Integrative & Comparative Biology, 2021, pp. 1–8

doi: DOI HERE

Advance Access Publication Date: Day Month Year Paper

PAPER

Artificial Intelligence for Biology

Soha Hassoun[1], Felicia Jefferson[2], Xinghua Shi[3], Brian Stucky[4], Jin Wang[5] and Epaminondas Rosa Jr[6]*

[1]Department of Computer Science, Tufts University, Medford, MA 02155, USA, [2]Biology Academic Department, Fort Valley State University, Fort Valley, GA 31030, USA, [3]Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122, USA, [4]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA, [5]Department of Mathematics, University of Tennessee at Chattanooga, Chattanooga, TN 37403, USA and [6]Department of Physics and School of Biological Sciences, Illinois State University, Normal, IL 61790, USA

*Corresponding author: erosa@ilstu.edu

FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

Despite efforts to integrate research across different subdisciplines of biology, the scale of integration remains limited. We hypothesize that future generations of Artificial Intelligence (AI) technologies specifically adapted for biological sciences will help enable the reintegration of biology. AI technologies will allow us not only to collect, connect and analyze data at unprecedented scales, but also to build comprehensive predictive models that span various subdisciplines. They will make possible both targeted (testing specific hypotheses) and untargeted discoveries. AI for biology will be the cross-cutting technology that will enhance our ability to do biological research at every scale. We expect AI to revolutionize biology in the 21st century much like statistics transformed biology in the 20th century. The difficulties, however, are many, including data curation and assembly, development of new science in the form of theories that connect the subdisciplines, and new predictive and interpretable AI models that are more suited to biology than existing machine learning and AI techniques. Development efforts will require strong collaborations between biological and computational scientists. This white paper provides a vision for AI for Biology and highlights some challenges.

Key words: biology, reintegrate, artificial intelligence

1. Introduction

Artificial intelligence as an idea is old. It can be dated back to ancient times around 700 B.C. in Greek mythology, for example, with the giant Talus made of bronze and created, not born, to protect Europa, the mother of King Minos in Crete [Mayor, 2018]. From then to more modern and scientific times, the main restriction to produce machines capable of thinking has been technology, as recognized by Alan M. Turing [Turing, 1936] who, ahead of his time, was asking questions about machines, behavior, consciousness and using discrete processes to mimic nervous systems that operate continuously. John von Neuman [von Neuman, 1958], also ahead of his time, proposed in 1945 a computer architecture in which both the program instructions and the data are located in random-access memory. This design was the precursor of the modern computer, but it was not until the advent of the fast microchip that AI became a practical reality. Since its early beginnings in 1956 [McCarthy *et al.*, 2006, Kaplan and Haelein, 2019] as a field of research and development, AI has evolved and suffered setbacks, until early in the 21st century when it finally flourished with successful applications in academia and industry. A combination of new methods and availability of powerful computers along with vast collections of data brought large investment and widespread interest in AI.

In biology, AI has evolved from the symbolic approach where complex rules are coded in computer language to enable machines to execute coordinated sequences of operations. A typical example of symbolic AI is the game of chess, with relatively simple rules, but a wide range of possible outcomes after the move of one piece. In this case, the rules are set, and a computer can be programmed to analyze all the possibilities before the next move, and then choose the option that produces the most beneficial outcome. A well-known example of successful symbolic AI is IBM's Deep Blue computer that in 1997 beat the then world chess champion Garry Kasparov. Before Deep Blue, computers were not capable of performing computations fast enough to outpace a well-trained human brain. Just as a reference, a smart phone today has a computational speed comparable to that of the Deep Blue.

Powerful as it is, symbolic AI is limited to systems that operate by well-defined sets of rules [Haugeland, 1985], which is not necessarily the case in the realm of living systems. Additionally, there is not much resemblance between symbolic AI and biological intelligence as far as their operation and functioning are concerned. Symbolic AI can only make choices based on an *a priori* established set of rules. Biological intelligence, however, can learn on the fly and make decisions based on information acquired by experience and by seeing objects, for example.

A similar feature was introduced with Artificial Neural Networks (ANNs) and Machine Learning (ML), inspired by the networked neurons of the biological brain. For instance, the mechanism behind memory in the biological brain is known to be related to the strength of the connections, or synapses, between neurons [Hebb, 1949]. The remarkable Hopfield network model with associative memory [Hopfield, 1982] has provided essential insights into neuronal computation. In the Hopfield model, each node is assigned a binary unit, and the strengths of the connections between nodes are quantified in terms of weights, and it has been successfully implemented in a number of applications, including the enhancement of the network capability for coding and information retrieval [Follmann *et al.*, 2014].

Another branch in AI is the Hidden Markov Model (HMM), applicable to stochastic processes occurring in systems with behaviors displaying no recurrence of fixed patterns. It has been implemented for example in unequal and unknown evolution rates at different sites in molecular sequences where the HMM allows for rates to differ between sites and for correlations between the rates of neighboring sites [Felsenstein and Churchill, 1996].

A noticeable step up in AI is Machine Learning (ML), where the computer is given samples of data with different but related patterns in connection with a topic of interest. The computer then learns about those patterns by searching features that will distinguish between diverse categories of patterns or tries to identify features that are common among the various categories. After this learning phase the computer's task is to classify a given new pattern that it is presented with, or to predict a future behavior of system being studied [Rawlings and Fox, 1994, Follmann and Rosa Jr, 2019]. The network used in ML has been extended in Reservoir Computing to include layers of connections which makes the process more efficient.

Major recent advances in AI are due to Deep Learning (DL), consisting of multiple processing layers in artificial neuronal networks aimed at pattern recognition and modeling complex relationships between input and output. In addition, DL has enhanced the potential for using computer-assisted discovery in prediction of protein structure, molecular design and macromolecular target identification for drug discovery [J. Gimenez-Luna *et al.*, 1985].

2. The need for AI to reintegrate biology

Concern about the fragmentation of biology into specialized subdisciplines, and calls for its reintegration, have been appearing in the scientific literature for years [Sukumaran and Knowles, 2018, Noble, 2013, Hayes, 2005, Drew and Henne, 2006]. So far, though, a grand reunification has remained elusive. Human intellectual limits in collecting data, integrating data, and testing hypotheses spanning multiple subdisciplines are the primary reason biology became fragmented in the first place. Reintegration will be impossible without overcoming these limitations. Stated differently, key biological systems and related information, at all levels of biological organization, are simply too complex for humans to understand with sufficient depth to elicit generalized, human-driven reintegration. Here, we make the case that advances in AI methods and technologies will provide our best hope for overcoming the human cognitive limitations that have splintered biology into ever-more-specialized subdisciplines.

Our vision for reintegrating biology recognizes the enormous potential of existing AI techniques to accelerate biological research. Current AI and ML methods are already having an impact in biology (discussed in more detail below), but there is room for improvement on the existing methods and techniques for data integration. While technological advances have made great strides in hardware for processing speed, inadequate input/output performance in the case of large amounts of data may result in severe limitations on the overall process [Isakov, 2020].

We envision new suites of AI tools, developed for biological inquiry and perhaps even inspired by biological systems [Yanguas-Gil *et al.,* 2019, Drumond *et al*., 2019, Chance *et al*., 2020, Follmann and Rosa Jr, 2019], powering biological investigation at unprecedented scales.

3. What is the potential impact?

The development of statistics and electronic computers transformed 20th-century biology, and we foresee AI

having a similarly transformative impact on 21st-century biology [Yu and Kumbier, 2018]. AI-driven reintegration of biological disciplines will establish a new kind of biology that will allow us to answer deep biological questions in ways that are impossible today. Such questions will cut across biological subdisciplines and integrate across the scales of biological inquiry (spatial, temporal, and organizational). We offer some examples as illustrations, arranged in approximate order of increasing difficulty of implementation.

Example 1: Biological knowledge discovery and assembly

Surely all research biologists have at some point spent countless hours searching for relevant literature and sifting through various data sources to assemble information relevant to a particular research question. As the volume of published literature and data continues to grow at a nearly exponential rate, this process becomes increasingly difficult and frustrating. In fact, for human researchers, comprehensive collection, assembly, integration, and analysis of published literature and data at even modest scales is nearly impossible today. We predict that AI-driven data generation and integration across the spectrum of data modalities and sources will eventually largely solve this problem. AI will utilize a variety of known and new techniques to collect and assemble these data: text mining [Cohen and Hunter, 2008], semantic analysis [Berners-Lee *et al*., 2001], and missing link prediction [Ahmad *et al*., 2020] in existing multilevel and hierarchical knowledge graphs. Simply put, we need a next-generation search engine capable of unearthing known and predicted biological knowledge. Ultimately, we envision a system that can aid biological research by retrieving all known information relevant to a particular question, organized and visualized in a coherent and potentially customizable way, while also highlighting missing information. We do not anticipate AI to perform biological research totally independent from human supervision and control. However, there is potential for AI to become a powerful and necessary tool for information discovery.

Example 2: Behavioral ecology

Suppose that, for some species of bird, we would like to understand the relationship between individual fitness and environment, including the birds' social environment [Hawkins and DuRant, 2020]. Ideally, this task would draw upon data from a wide range of biological and spatial scales (*e.g*., vocalizations and communication, social networks, movement, morphometrics, parasite loads, genetics, biomarkers, etc.) and sources (*e.g.*, images, videos, audio recordings, tracking tags, DNA sequencers, etc.). Currently, such analysis is usually done using one or a few data modalities with relatively small numbers of individuals (*e.g.*, using radio-frequency identification (RFID) tags to collect movements and social network analysis to understand social behaviors of birds). We hypothesize that simultaneous advances in AI and automated data collection will make it possible to answer these questions using a holistic approach that goes far beyond current capabilities, which will allow us to answer ever more complicated biological questions; for example: How does genetics affect social behaviors that in turn affect collective behaviors like migration [Sukumaran *et al*., 2016]? Another example would be the integration of AI in hierarchical decision-making models of behavior extended to the foraging of large herbivores [Saarenmaa, 1988].

Example 3: Genes to phenotypes

Predicting an organism's phenotype is extraordinarily difficult because it requires integrating processes and information across multiple scales of biological organization, from molecules to an organism's environment [Burnett *et al*., 2020]. The general solutions to this problem are beyond the grasp of today's AI technologies, but future advances in machine reasoning, learning, and causal inference, combined with continual growth in data, collection, and computational capacity, will help transform our understanding of how phenotypes emerge. Specifically, these technologies will allow us to use heterogeneous data (e.g., DNA sequence data, phylogenetic information, environmental data) and knowledge (e.g., gene function, results of prior experiments) to elucidate and test hypotheses about the inputs that shape phenotypes. For instance, we could investigate how data collected over diverse labs and fields (*e.g*., imaging of cells, genomics, epigenomics, proteomics, metabolomics, metagenomics in soils) can predict the cellular decision making or phenotypic changes that affect productivity of crops like corn.

Example 4: Prediction, evolution and control of infectious diseases

Infectious diseases are caused by pathogenic microorganisms, and their spread may be based on direct (*i.e.*, human-to-human) and/or indirect (such as environment-to-human and vector-to-human) transmission routes. Infectious diseases can be deadly, very contagious, and display incubation periods of days or weeks with no visible symptoms. Add to this equation the lack of knowledge or means to detect and treat novel diseases, and we have a problem that can be as big as the situation we are living today with the COVID-19 pandemic. While traditional mathematical and statistical models are capable of making predictions, albeit limited, developing strategies for disease control may require more elaborate approaches for making well informed decisions. A number of recent studies have already started applying AI and ML methods to the investigation of COVID-19 [Lalmuanawma *et al*., 2020, Abd-Alrazaq *et al*., 2020].

COVID-19 in particular, as a current dramatic example, not only has led to unprecedented cases and deaths, but also exhibited a high level of unpredictability from the classical modeling point of view. Most (if not all) of the traditional epidemic models based on early COVID-19 data have failed to correctly predict the pandemic progression, often by an order of magnitude [Kuhl, 2020]. These traditional modeling and computing techniques do not possess the capability to react or adapt when an unexpected situation is encountered, and they generally have difficulty in handling heterogeneous sources of data. In contrast, AI could enable machines to better act or react to evolving and heterogeneous pandemic data [Wiemken and Kelly, 2020, Agrebi and Larbi, 2020]. With the fast improvement of computational power and wide availability of demographic, epidemic and human mobility data, the application of AI to infectious diseases, particularly COVID-19, has become increasingly popular and practically indispensable. Furthermore, AI and machine learning methods can be integrated with classical mechanistic models to infer critical disease parameters in real time from reported case data, which could lead to more accurate forecasts of the pandemic progression and, consequently, more effective policy making. Given all these new developments, we believe that AI has become a vital tool in epidemiology where potential breakthroughs will soon take place with the application of AI and its integration with other cutting-edge computational, mathematical and statistical approaches. However, we also note that many recently published applications of AI techniques to COVID-19 are of limited use due to methodological flaws or bias issues [Roberts *et al*. 2020. Nevertheless, facing a sea of data in the digital age, it is imperative that we leverage the power of AI to deepen our understanding of infectious diseases, to improve our practice in the control and management of disease outbreaks, and to help promote public health. This is especially important for the prevention of and intervention on future pandemics.

Meanwhile, state-of-the-art supercomputing models can give us a glimpse of what to expect from the implementation of AI in epidemiological studies [ALCF]. Given the recent technological advances in capability for data collection, analysis and storage, AI has the potential not only for forecasting the outbreak of new diseases but also for helping in the implementation of methods and techniques for tracking [AlGaradi *et al*., 2016], diagnosis and treatment, leading to effective control and potential termination of a pandemic.

In summary, the new AI-augmented biology we envision will generate tools, methods, and knowledge that will translate to a host of biology-adjacent disciplines, such as bioengineering, biophysics, biochemistry and medicine. In particular, new developments in drug discovery using AI will play a seminal role in disease prevention and treatment [Fleming, 2018, Smith, 2018]. Additionally, we anticipate that new AI tools, in concert with open data, will help democratize participation in biology, allowing researchers at institutions with more limited resources to participate in cutting-edge biological research.

4. Why now?

The time for AI in biology has arrived. There are now sensors, Internet of Things (IoT), and environmental monitors that allow the collection of data at unprecedented scales. Large, heterogeneous datasets at the confluence of multiple information streams are rapidly growing in size. We now have multivariate data across time, space, and biological scales that need to be analyzed in an integrated manner to discover system-wide, multiscale phenomena that can lead us to understand fundamental rules of life and their application to other systems. The AI infrastructure to support these efforts is beginning to emerge. There are now unprecedented computational capabilities in the form of storage, CPU/GPU computing, and large-scale distributed computing which, combined with the increasing availability of software tools for AI, is enabling the rapid exploration and development of novel techniques and applications. These resources continue to grow and will enable the next generation of AI for the most complex problems in biology. However, all these features are not free from challenges which include, for example, still limited computational input/output capability [Meena, 2014, Ben-David, 2016] as well as critical ethical issues [Tonkens, 2009]. Both these topics are further discussed below.

5. State-of-the-art technologies and applications

Although machine learning (ML) has recently entered the popular lexicon and is often conflated with AI in general, AI is a broad field with a long history, and it provides a diverse set of tools and approaches that encompass much more than ML. A variety of these tools have already been used to help solve some biological problems. For example, methods from symbolic AI have been used to develop sophisticated software pipelines for integrating highly heterogeneous sources of information about plant development and to help elucidate possible links between gene function and phenotype [Stucky *et al.*, 2018, Braun and Lawrence-Dill, 2020, Edmunds *et al*., 2015]. Statistical learning, and deep learning [Lamba *et al*., 2019] in particular, have recently found application in the automated analysis of biological imagery at various scales including unmanned aerial vehicle (UAV) and field photographs of plants [Gao, 2020], satellite imagery [Kislov *et al.*, 2020], biomedicine [Tian, 2021] bioacoustic data [Bermant *et al.*, 2019], genomic analyses [Libbrecht and Noble, 2015], and classifying protein

function from amino acid sequences [Nikam and Gromiha, 2019].

6. Barriers

Many important barriers need to be addressed to enable the next generation of AI for biology.

6.1. Data are critical to all aspects of this vision

New technologies need to be developed for the automatic collection of biological data with varied data modalities (*e.g.,* images, videos, molecular profiles) and comprehensive measurements of biological systems at various biological, spatial and temporal scales. Furthermore, data quality is a concern with large, noisy datasets, so data scientists must work with biologists to ensure the data we generate are as useful as possible. Key challenges include identifying outliers and biases, mitigating known biases, understanding variation, and improving signal-to-noise ratios. To enable the open sharing of data, tools should be developed to allow for transparent data sharing, with consideration of provenance, security, privacy, and fairness. Other researchers can use these shared data to form new hypotheses and build new theories. Beyond new technologies for gathering biological data, high-quality reference datasets for benchmarking AI applications in biology will also be critical. For example, over the last decade, the availability of the ImageNet dataset has been a major factor in the development of new AI methods for image processing [Deng, 2009, Russakovsky, 2015]. Similarly, reference datasets for evaluating AI methods across a range of biological applications will be needed to support future innovation in the biological domain.

6.2. Theory

Development of theory from multiple disciplines will enable the development of new AI technologies for biology. For example, theory in biology, chemistry, physics, and social sciences could be utilized to develop more appropriate AI models for understanding biological systems. Mathematical and statistical theory should be developed to not only design new AI methods but also further our understanding of the fundamental principles [Deisenroth *et al*., 2020] underlying current and emerging AI technologies. Novel development and incorporation of evolving and updated theory will be conducted in a feedback loop, with AI data analysis and evaluation leading to the development of improved methods.

6.3. Models

Novel AI models need to be developed that are bio-meaningful, bio-inspired, and bio-integrated at scale [Alber *et al*., 2019]. AI models should incorporate biological hierarchical structures and feedback/loops. Notably, deep learning, which dominates current AI research, arose from biological inspiration. Deep learning systems are based on artificial neural networks, which originated with efforts to mimic the way computation happens in biological brains. Many other biological systems are characterized by highly complex interactions leading to system-level emergent properties and behaviors, and we suspect the mechanisms behind such systems might present opportunities for new approaches to AI. Although black-box models are appropriate for some types of modeling tasks, AI models that are interpretable, explainable, and visualizable should be encouraged. AI models should be robust and resilient, allowing for redundancy and plasticity. AI models should enable unsupervised learning or semi-supervised learning when labeled data are missing, limited or insufficient.

AI models and software should be open-source to allow not only accessibility for all but also for taking advantage of collaborative public efforts that can bring a plethora of perspectives and development contributions. Open availability of scientific data will directly benefit society as a whole by promoting transparency, reproducibility and more efficient use of information. However, challenges exist including limited control over how the data will be used, and lack of recognition and of incentive to the generators of data. These challenges are not simple problems and will take some time to resolve [Molloy, 2011].

6.4. Computing Infrastructure

Current computing storage and throughput will be challenged by the amount and scale of future biological data. Accordingly, storage and performance of computing systems must also scale. Traditional computing models (von Neumann architectures) [von Neumann, 1958] may not be well suited for biological tasks. Emerging technologies such as quantum and neuromorphic computing might provide appropriate alternatives. Focusing AI on biology will open up novel opportunities for developing hardware, software, and new computing mediums that are more appropriate for biological applications. There are also exciting opportunities to explore novel computing-biological interfaces at the intersection of biology and computing.

Whatever new technologies might be realized in the future, it will be critical to ensure that leading-edge computing infrastructure is available to as many researchers as possible, not just researchers fortunate enough to be affiliated with the most well-funded universities, government agencies, and NGOs. As an example, the NSF-funded Extreme Science and Engineering Discovery Environment (XSEDE - https://www.xsede.org) is a virtual organization that provides advanced computing infrastructure to researchers across the United States, including many who might not otherwise have access to high-performance computing resources. Efforts like XSEDE will be crucial in the future to help democratize access to AI-related computing

tools and to facilitate the pooling of resources required for extremely large-scale projects. The cost associated with the development of this infrastructure is expected to be a barrier for its implementation, unless private investors and public sectors can foresee the benefits of the investment.

In the context of the last two subsections, it is imperative for a mechanism to be created to ensure long term maintenance and updating of data storage and coding. This should guarantee reproducibility of results and also that the scientific community as a whole will have easy access to the methods and tools to stay up-to-date with potentially fast-paced developments.

6.5. Ethics

In a wide range of fields, biology included, a growing number of functions are being outsourced to AI with less direct human participation and control. This raises concerns about biases, unfairness and discrimination, and effort must be made to guarantee equitability [Piano, 2020]. Central to this effort is to develop mechanisms that ensure transparency, fairness, access, equity, diversity, shared governance, privacy and security of data at all development stages. There are already well-known cases of biases in ML data and algorithms [Garcia, 2016], which can then be exacerbated as data and models become more complicated. Black box models, for example, restrict shared-decision and make it difficult to effectively implement real-time error-checking [Rudin *et al.*, 2021]. One venue to tackle ethics in AI would be through governance. However, while AI is evolving rather quickly, the governance of AI is in its infancy [Renda, 2019, Taeihagh, 2021]. Ethical issues in AI must be addressed head-on as a first-class concern. Developers and users need to be trained to be aware of these issues, and our workforce must be sufficiently diversified to ensure no one is left behind. Further, we all should be aware of potential misuse of AI to harm humans or the environment and the utmost care must be taken to assess and address these issues.

6.6. Training

Training must be addressed in a more systematic and cross-institutional/disciplinary manner. A new generation of diverse scientists must be trained at the intersection of biology and computer science, starting with undergraduate studies and through graduate and postdoctoral opportunities. In line with much of the recent NSF-funded STEM educational research, training of future AI/ML researchers may need to commence even earlier [Jones et. al, 2020; Paul and Jefferson, 2019]. According to a Brookings Institute Report on the Future of Education in the AI Age, America's early education must reflect a deliberately tuned and calibrated system that proactively emphasizes AI/ET, big data analytics, and super-computing. [J.R. Allen, 2019]. As energy-efficient neural coding is required to control individual neurons and brain circuits, so too is balance and inputting-outputting of AI and ML data. Balance requires distribution and diversity. Users of AI systems must be trained to interpret the results and use the various tools judiciously. Vocational pathways need to reward cross-disciplinary work.

7. Concluding remarks

The rapid growth and consequent fragmentation of biology has created a wealth of subdisciplines that would benefit greatly from being part of an integrated collective rather than remaining individualized. Given the overwhelming complexity of contemporary biological knowledge, placing subdisciplines of biology under a single umbrella has become a task of insurmountable proportions. Nevertheless, certain technological tools available today and still evolving, can help amalgamate different subdisciplines of biology, each realizing the inherent advantages of working in unison with the others. AI is one such a tool. It has the potential for broad and long-lasting impacts on biological science and beyond. AI will equip biologists with powerful tools to ask and solve ambitious questions, such as investigating and integrating complex mechanisms across a wide range of scales (from genes, to cells, to organisms, populations, and ecosystems), and developing theoretical machines to understand biological and ecological systems at extremely large scales, all of which would be severely limited without AI. Meanwhile, feedback from biology will help to re-define AI concepts and improve AI computing. We expect these developments will lead to better integration of biological knowledge and enable exciting new collaborations among researchers across biology and adjacent disciplines, including computer science and engineering. Such interdisciplinary collaborations are critical in promoting the next generation of AI in biology, and in addressing the barriers of data, theory, model development and various other challenges the AI field is currently facing.

Acknowledgements

The initial development of this vision paper took place during the December 2019 NSF-funded Jumpstart Meeting for Reintegrating Biology.
Soha Hassoun was supported by NSF, Award CCF-1909536. Felicia Jefferson was supported by NSF Awards 1939739 and 1900572. Xinghua Shi was supported by NSF Award 1750632. Jin Wang was supported by NSF Award 1951345.

Competing interests

There are NO Competing Interests.

References

A. Abd-Alrazaq, M. Alajlani, D. Alhuwail, J. Schneider, S. AlKuwari, Z. Shah, M. Hamdi, and M. Househ. Artificial intelligence in the fight against covid-19: Scoping review. *Journal of medical Internet research*, 22(12): e20756, 2020.

I. Ahmad, M.U. Akhtar, S. Noor, and A. Shahnaz. Missing link prediction using common neighbor and centrality based parameterized algorithm. *Scientific reports*, 10(1):1–9, 2020.

S. Agrebi and A. Larbi, Use of artificial intelligence in infectious diseases, Artificial Intelligence in Precision Health, 2020: 415-438.

M.A. Al-Garadi, M.S. Khan, K.D. Varathan, G. Mujtaba, and A.M. Al-Kabsi. Using online social networks to track a pandemic: A systematic review. *Journal of biomedical informatics*, 62:1–11, 2016.

M. Alber, A.B. Tepole, W. R. Cannon, S. De, S. Dura-Bernal, K. Garikipati, G. Karniadakis, W.W. Lytton, P. Perdikaris, L. Petzold, *et al*. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ digital medicine*, 2(1):1–11, 2019.

ALCF. Combating covid-19 at the Argonne leadership computing facility. *https://www.alcf.anl.gov/combatingcovid-19-alcf*.

J.R. Allen. Why we need to rethink education in the artificial intelligence age. The Brookings Institute, January 31, 2019.

N. Ben-David, *et al.* Parallel algorithms for asymmetric read-write costs. *Proceedings of the 28th ACM Symposium on Parallelism in Algorithms and Architectures*, 2016.

P.C. Bermant *et al*. Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. *Scientific reports* 9.1: 1-10, 2019.

T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):34–43, 2001.

I.R. Braun and C.J. Lawrence-Dill. Automated methods enable direct computation on phenotypic descriptions for novel candidate gene prediction. *Frontiers in plant science*, 10:1629, 2020.

K.G. Burnett, D.S. Durica, D.L. Mykles, J. H. Stillman, and C. Schmidt. Recommendations for advancing genome to phenome research in non-model organisms. *Integrative and Comparative Biology*, 60(2):397–401, 2020.

F.S. Chance, J.B. Aimone, S. S. Musuvathy, M. R. Smith, C. M. Vineyard, and F. Wang. Crossing the cleft: Communication challenges between neuroscience and artificial intelligence. *Frontiers in Computational Neuroscience*, 14:39, 2020.

K.B. Cohen and L. Hunter. Getting started in text mining. *PLoS Comput Biol*, 4(1):e20, 2008.

M.P. Deisenroth, A.A. Faisal, and C. S. Ong. *Mathematics for machine learning*. Cambridge University Press, 2020.

J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, ImageNet: A large-scale hierarchical image database, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248-255, 2009.

J.A. Drew and AP. Henne. Conservation biology and traditional ecological knowledge: integrating academic disciplines for better conservation practice. *Ecology and Society*, 11(2), 2006.

T.F. Drumond, T. Vi´eville, and F. Alexandre. Bio-inspired analysis of deep learning on not-so-big data using dataprototypes. *Frontiers in computational neuroscience*, 12: 100, 2019.

R.C. Edmunds, B. Su, J.P. Balhoff, B. F. Eames, W. M. Dahdul, H. Lapp, J. G. Lundberg, T. J. Vision, R. A. Dunham, P. M. Mabee, *et al*. Phenoscape: identifying candidate genes for evolutionary phenotypes. *Molecular biology and evolution*, 33(1):13–24, 2015.

J. Felsenstein and G.A. Churchill. A Hidden Markov Method approach to variation among sites in rate of evolution. Molecular Biology and Evolution 13: 93-104, 1996.

N. Fleming. How artificial intelligence is changing drug discovery. *Nature* 557.7706: S55-S55, 2018.

R. Follmann, E.E.N. Macau, E. Rosa Jr., J.R.C. Piqueira. Phase Oscillatory Network and Visual Pattern Recognition. IEEE Trans. Neural Networks 26: 1539-1544, 2014.

R. Follmann and E. Rosa Jr. Predicting slow and fast neuronal dynamics with machine learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(11): 113119, 2019.

Z. Gao *et al*. Deep learning application in plant stress imaging: a review. *AgriEngineering* 2.3: 430-446, 2020.

M. Garcia. Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal*, 33(4): 111–117, 2016.

J. Gimenez-Luna, F. Grisoni and G. Schneider. Drug Discovery with Explainable Artificial Intelligence. Nature Machine Intelligence 2: 573-584, 2020.

J. Haugeland. Artificial Intelligence: The Very Idea. Cambridge, Mass: MIT Press, 1985.

W.D. Hawkins and S.E. DuRant. Applications of machine learning in behavioral ecology: Quantifying avian incubation behavior and nest conditions in relation to environmental temperature. *Plos One*, 15(8): e0236925, 2020.

T.B. Hayes. Welcome to the revolution: integrative biology and assessing the impact of endocrine disruptors on environmental and public health. *Integrative and comparative biology*, 45(2):321–329, 2005.

D.O. Hebb. The Organization of Behavior: A Neuropsychological Theory. Wiley, 1949.

J.J. Hopfield. Neural networks and physical systems with

emergent collective computational abilities. Proc. Natl. Acad. Sci., 79: 2554–2558, 1982.
M. Isakov *et al*., HPC I/O Throughput Bottleneck Analysis with Explainable Local Models, *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1-13, 2020.
E.C. Jones, G. Azeem, E.C. Jones, F. Jefferson. Impacting at Risk Communities using AI to optimize the COVID-19 Pandemic Therapeutics Supply Chain. *International Supply Chain Technology Journal*, 6(9), 2020.
A. Kaplan and M. Haenlein. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, 62:15–25, 2019.
Kislov, D.E. and K.A. Korznikov. Automatic windthrow detection using very-high-resolution satellite imagery and deep learning. *Remote Sensing* 12.7: 1145, 2020.
E. Kuhl, Data-driven modeling of COVID-19 — Lessons learned, Extreme Mechanics Letters, 40: 100921, 2020.
S. Lalmuanawma, J. Hussain, and L. Chhakchhuak. Applications of machine learning and artificial intelligence for covid-19 (sars-cov-2) pandemic: A review. *Chaos, Solitons & Fractals*, page 110059, 2020.
A. Lamba, P. Cassey, R.R. Segaran, and L. P. Koh. Deep learning for environmental conservation. *Current Biology*, 29(19): R977–R982, 2019.
M. W. Libbrecht and W. S. Noble. Machine learning applications in genetics and genomics. *Nature Reviews Genetics* 16.6: 321-332, 2015.
A. Mayor. Gods and Robots: Myths, Machines, and Ancient Dreams of Technology, Princeton University Press, 2018.
J. McCarthy *et al*. A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine* 27.4: 12-12, 2006.
Meena, J.S., Sze, S.M., Chand, U. *et al.* Overview of emerging nonvolatile memory technologies. *Nanoscale Res Lett* 9**,** 526, 2014.
Molloy, J. C. The open knowledge foundation: open data means better science. *PLoS biology* 9.12: e1001195, 2011.
Nikam, Rahul, and M. Michael Gromiha. Seq2Feature: a comprehensive web-based feature extraction tool. *Bioinformatics* 35.22: 4797-4799, 2019.
D. Noble. Physiology is rocking the foundations of evolutionary biology. *Experimental Physiology*, 98(8):1235–1243, 2013.
J. Paul and F. Jefferson. A Comparative Analysis of Student Performance in an Online vs. Face-to-Face Environmental Science Course From 2009 to 2016. *Frontiers in Computer Science*, v.1, 2019
S.L. Piano. Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and Social Sciences Communications* 7.1: 1-7, 220.
C.J. Rawlings and J.P. Fox. Artificial Intelligence in Molecular Biology: A Review and Assessment. Phil. Trans. R. Soc. London B 344: 353-363, 1994.
Renda, Andrea. Artificial Intelligence. Ethics, governance and policy challenges. CEPS Centre for European Policy Studies, 2019.
Roberts, Kirk, *et al*. TREC-COVID: rationale and structure of an information retrieval shared task for COVID-19. *Journal of the American Medical Informatics Association* 27.9: 1431-1436, 2020.
C. Rudin *et al.* Interpretable machine learning: Fundamental principles and 10 grand challenges. *arXiv: 2103.11251* (2021).
Russakovsky, O., Deng, J., Su, H. *et al.* ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis* **115,** 211–252, 2015.
H. Saarenmaa *et al.* An artificial intelligence modelling approach to simulating animal/habitat interactions. *Ecological Modelling* 44.1-2: 125-141, 1988.
J.S. Smith *et al.* Transforming Computational Drug Discovery with Machine Learning and AI, ACS Medicinal Chemistry Letters 9 (11), 1065-1069, 2018.
B.J. Stucky *et al.* The plant phenology ontology: A new informatics resource for large-scale integration of plant phenology data. Frontiers in Plant Science 9:517, 2018.
J. Sukumaran and L.L. Knowles. Trait-dependent biogeography: (re)integrating biology into probabilistic historical biogeographical models. *Trends in ecology & evolution*, 33(6):390–398, 2018.
J. Sukumaran, E. P. Economo, and L. Lacey Knowles. Machine learning biogeographic processes from biotic patterns: a new trait-dependent dispersal and diversification model with model choice by simulation-trained discriminant analysis. *Systematic Biology*, 65(3):525–545, 2016.
A. Taeihagh. Governance of artificial intelligence. *Policy and Society* 1-21 (2021).
J. Tian et al. Modular machine learning for Alzheimer's disease classification from retinal vasculature. *Scientific Reports* 11.1: 1-11, 2021.
R. Tonkens. A challenge for machine ethics. *Minds and Machines* 19.3: 421, 2009.
A.M. Turing. Computing Machinery and Intelligence. Mind 49: 433:460, 1936.
J. von Neumann. *The computer and the brain*. Yale University Press, 1958.
T.L. Wiemken and R.R. Kelley, Machine learning in epidemiology and health outcomes research, Annual Review of Public Health, 41: 21-36, 2020.
A. Yanguas-Gil, A. Mane, J. W. Elam, F. Wang, W. Severa, A.R. Daram, and D. Kudithipudi. The insect brain as a model system for low power electronics and edge processing applications. In *2019 IEEE Space Computing Conference (SCC)*, pages 60–66. IEEE, 2019.
A. Yu and K. Kumbier. Artificial intelligence and

statistics. *Frontiers of Information Technology & Electronic Engineering*, 19(1):6–9, 2018.