10,222
52. Nikolaev P, Hooper D, Webber F, Rao R, Decker K, Krein M, Poleski J, Barto R, Maruyama B (2016) Autonomy in materials research: a case study in carbon nanotube growth. *npj Comput Mater* 2:16,031
53. Oses C, Toher C, Curtarolo S (2018) Autonomous data-driven design of inorganic materials with AFLOW. arXiv:1803.05035
54. Zhang Y, Godaliyadda G, Ferrier N, Gulsoy EB, Bouman CA, Phatak C (2018) Slads-net: supervised learning approach for dynamic sampling using deep neural networks. arXiv:180302972
55. Kraus OZ, Grys BT, Ba J, Chong Y, Frey BJ, Boone C, Andrews BJ (2017) Automated analysis of high-content microscopy data with deep learning. *Mol Syst Biol* 13(4):924. <https://doi.org/10.15252/msb.20177551>. <http://msb.embopress.org/content/13/4/924.full.pdf>
56. Jiang YG, Wu Z, Wang J, Xue X, Chang SF (2018) Exploiting feature and class relationships in video categorization with regularized deep neural networks. *IEEE Trans Pattern Anal Mach Intell* 40(2):352–364. <https://doi.org/10.1109/TPAMI.2017.2670560>
57. Dhanaraj K, Kannan R (2018) Capitalizing the collective knowledge for video annotation refinement using dynamic weighted voting. *International Journal for Research in Science Engineering & Technology*, p 4
58. Kaspar A, Patterson G, Kim C, Aksoy Y, Matusik W, Elgharib MA (2018) Crowd-guided ensembles: how can we choreograph crowd workers for video segmentation? In: *CHI'18 Proceedings of the 2018 CHI conference on human factors in computing systems*, Paper No. 111. Montreal QC, Canada. April 21–26, 2018 ACM New York, NY, USA ©2018 table of contents ISBN: 978-1-4503-5620-6. <https://doi.org/10.1145/3173574.3173685>
59. Wang K, Guo Z, Sha W, Glicksman M, Rajan K (2005) Property predictions using microstructural modeling. *Acta Materialia* 53(12):3395–3402
60. Fullwood DT, Niezgodna SR, Adams BL, Kalidindi SR (2010) Microstructure sensitive design for performance optimization. *Prog Mater Sci* 55(6):477–562
61. Niezgodna SR, Kanjarla AK, Kalidindi SR (2013) Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data. *Integr Mater Manuf Innov* 2(1):3
62. Liu R, Kumar A, Chen Z, Agrawal A, Sundararaghavan V, Choudhary A (2015) A predictive machine learning approach for microstructure optimization and materials design. *Scientific Reports* 5:11,551
63. Xu H, Liu R, Choudhary A, Chen W (2015) A machine learning-based design representation method for designing heterogeneous microstructures. *J Mech Des* 137(5):051,403
64. Kalidindi SR, Brough DB, Li S, Cecen A, Blekh AL, Congo FYP, Campbell C (2016) Role of materials data science and informatics in accelerated materials innovation. *MRS Bull* 41(8):596–602
65. Rajan K (2015) Materials informatics: the materials gene and big data. *Annu Rev Mater Res* 45:153–169
66. Reddy N, Krishnaiah J, Young HB, Lee JS (2015) Design of medium carbon steels by computational intelligence techniques. *Comput Mater Sci* 101:120–126
67. Oliynyk AO, Adutwum LA, Harynyk JJ, Mar A (2016) Classifying crystal structures of binary compounds AB through cluster resolution feature selection and support vector machine analysis. *Chem Mater* 28(18):6672–6681
68. Seshadri R, Sparks TD (2016) Perspective: interactive material property databases through aggregation of literature data. *APL Mater* 4(5):053,206
69. Sparks TD, Gaultois MW, Oliynyk A, Brgoch J, Meredig B (2016) Data mining our way to the next generation of thermoelectrics. *Scr Mater* 111:10–15
70. Oliynyk AO, Antono E, Sparks TD, Ghadbeigi L, Gaultois MW, Meredig B, Mar A (2016) High-throughput machine-learning-driven synthesis of full-Heusler compounds. *Chem Mater* 28(20):7324–7331
71. Citrine Informatics Inc (2017) Citrine Informatics inc. <https://citrine.io/>. Accessed 28 June 2018
72. Brough DB, Wheeler D, Kalidindi SR (2017) Materials knowledge systems in python a data science framework for accelerated development of hierarchical materials. *Integr Mater Manuf Innov* 6(1):36–53
73. Materials Resources LLC (2017) Materials resources LLC. <http://www.icmrl.net/>. Accessed 28 June 2018
74. BlueQuartz Software LLC (2017) Bluequartz Software, LLC. <http://www.bluequartz.net/>. Accessed 28 June 2018
75. Liu X, Furrer D, Kusters J, Holmes J (2018) Vision 2040: a roadmap for integrated, multiscale modeling and simulation of materials and systems. Tech rep NASA. <https://ntrs.nasa.gov/search.jsp?R=20180002010>. Accessed 28 June 2018
76. Ghaboussi J, Garrett J Jr, Wu X (1991) Knowledge-based modeling of material behavior with neural networks. *J Eng Mech* 117(1):132–153
77. Haj-Ali R, Pecknold DA, Ghaboussi J, Voyiadjis GZ (2001) Simulated micromechanical models using artificial neural networks. *J Eng Mech* 127(7):730–738
78. Lefik M, Boso D, Schrefler B (2009) Artificial neural networks in numerical modelling of composites. *Comput Methods Appl Mech Eng* 198(21–26):1785–1804
79. Schooling J, Brown M, Reed P (1999) An example of the use of neural computing techniques in materials science the modelling of fatigue thresholds in Ni-base superalloys. *Mat Sci Eng: A* 260(1–2):222–239

80. Yun GJ, Ghaboussi J, Elnashai AS (2008a) A new neural network-based model for hysteretic behavior of materials. *Int J Numer Methods Eng* 73(4):447–469
81. Yun GJ, Ghaboussi J, Elnashai AS (2008b) Self-learning simulation method for inverse nonlinear modeling of cyclic behavior of connections. *Comput Methods Appl Mech Eng* 197(33–40):2836–2857
82. Oishi A, Yamada K, Yoshimura S, Yagawa G (1995) Quantitative nondestructive evaluation with ultrasonic method using neural networks and computational mechanics. *Comput Mech* 15(6):521–533
83. Ghaboussi J, Joghataie A (1995) Active control of structures using neural networks. *J Eng Mech* 121(4):555–567
84. Bani-Hani K, Ghaboussi J (1998) Nonlinear structural control using neural networks. *J Eng Mech* 124(3):319–327
85. Pei JS, Smyth AW (2006) New approach to designing multilayer feedforward neural network architecture for modeling nonlinear restoring forces. II: applications. *J Eng Mech* 132(12):1301–1312
86. Singh V, Willcox KE (2017) Methodology for path planning with dynamic data-driven flight capability estimation. *AIAA J*, pp 1–12
87. Yun GJ, Saleeb A, Shang S, Binienda W, Menzemer C (2011) Improved selfsim for inverse extraction of nonuniform, nonlinear, and inelastic material behavior under cyclic loadings. *J Aerosp Eng* 25(2):256–272
88. Yun GJ (2017) Integration of experiments and simulations to build material big-data. In: *Proceedings of the 4th world congress on integrated computational materials engineering (ICME 2017)*. Springer, pp 123–130
89. Farrokh M, Dizaji MS, Joghataie A (2015) Modeling hysteretic deteriorating behavior using generalized Prandtl neural network. *J Eng Mech* 141(8):04015,024
90. Wang B, Zhao W, Du Y, Zhang G, Yang Y (2016) Prediction of fatigue stress concentration factor using extreme learning machine. *Comput Mater Sci* 125:136–145
91. Versino D, Tonda A, Bronkhorst CA (2017) Data driven modeling of plastic deformation. *Comput Methods Appl Mech Eng* 318:981–1004
92. Nutonian (2017) Nutonian. <https://www.nutonian.com/products/eureka/>. Accessed 28 June 2018
93. Bessa M, Bostanabad R, Liu Z, Hu A, Apley DW, Brinson C, Chen W, Liu WK (2017) A framework for data-driven analysis of materials under uncertainty: countering the curse of dimensionality. *Comput Methods Appl Mech Eng* 320:633–667
94. DeCost BL, Francis T, Holm EA (2017) Exploring the microstructure manifold: image texture representations applied to ultrahigh carbon steel microstructures. *Acta Mater* 133:30–40
95. DeCost BL, Holm EA (2015) A computer vision approach for automated analysis and classification of microstructural image data. *Comput Mater Sci* 110:126–133
96. Adachi Y, Taguchi M, Hirokawa S (2016) Microstructure recognition by deep learning. *Tetsu To Hangne-Journal of the Iron and Steel Institute of Japan* 102(12):722–729. <https://doi.org/10.2355/tetsutohagane.TETSU-2016-035>
97. Adachi Y, Taguchi S, Kohkawa S (2016) Microstructure recognition by deep learning. *Tetsu-to-Hagane* 102(12):722–729
98. Chowdhury A, Kautz E, Yener B, Lewis D (2016) Image driven machine learning methods for microstructure recognition. *Comput Mater Sci* 123:176–187
99. DeCost BL, Hecht MD, Francis T, Picard YN, Webler BA, Holm EA (2017) UHCSDB: UltraHigh carbon steel micrograph database. *Integr Mater Manuf Innov* 6:97–205
100. DeCost BL, Holm EA (2017) Characterizing powder materials using keypoint-based computer vision methods. *Comput Mater Sci* 126:438–445
101. DeCost BL, Jain H, Rollett AD, Holm EA (2017) Computer vision and machine learning for autonomous characterization of am powder feedstocks. *JOM* 69(3):456–465
102. Van Valen DA, Kudo T, Lane KM, Macklin DN, Quach NT, DeFelice MM, Maayan I, Tanouchi Y, Ashley EA, Covert MW (2016) Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS Comput Biol* 12(11):e1005177
103. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H (2017) Brain tumor segmentation with deep neural networks. *Med Image Anal* 35:18–31
104. Zoph B, Vasudevan V, Shlens J, Le QV (2017) Learning transferable architectures for scalable image recognition. [arXiv:170707012](https://arxiv.org/abs/1707.07012)
105. citrination (2017) Citrination. <http://help.citrination.com/knowledgebase/articles/1804297-citrine-deep-learning-micrograph-color>. Accessed 28 June 2018
106. kitware (2018) Kitware. <https://www.kitware.com/platforms/>. Accessed 28 June 2018
107. DeCost BL, Francis T, Holm EA (2018) High throughput quantitative metallography for complex microstructures using deep learning: a case study in ultrahigh carbon steel. [arXiv:1805.08693](https://arxiv.org/abs/1805.08693)
108. Jain A, Persson KA, Ceder G (2016) Research upyear: the materials genome initiative: data sharing and the impact of collaborative AB initio databases. *APL Mater* 4(5):053,102
109. Wodo O, Broderick S, Rajan K (2016) Microstructural informatics for accelerating the discovery of processing–microstructure–property relationships. *MRS Bull* 41(8):603–609
110. Agrawal A, Choudhary A (2016) Perspective: materials informatics and big data: realization of the fourth paradigm of science in materials science. *APL Mater* 4(5):053,208
111. Mulholland GJ, Paradiso SP (2016) Perspective: materials informatics across the product lifecycle: selection, manufacturing, and certification. *APL Mater* 4(5):053,207
112. McGinn PJ (2015) Combinatorial electrochemistry–processing and characterization for materials discovery. *Mater Discov* 1:38–53
113. Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C (2017) Machine learning in materials informatics: recent applications and prospects. *npj Comput Mater* 3(1):54
114. Ryan K, Lengyel J, Shatruk M (2018) Crystal structure prediction via deep learning. *J Am Chem Soc* 140(32):10158–10168. <https://doi.org/10.1021/jacs.8b03913>
115. Graser J, Kauwe SK, Sparks TD (2018) Machine learning and energy minimization approaches for crystal structure predictions: a review and new horizons. *Chem Mater* 30(11):3601–3612
116. Oliynyk AO, Mar A (2017) Discovery of intermetallic compounds from traditional to machine-learning approaches. *Acc Chem Res* 51(1):59–68
117. Furmanchuk A, Saal JE, Doak JW, Olson GB, Choudhary A, Agrawal A (2018) Prediction of seebeck coefficient for compounds without restriction to fixed stoichiometry: a machine learning approach. *J Comput Chem* 39(4):191–202
118. Curtarolo S, Hart GL, Nardelli MB, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. *Nat Mater* 12(3):191
119. De Jong M, Chen W, Angsten T, Jain A, Notestine R, Gamst A, Sluiter M, Ande CK, Van Der Zwaag S, Plata JJ et al (2015) Charting the complete elastic properties of inorganic crystalline compounds. *Scientific Data* 2:150,009
120. Oliynyk AO, Adutwum LA, Rudyk BW, Pisavadia H, Lotfi S, Hlukhyy V, Harynyk JJ, Mar A, Brgoch J (2017) Disentangling structural confusion through machine learning: structure

- prediction and polymorphism of equiatomic ternary phases abc. *J Am Chem Soc* 139(49):17,870–17,881
121. Oliynyk AO, Gaultois MW, Hermus M, Morris AJ, Mar A, Brgoch J (2018) Searching for missing binary equiatomic phases: complex crystal chemistry in the Hf[–]In system. *Inorg Chem* 57(13):7966–7974. <https://doi.org/10.1021/acs.inorgchem.8b01122>
 122. Ash J, Fourches D (2017) Characterizing the chemical space of ERK2 kinase inhibitors using descriptors computed from molecular dynamics trajectories. *J Chem Inf Model* 57(6):1286–1299
 123. Mannodi-Kanakthodi A, Pilania G, Ramprasad R (2016) Critical assessment of regression-based machine learning methods for polymer dielectrics. *Comput Mater Sci* 125:123–135
 124. Bereau T, Andrienko D, Kremer K (2016) Research upyear: computational materials discovery in soft matter. *APL Materials* 4(5):053,101
 125. Zhuo Y, Mansouri Tehrani A, Brgoch J (2018) Predicting the band gaps of inorganic solids by machine learning. *J Phys Chem Lett* 9(7):1668–1673
 126. Gaultois MW, Oliynyk AO, Mar A, Sparks TD, Mulholland GJ, Meredig B (2016) Perspective: web-based machine learning models for real-time screening of thermoelectric materials properties. *APL Mater* 4(5):053,213
 127. Kauwe SK, Graser J, Vazquez A, Sparks TD (2018) Machine learning prediction of heat capacity for solid inorganics. *Integrating Materials and Manufacturing Innovation*, pp 1–9
 128. Xie T, Grossman JC (2018) Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys Rev Lett* 120(14):145,301
 129. Baumes LA, Collet P (2009) Examination of genetic programming paradigm for high-throughput experimentation and heterogeneous catalysis. *Comput Mater Sci* 45(1):27–40
 130. Kim C, Pilania G, Ramprasad R (2016) From organized high-throughput data to phenomenological theory using machine learning: the example of dielectric breakdown. *Chem Mater* 28(5):1304–1311
 131. Bicevska Z, Neimanis A, Oditis I (2016) NoSQL-based data warehouse solutions: sense, benefits and prerequisites. *Baltic J Mod Comput* 4(3):597
 132. Gagliardi D (2015) Material data matterstandard data format for engineering materials. *Technol Forecast Soc Chang* 101:357–365
 133. Takahashi K, Tanaka Y (2016) Materials informatics: a journey towards material design and synthesis. *Dalton Trans* 45(26):10,497–10,499
 134. Blaiszik B, Chard K, Pruyne J, Ananthkrishnan R, Tuecke S, Foster I (2016) The materials data facility: data services to advance materials science research. *JOM* 68(8):2045–2052
 135. O'Mara J, Meredig B, Michel K (2016) Materials data infrastructure: a case study of the citrination platform to examine data import, storage, and access. *JOM* 68(8):2031–2034
 136. Puchala B, Tarcea G, Marquis EA, Hedstrom M, Jagadish H, Allison JE (2016) The materials commons: a collaboration platform and information repository for the global materials community. *JOM* 68(8):2035–2044
 137. Jacobsen MD, Fourman JR, Porter KM, Wirrig EA, Benedict MD, Foster BJ, Ward CH (2016) Creating an integrated collaborative environment for materials research. *Integr Mater Manuf Innov* 5(1):12
 138. The Minerals Metals & Materials Society (TMS) (2017) Building a materials data infrastructure: opening new pathways to discovery and innovation in science and engineering. TMS, Pittsburgh, PA. https://doi.org/10.7449/mdistudy_1
 139. AFRL (2018) Air Force Research Laboratory (AFRL) additive manufacturing (AM) modeling challenge series. <https://materials-data-facility.github.io/MID3AS-AM-Challenge/>. Accessed 28 June 2018
 140. Kaggle (2018) Kaggle competitions. <https://www.kaggle.com/competitions>. Accessed 28 June 2018
 141. Rose F, Toher C, Gossett E, Oses C, Nardelli MB, Fornari M, Curtarolo S (2017) Aflux: the lux materials search API for the afflow data repositories. *Comput Mater Sci* 137:362–370
 142. Balachandran PV, Young J, Lookman T, Rondinelli JM (2017) Learning from data to design functional materials without inversion symmetry. *Nat Commun* 8:14,282
 143. Jain A, Hautier G, Ong SP, Persson K (2016) New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships. *J Mater Res* 31(8):977–994
 144. Ling J, Hutchinson M, Antono E, Paradiso S, Meredig B (2017) High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates. *Integr Mater Manuf Innov* 6(3):207–217
 145. Godaliyadda G, Ye DH, Uchic MD, Groeber MA, Buzzard GT, Bouman CA (2016) A supervised learning approach for dynamic sampling. *Electron Imag* 2016(19):1–8
 146. Godaliyadda GDP, Ye DH, Uchic MD, Groeber MA, Buzzard GT, Bouman CA (2018) A framework for dynamic image sampling based on supervised learning. *IEEE Trans Comput Imag* 4(1):1–16
 147. Bjerrum EJ (2017) Molecular generation with recurrent neural networks. arXiv:170504612
 148. Segler MH, Kogej T, Tyrchan C, Waller MP (2018) Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent Sci* 4(1):120–131. <https://doi.org/10.1021/acscentsci.7b00512>
 149. Wei JN, Duvenaud D, Aspuru-Guzik A (2016) Neural networks for the prediction of organic chemistry reactions. *ACS Central Sci* 2(10):725–732
 150. Goh GB, Hodas NO, Vishnu A (2017) Deep learning for computational chemistry. *J Comput Chem* 38(16):1291–1307. <https://doi.org/10.1002/jcc.24764>
 151. Segler MH, Preuss M, Waller MP (2017) Learning to plan chemical syntheses. arXiv:170804202
 152. Liu R, Ward L, Wolverton C, Agrawal A, Liao W, Choudhary A (2016) Deep learning for chemical compound stability prediction. In: Proceedings of ACM SIGKDD workshop on large-scale deep learning for data mining (DL-KDD), pp 1–7
 153. Wu H, Lorensen A, Anderson B, Witteman L, Wu H, Meredig B, Morgan D (2017) Robust FCC solute diffusion predictions from ab-initio machine learning methods. *Comput Mater Sci* 134:160–165
 154. Carter EA (2008) Challenges in modeling materials properties without experimental input. *Science* 321(5890):800–803
 155. Carleo G, Troyer M (2017) Solving the quantum many-body problem with artificial neural networks. *Science* 355(6325):602–606
 156. Hush MR (2017) Machine learning for quantum physics. *Science* 355(6325):580–580
 157. Faber FA, Hutchison L, Huang B, Gilmer J, Schoenholz SS, Dahl GE, Vinyals O, Kearnes S, Riley PF, Von Lilienfeld OA (2017) Fast machine learning models of electronic and energetic properties consistently reach approximation errors better than DFT accuracy. arXiv:170205532
 158. Schütt K, Kindermans PJ, Felix HES, Chmiela S, Tkatchenko A, Müller KR (2017) SchNet: a continuous-filter convolutional neural network for modeling quantum interactions. In: Advances in neural information processing systems, pp 992–1002

159. Brockherde F, Vogt L, Li L, Tuckerman ME, Burke K, Müller KR (2017) Bypassing the Kohn-Sham equations with machine learning. *Nat Commun* 8(1):872
160. Kolb B, Lentz LC, Kolpak AM (2017) Discovering charge density functionals and structure-property relationships with prophet: a general framework for coupling machine learning and first-principles methods. *Scientific Reports* 7(1):1192
161. Geers M, Yvonnet J (2016) Multiscale modeling of microstructure–property relations. *MRS Bull* 41(8):610–616
162. McDowell DL, LeSar RA (2016) The need for microstructure informatics in process–structure–property relations. *MRS Bull* 41(8):587–593
163. Behler J (2016) Perspective: machine learning potentials for atomistic simulations. *J Chem Phys* 145(17):170,901
164. Dragoni D, Daff TD, Csányi G, Marzari N (2018) Achieving DFT accuracy with a machine-learning interatomic potential: thermomechanics and defects in bcc ferromagnetic iron. *Phys Rev Mater* 2(1):013,808
165. González D, Badías A, Alfaro I, Chinesta F, Cueto E (2017) Model order reduction for real-time data assimilation through extended kalman filters. *Comput Methods Appl Mech Eng* 326:679–693
166. Matouš K, Geers MG, Kouznetsova VG, Gillman A (2017) A review of predictive nonlinear theories for multiscale modeling of heterogeneous materials. *J Comput Phys* 330:192–220
167. Oishi A, Yagawa G (2017) Computational mechanics enhanced by deep learning. *Comput Methods Appl Mech Eng* 327:327–351
168. Kani JN, Elsheikh AH (2017) Dr-rnn: a deep residual recurrent neural network for model reduction. arXiv:170900939
169. Acharya A (2017) Personal communication, 19 July 2017
170. Bertin M, Du C, Hoefnagels JP, Hild F (2016) Crystal plasticity parameter identification with 3D measurements and integrated digital image correlation. *Acta Mater* 116:321–331
171. Chen Z, Daly S (2017) Active slip system identification in polycrystalline metals by digital image correlation (dic). *Exp Mech* 57(1):115–127
172. Papanikolaou S, Tzimas M, Song H, Reid AC, Langer SA (2017) Learning crystal plasticity using digital image correlation: examples from discrete dislocation dynamics. arXiv:170908225
173. Papanikolaou S (2017) Personal communication, 9 November 2017
174. Shakirov V (2016) Review of state-of-the-arts in artificial intelligence with application to ai safety problem. arXiv:160504232
175. Grace K, Salvatier J, Dafoe A, Zhang B, Evans O (2017) When will AI exceed human performance? Evidence from ai experts. arXiv:170508807
176. Doshi-Velez F, Kim B (2017) Towards a rigorous science of interpretable machine learning. arXiv:170208608
177. Lipton ZC (2016) The mythos of model interpretability. arXiv:160603490
178. Voosen P (2017) The AI detectives. *Science* 357(6346):22–27. <https://doi.org/10.1126/science.357.6346.22>
179. Bornstein AM (2016) Is artificial intelligence permanently inscrutable? *Nautilus*, issue 040, chapter one. <http://nautil.us/issue/40/Learning/is-artificial-intelligence-permanently-inscrutable>
180. Islam R, Henderson P, Gomrokchi M, Precup D (2017) Reproducibility of benchmarked deep reinforcement learning tasks for continuous control. arXiv:1708.04133
181. Henderson P, Islam R, Bachman P, Pineau J, Precup D, Meger D (2017) Deep reinforcement learning that matters. arXiv:170906560
182. Warden P (2018) The machine learning reproducibility crisis. <https://petewarden.com/2018/03/19/the-machine-learning-reproducibility-crisis/>. Accessed 30 Jul 2018
183. Ghanta (2018) A systems perspective to reproducibility in production machine learning domain, 2nd Reproducibility in Machine Learning Workshop at ICML 2018
184. Grazzini (2018) “Show me your code, and then I will trust your figures”: towards software-agnostic open algorithms in statistical production, 2018 European Conference on Quality in Official Statistics
185. Khetarpal (2018) RE-EVALUATE: reproducibility in evaluating reinforcement learning algorithms, ICML 2018 RML
186. Schelter (2018) Declarative metadata management: a missing piece in end-to-end machine learning, SysML 2018
187. Bau D, Zhou B, Khosla A, Oliva A, Torralba A (2017) Network dissection: quantifying interpretability of deep visual representations. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 3319–3327
188. Benhenda M (2017) Chemgan challenge for drug discovery: can ai reproduce natural chemical diversity? arXiv:170808227
189. Mangal A, Holm EA (2018) A comparative study of feature selection methods for stress hotspot classification in materials. *Integr Mater Manuf Innov* 7(3):19. <https://doi.org/10.1007/s40192-018-0109-8>
190. Singh S, De Graef M (2017) Dictionary indexing of electron channeling patterns. *Microsc Microanal* 23(1):110. <https://doi.org/10.1017/S1431927616012769>
191. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
192. Collobert R, Kavukcuoglu K, Farabet C (2011) Torch7: a matlab-like environment for machine learning. In: BigLearn, NIPS workshop, EPFL-CONF-192376
193. Tokui S, Oono K, Hido S, Clayton J (2015) Chainer: a next-generation open source framework for deep learning. In: Proceedings of workshop on machine learning systems (LearningSys) in the twenty-ninth annual conference on neural information processing systems (NIPS), vol 5
194. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M et al (2016) Tensorflow: a system for large-scale machine learning. In: OSDI, vol 16, pp 265–283