The Role and Promise of Artificial Intelligence in Medical Toxicology

Michael A. Chary^{1,2} · Alex F. Manini³ · Edward W. Boyer² · Michele Burns¹

Received: 26 September 2019 / Revised: 3 February 2020 / Accepted: 5 March 2020 / Published online: 25 March 2020
© American College of Medical Toxicology 2020

Abstract

Artificial intelligence (AI) refers to machines or software that process information and interact with the world as understanding beings. Examples of AI in medicine include the automated reading of chest X-rays and the detection of heart dysrhythmias from wearables. A key promise of AI is its potential to apply logical reasoning at the scale of data too vast for the human mind to comprehend. This scaling up of logical reasoning may allow clinicians to bring the entire breadth of current medical knowledge to bear on each patient in real time. It may also unearth otherwise unreachable knowledge in the attempt to integrate knowledge and research across disciplines. In this review, we discuss two complementary aspects of artificial intelligence: deep learning and knowledge representation. Deep learning recognizes and predicts patterns. Knowledge representation structures and interprets those patterns or predictions. We frame this review around how deep learning and knowledge representation might expand the reach of Poison Control Centers and enhance syndromic surveillance from social media.

Keywords Artificial intelligence · Machine learning · Knowledge representation · Big data

A Brief History of Artificial Intelligence

We use the term *artificial intelligence* to refer to efforts to imbue inanimate objects with the ability to reason about the world. This theme of animating inanimate objects recurs throughout history. Hephaestus, the Greek god of fire and metallurgy, made Talos, a bronze sentry to guard Crete against unwanted visitors [1]. Talos, an early example of an intelligent firewall, interrogated visitors to Crete to judge whether they were expected. If not, Talos engulfed the visitor in flames. The Argonauts bypassed Talos by asking Medea to cloak them, just as computer viruses may hide in innocuous programs.

Giving inanimate objects the ability to reason about the world became more realistic when Blaise Pascal implemented an early calculator, at its core an elaborate clock, in 1642 [1, 2]. Pascal's machine, like Charles Babbage's difference engine 130 years later, could only perform one calculation and was not reprogrammable. In 1808, Joseph-Marie Jacquard described the first reprogrammable machine, the Jacquard loom, which could be directed with punch cards to weave different patterns on textiles [3]. In 1879, Gottlieb Frege described a formal notation system for logical proofs [4]. This unified and simplified the representation of logical proofs, making it easier to spot errors and then debug them. In 1943, McCulloch and Pitts developed the artificial neuron to endow machines the ability to process information. An artificial neuron is a conceptual construct inspired by how the nervous system distributes information processing over a densely interconnected network of simple functional units [5]. In 1980, Haskell Curry and William Alvin Howard demonstrated that logical proofs written in a notation derived from Frege's could be automatically translated into computer programs [6]. Since 2000, the growth in computing power has allowed networks of artificial neurons (usually called *neural networks*) to approach the size and intricate connectivity patterns of the human nervous system.

Frege's symbolic representation of logic and Jacquard's reprogrammable machines represent two themes in artificial intelligence: knowledge representation and statistical learning.

Supervising Editor: Michael Hodgman, MD

✉ Michael A. Chary
michael.chary@childrens.harvard.edu

¹ Harvard Medical Toxicology Program, Department of Emergency Medicine, Boston Children's Hospital, Boston, MA, USA

² Department of Emergency Medicine, Brigham and Women's Hospital, Boston, MA, USA

³ Department of Emergency Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Knowledge representation depicts knowledge in terms of explicit relationships between data. Each relationship is written as a logical function. For example, the logical function *made_of(metal,satellite)* could represent “the satellite is made of metal” and *backing_up(car)* could represent “the car is backing up.” The field of knowledge representation can be understood as an effort to identify the proper construction of these logical functions as well as how to compose one logical function from previously defined functions.

Statistical learning represents information in terms of equations that make quantitative predictions, for example estimating the velocity of the car backing up at a certain time. Statistical relationships implicitly represent knowledge. For example, a negative velocity could indicate that the car is backing up. The combination of symbolic representation and statistical learning is termed statistical relational learning.

Deep Learning

Deep learning refers to the training of deep neural networks to cluster data or recognize patterns in data. A neural network is *deep* if it contains one or more hidden layers (Fig. 1). A neural network is *shallow* (usually called *simple*) if it contains no hidden layers (goes straight from green to red in Fig. 1). A shallow neural network resembles a spinal reflex, mapping sensory input to motor output in one step (synapse). A *layer* of the neural network refers to a level of processing, similar to how retinal ganglion cells constitute one layer of sensory processing and neurons in the primary visual cortex represent a deeper (or, sometimes *higher*) level of processing. A layer of the neural networks is *hidden* if it neither receives input from the external world nor produces output that is presented to the external world.

Fig. 1 Schematic of a deep neural network. Green neurons receive external input. Blue intermediary neurons combine the input with their own state according to an activation function. The states of some of the intermediary neurons contribute to the red output neuron.

Examples of single-layer (shallow) neural networks include all clinical scoring algorithms (qSOFA, PESI, PORT, TIMI, ABCD2, CURB-65). Each algorithm takes an input and calculates output from one round of processing. One must balance this recognition of the general in the specific with an appreciation for the clinical differences, for example, between sepsis and myocardial ischemia.

A deep neural network is, typically, *trained* on one set of data and *evaluated* (or *tested*) on a second set. The second set of data should be statistically indistinguishable from the first with respect to the outcomes of interest. Training a neural network refers to adjusting the strengths of connections between neurons, similar to determining the coefficients for each variable in linear regression. Training a neural network on one data set and evaluating it on an independent set resembles deriving a clinical decision rule on retrospective data, and evaluating it on prospective data. The perils of overfitting and applying the rule indiscriminately to disparate populations apply to deep learning just as to the derivation of clinical decision rules.

A neural network is conceptually similar to nonlinear regression. A neural network that is deeper than one layer resembles a chain of regressions, where the dependent variable from one regression is treated as an independent variable in the next regression layer. A deep neural network can represent more complex functions than a shallow neural network. Neural networks may predict quantitative or qualitative variables. Neural networks that predict qualitative variables are often called classifiers. Just as with logistic regression, the choice of the number of categories and the relative amounts of data for each category can influence the accuracy and generalizability of the classifier. A neural network that is trained to be a binomial classifier should be retrained on independent data before being used as a multinomial classifier.

Splitting data into testing and training sets faces unique issues related to sampling. Some training algorithms perform better if the training set is biased to include certain events. This resembles how a resident or fellow trains on a biased sample of the general population to deliberately develop a narrow expertise. This analogy also demonstrates the metaphysical tension between training “generalist” and “specialist” neural networks. A network that identifies a botulism exposure may not work as well for identifying a buckthorn exposure. In evaluating the performance of the neural network, we must consider that it is more important to identify some exposures over others.

There are multiple ways to evaluate the performance of a neural network. Clinically relevant metrics include the sensitivity and specificity. Machine learning traditionally uses precision and recall. Precision is the positive predictive value. The quantity called *recall* in machine learning is mathematically equivalent to specificity, if in calculating specificity, true positive is taken to mean when the

neural network's classification is the same as the expert panel's. Cross-validation refers to quantifying the average performance over repeated divisions of the data into training and testing sets. It estimates how much a neural network's performance depends on the idiosyncrasies of the current data set. If the splitting of the data must be biased for the algorithm to achieve a certain performance, then partitions of the data are not independent, which will confound the cross-validation's estimate of performance.

Just as large multivariate models require large data sets with precise measurements to accurately estimate the weighting coefficients for each variable, so too do deep neural networks require large models to estimate the interaction coefficients between neurons. A neural network may need hundreds of thousands of images or millions of images to learn to recognize stigmata of one condition on X-ray, in contrast to the few hundreds a radiologist uses to learn during his or her training.

A deep neural network developed to detect diabetic retinopathy or hemorrhage from retinal fundograms was trained on 128,175 and then validated on 9963 retinal images [7]. The labels for the training set were provided by 54 licensed ophthalmologists and ophthalmology senior residents. The validation consisted of comparing the deep neural network's assessment to the consensus rating of 7 board-certified ophthalmologists. Another deep network was trained on 680 two-view chest X-rays to determine whether or not they contained lesions consistent with tuberculosis and validated on 320 X-rays [8].

The weights in hidden layers of neurons are hard to interpret, as can be the coefficients in regression models. This approach creates two difficulties. First, it is unclear how to combine the output of multiple neural networks, just as it can be unclear how to combine clinical decision rules that use the same variables to predict different things. Clinical context, which is external to the model, guides, for example, whether to consider sinus tachycardia as suggestive of sepsis or a pulmonary embolus. Second, it is often difficult to understand how changes in inputs to neural networks affect a neural network's prediction, or whether those changes make clinical sense. Knowledge representation, by ontologies, discussed in the next section, provides a way to make the output of neural networks more comprehensible to clinicians by casting their output in the context of biomedical knowledge.

A barrier to clinicians using deep learning algorithms is the requirement to enter extensive amounts of data. A hypothetical deep neural network to predict pneumonia severity might require the clinician to enter all data for the PORT and CURB-65 scores. PORT and CURB-65 both use age, which introduces collinearity. PORT and CURB-65, moreover, answer different but related questions. It may not be appropriate to always combine them. In addition, each measurement is associated with uncertainty. The imprecision of the final score reflects the combined uncertainty in all the underlying

measurements. A model that combines the PORT and CURB-65 scores could, thus, be less accurate than either score alone. For this software to run unobtrusively in the background without clinician input, the administrative and logistic barriers to abstracting information from any chart in any electronic medical system would have to be overcome.

One can critically evaluate neural networks using the same principles as one would use to evaluate a large multivariate regression model. Both models attempt to predict an unknown or future value based on uncertain input. Both can mistake a correlation arising from the peculiarity of one data set as a meaningful generalizable correlation. Both assume that variables interact in certain ways, and that the nature of these interactions does not vary across groups of people.

To demonstrate how neural networks could advance toxicology, consider designing a neural network to identify the most likely exposure associated with a phone call to a poison center. Once trained and tested, this network could extend the expertise of on-call toxicology to the Web and areas with fewer staff. The neural network would be trained and evaluated on poison center cases. In training, the group of toxicologists would label each poison center record with their consensus on the exposure. The neural network would uncover features in the records that identify the consensus exposure.

This example demonstrates some of the issues with applying neural networks to clinical data. Neural networks parse text, but do not (yet) understand it. A neural network, for example, could consider *apneic* and *dyspneic* as more similar than *apneic* and *not breathing* because *apneic* and *dyspneic* share more characters. This is the difference between natural language *processing* and natural language *understanding*. A neural network trained on poison center cases looks for patterns in the text. Neural networks, unless programmed to do so, cannot bring prior knowledge to bear. An ontology (discussed in the immediately following section “[Knowledge Representation](#)”) provides a way to bring in prior knowledge.

Knowledge Representation

The goal of knowledge representation is to explicitly represent relationships between entities that describe a portion of reality. An unambiguous and logically consistent representation is called an *ontology*. A description is *unambiguous* if it has only one meaning. A representation (collection of descriptions) is *logically consistent* if it contains no two descriptions that contradict each other.

Logically consistent ontologies allow computers to perform automated inference. Automated inference refers to the ability of a computer to generate statements that follow from the stated facts, but are not themselves explicitly stated. To follow a classic example, a computer using an ontology could conclude that *Socrates is mortal*, given the statements that (1)

All men are mortal and (2) *Socrates is a man*. In medical toxicology, automated inference could increase the medical sophistication of Web interfaces, such as webPoisonControl [9], text services, or even smart phone apps. It could generate a ranked list of possibilities for an unknown exposure, bringing all toxicologic knowledge, and even knowledge from other fields, to bear.

An ontology that describes the knowledge base of medical toxicology could help automate the management of routine intoxications and exposures and combine knowledge across disparate digital sources to inform the management of complex or less common exposures. It might even identify gaps between clinical guidelines and basic research, and automatically generate hypotheses to reconcile those inconsistencies. One could, for example, use an ontology to direct the combination of pharmacokinetic/pharmacodynamics and gene expression to hypothesize at scale about genetic variations in drug metabolism. Physician participation in constructing ontologies can accelerate the representation of clinically relevant concepts in ontologies.

The OBO (Open Biomedical Ontology) Foundry provides a listing of ontologies for science or medicine [10], for example the Foundational Model of Anatomy [11], ChEBI (Chemical Entities of Biological Interest) [12], or Infectious Disease Ontology [13]. Examples closer to medical toxicology include DINTO, the drug interactions ontology [14], and DIDEO, the drug interaction evidence ontology [15]. It is considered best practice to refer to a central repository of ontologies, such as the OBO Foundry, before creating a new ontology, to identify which entities and relationships already have formal definitions.

One currently implemented knowledge representation in medicine is MeSH, the National Library of Medicine (NLM) Medical Subject Headings. MeSH is not logically consistent because it uses the same term to refer to multiple entities, multiple terms to refer to a single entity, and the same term to refer to an entity and a relationship. MeSH is designed to catalogue medical research to help researchers and clinicians access specific articles and browse related articles. It reflects, rather than clarifies, the ambiguity of human language. MeSH's lack of a one-to-one relationship reflects the difference between the human and computer representation of language. For example, the MeSH term *disease* is used to mean a specific disease, such as appendicitis, as well as the general concept of disease as distinct from health. In a subtler example, the MeSH term *cardiomegaly* is stated to be equivalent to “enlarged heart” or “heart enlargement.” The context that “enlarged” means “pathologically enlarged” is missing; “enlarged heart” is not the same as “heart enlargement.” The former describes a structure. The latter describes a process. Humans can recognize this difference between a structure and a process and abstract beyond it. Without explicit instructions, a computer cannot.

Creating an Ontology

To construct a representation of knowledge in a domain, we begin by identifying relevant entities and relationships in a domain of interest. A good starting point is to list the most common nouns and verbs specific to that domain of interest. Nouns correspond to entities and verbs to relationships, to a first approximation. Part-of-speech tagging is a natural language processing (NLP) technique that can identify nouns and verbs in text. There are many implementations of part-of-speech tagging, for example nltk (the natural language toolkit) [16], Spacy [17], or WEKA [18].

Basic Formal Ontology provides a set of guidelines for constructing ontologies. Most ontologies in the OBO Foundry adhere to Basic Formal Ontology (BFO) standards. Constructing a new ontology that follows BFO standards allows immediate interoperability with all other BFO-compliant ontologies. Most current BFO-compliant ontologies represent concepts in terms of a triple, a combination of three concepts, such as “tachycardia,” “is symptom of,” and “anticholinergic toxidrome.” The elements of a triple are called the subject, verb, and object, respectively, reflecting the terms as used in linguistics. Representing concepts as triples facilitates fast and strict type-checking. Type-checking refers to the validation that the data source is providing is what the source claims. An example from healthcare of type-checking is knowing the reference ranges for vital signs. An adolescent male is less likely than a 1-year-old to have a baseline heart rate of 160 beats/minute. In ontological terms, we could consider the triple “adolescent male,” “has heart rate,” and “X beats per minute,” and restrict values of *X* to physiologic values, allowing the automatic flagging of other values.

There is tension between the ambiguity and metaphorical natures of human language and the linguistic precision an ontology requires. In an ontology, each entity or relationship can have only one meaning. Words in English often have multiple meanings and multiple words can have the same meaning. The word *lead* can refer to element (an entity), a person in a play (an entity, by ellipsis [omission of words that probably can be understood from context] of the phrase “lead performer”), or a verb (a relationship). Type-checking reduces the ambiguity by restricting meanings of a word to only certain uses. Type-checking would allow the word *lead* to acquire the meaning of “to guide or conduct someone to a goal” only if *lead* was being used as a verb.

Type-checking may not be sufficient to isolate one meaning of a word. Lead as an element (also written *qua* element) does not specify the oxidation state or isotope. Even the symbol $^{204}\text{Pb}^{2+}$ is not unique. It refers to any atom of ^{204}Pb in the 2+ oxidation state, not to a particular atom. This distinction between the general and particular is crucial to represent, specifically for, concepts in toxicology such as compounding errors or environmental exposures. It also explains the

difficulty machines have understanding everyday language as opposed to miming it. The phrase “the patient took quetiapine” is confusing for a machine. A machine will infer the meaning of that sentence as “the patient took a [certain dose] of a pill that contains the active ingredient quetiapine” only if the machine understands figures of speech such as ellipsis or synecdoche [figure of speech where part of something symbolizes the whole].

Ontologies are usually written in OWL (the Web Ontology Language) [19]. Each entity or relationship is delimited by tags. Each tag in this triple denotes a concept, not a markup. An OWL entry in an ontology about toxidromes might read “<tachycardia> <is symptom of> <anticholinergic toxidrome>.” OWL is a dialect of XML (Extensible Markup Language). XML is a generalization of HTML (Hypertext Markup Language) that allows users to define tags, for example <tachycardia>.

Many applications exist to facilitate ontology creation. Protégé is a common free application with desktop and Web versions that provide an interface similar to Windows Explorer [20]. One can also create an ontology via spreadsheet applications such as Microsoft Excel, Apple Numbers, or LibreOffice, using rows to denote entities and columns to denote relationships. Editors such as Protégé can convert spreadsheets, that is XLS (Excel Spreadsheet 1997–2000 format), XLSX (Excel Spreadsheet Extended), or CSV (comma-separated value) files, to OWL format. XLSX, as the second X suggests, is a dialect of XML.

To construct an ontology of toxidromes, for example, one could begin with entities such as *tachycardia*, *sympathomimetic toxidrome*, or *mydriasis*, and relationships such as *hypotensive*, *symptomatic*, or *seizing*. This ontology of toxidromes would recognize a patient demonstrating the combination of [hyperthermia, mydriasis, ileus, urinary retention, xerostomia, tachycardia, and flushed dry skin] as demonstrating anti-muscarinic toxicity. The degree of the ontology’s confidence in its prediction could be a proportion of the number of anti-muscarinic symptoms the patient manifested.

One starting point for creating an ontology would be to apply part-of-speech tagging to consensus statements by The American College of Medical Toxicology (ACMT) or the American Academy of Clinical Toxicology (AACT). The additional layer in our neural network example could use an ontology to include the prior knowledge expressed in these consensus statements. The ontology would relate each specific exposure to actions to be taken. Querying the ontology would provide an explanation for each action recommended, for example “contact child protective services” because “opioid exposure” was identified in a 2-year-old.

Machine learning techniques are often more accurate on data sets larger than a consensus statement. A rule of thumb in statistics is that the confidence interval around a sample statistic is inversely proportional to the square root of the

sample size. Increasing the sample size by 4 would halve the confidence interval. The confidence interval around many sample statistics used in machine learning converges more slowly, for example proportional to the fourth root.

Representing medical knowledge with an ontology provides a way for the neural network to explain its medical decision making to a supervising medical toxicologist. The ontology could print out the exposure and axioms that the neural network invoked to come to its recommendation and, in doing so, explain itself. Calls that involve a common exposure and for which the ontology does not recommend further action might be able to be handled automatically. Without another layer, the ontology cannot prioritize its recommendations. Markov logic networks provide a means to implement that layer.

Combining Neural Networks and Knowledge Representation: Markov Logic Networks

A Markov logic network (MLN) combines symbolic reasoning and deep learning [21]. MLNs overcome a significant disadvantage of ontologies; namely, ontologies cannot represent shades of gray. They can, for example, represent “sick” or “not sick,” but not “kind of sick.” MLNs assign a probability to each triple and can calculate a probability for any inferred triple. This allows one to compare the likelihood of two strains of reasoning, as one might do when creating a differential diagnosis or comparing alternative hypotheses to explain experimental results. MLNs also overcome a significant disadvantage of neural networks; namely, neural networks do not represent their latent knowledge or information processing in a way that is easily interpretable by humans or that leverages prior knowledge.

Markov logic networks have been used to reconcile conflicting data when making a comprehensive map of drug metabolism in colorectal cancer [22] and to develop drugs targeted to particular symptoms of specific neurological diseases [23]. Preliminary data suggest that MLNs can identify toxidromes from patient presentations [24]. The field of MLNs is young. The paper introducing MLNs was published 12 years ago, with only 6 peer-reviewed manuscripts describing a clinical application of MLNs since then, of which only 1 describes original research.

Surveillance of Novel Psychoactive Substances

The rate of emergence of novel psychoactive substances, hundreds in the last decade, outstrips the ability of traditional methods of surveillance to identify and characterize each substance [25]. Social media provide a rich source of data on

emerging substances. Deep learning might be able to automatically identify high-level patterns in drug use by combining data from social media, poison control logs, published reports, and national surveys.

A neural network could continuously scan social media for mentions of novel substances and effects. An ontology could summarize those descriptions and automatically compare those descriptions with the effects of known drugs to generate hypotheses about the class and mechanism of action of these emerging substances. It could match substances with different names but nearly identical effects, providing a short list for manual curation of xenobiotics that might be closely related or the same substance referred to by different names.

Monitoring for the emergence of novel psychoactive substances (e.g., derivatives of tryptamine, phenethylamine, or fentanyl) provides a challenge for neural networks. The identification of an emerging trend is similar to anomaly detection. Neural networks are better at pattern recognition than anomaly detection.

Consider a neural network that monitors social media (or national registries) for mentions of substances. The neural networks classify each communication as mentioning substance use or not and then identifies which substance(s) in those communications that mention substance use. One straightforward approach to anomaly detection is to identify a time period or geographic area in which the neural network identifies communications as mentioning substance use, but cannot identify the substance, and then manually review that area. An ontology could accelerate that review by performing a deeper linguistic analysis of the communication. It could, for example, divide the communications in question, which could still number in the millions, into those discussing substance use as a phenomenon from those discussing the use of a particular substance. Then, it could identify candidate names for the substance from the communications and extract co-mentioned symptoms.

Combining Toxicologic Data with Other Streams for Toxicovigilance

The dynamics of the opioid epidemic demonstrate how a fuller understanding of opioid misuse arose from combining multiple sources of information. Opioid use initially grew as healthcare providers prescribed more opioids to treat pain. More judicious prescribing practices in the face of rising concern over adverse effects from opioids decreased the supply of prescription opioids and deaths due to prescription opioids. This apparent decrease was, on further analysis, a shift from consumption of prescription opioids to consumption of heroin, especially black tar heroin. People with untreated opioid use disorder substituted “street” opioids

for prescription opioids in response to the decreasing supply of prescription opioids. Legislative and law enforcement efforts led to a decrease in the heroin supply, which drove a shift toward the consumption of illicitly manufactured fentanyl and its derivatives. Federal agencies recognized that efforts to stem the supply of opioids were ineffective, motivating them to develop approaches to treat the psychiatric disorders that drive consumption.

The techniques discussed in this review could accelerate this process as follows. A neural network could monitor multiple sources of information—social media, official communications from legislative and law enforcement bodies, and economic information. It could then search for patterns and deviations from those patterns, for example discussions on social media of new substances that are compared to fentanyl. An ontology could (1) direct how to combine those sources of information and (2) summarize the information found by the neural network.

Ethical Considerations Artificial intelligence and machine learning are becoming ubiquitous. To make clinical decisions, these technologies must demonstrate a mastery of medical knowledge and a competency at acquiring the information relevant to each case. Neural networks, combined with ontologies and Markov logic networks, may develop into a digital colleague able to explain its rationale as one human colleague might to another. The integration of nonhuman reasoning into clinical decision making raises, beyond accountability and competency, issues of how to handle disagreements between a human and digital colleague and what to do if medicine becomes too complex for any human to understand.

Future Directions

This review introduces two aspects of AI: deep learning and knowledge representation with ontologies. Deep learning acts similarly to generalized linear models that use latent variables. Ontologies allow the computer to reason about information and explain its reasoning in terms a human can understand. Markov logic networks combine these two approaches. For medical toxicology to make full use of these techniques, efforts to curate ontologies of toxicologic knowledge and integrate these ontologies with existing ontologies are necessary. The maintenance of ontologies is most often done by hand. How to automatically update ontologies is an area of ongoing research.

Poison Control Centers allow one toxicologist via phone to provide care for more patients than he or she could physically visit. AI, similarly, will allow us to expand beyond human capacity the knowledge base and acumen we bring to each patient.

Sources of Funding Dr. Chary is supported by the Loan Repayment Program (National Institute on Drug Abuse, National Institutes of Health). Also supported by grants from the National Institutes of Health (R01-DA037317-05, to Dr. Manini, and R01-DA033769-01, to Dr. Boyer).

Compliance with Ethical Standards

Conflicts of Interest None

References

- Aspray W. Computing before computers, vol. 50010. Ames: Iowa State University Press; 1990. p. 2000.
- Randell B. The origins of digital computers: selected papers (monographs in computer science). Berlin: Springer-Verlag; 1982.
- Posselt EAE. The Jacquard machine analyzed and explained. Philadelphia: Dando Print and Publishing; 1888. [cited 2019 Jul 15]. Available from: https://www.forgottenbooks.com/en/books/TheJacquardMachineAnalyzedandExplained_10290845
- Frege G. Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. In: van Heijenoort J, editor. From Frege to Gödel: a source book in mathematical logic (1879–1931). Cambridge: Harvard University Press; 1977. p. 1–82.
- McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. Springer. 1943;5:115–33.
- Howard WA. The formulae-as-types notion of construction. In: Seldin J, Hindley J, editors. To HB Curry: essays on combinatory logic, lambda calculus and formalism. London: Academic Press; 1980. p. 479–90.
- Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316:2402–10.
- Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology. 2017;284:574–82.
- Litovitz T, Benson BE, Smolinske S webPOISONCONTROL: can poison control be automated? Am J Emerg Med. Elsevier. 2016;34:1614–9.
- Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. Nature Publishing Group. 2007;25:1251. Available from: <https://doi.org/10.1038/nbt1346>
- Rosse C, Mejino Jr JL V. A reference ontology for biomedical informatics: the foundational model of anatomy. J Biomed Inform. Elsevier. 2003;36:478–500.
- Degtyarenko K, De Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, et al. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. Oxford University Press. 2007;36:D344–D350.
- Cowell LG, Smith B. Infectious disease ontology. Infect Dis Inform. Springer. 2010. p. 373–95.
- Herrero-Zazo M, Segura-Bedmar I, Hastings J, Martínez P DINTO: using OWL ontologies and SWRL rules to infer drug–drug interactions and their mechanisms. J Chem Inf Model. ACS Publications. 2015;55:1698–707.
- Judkins J, Tay-Sontheimer J, Boyce RD, Brochhausen M. Extending the DIDEO ontology to include entities from the natural product drug interaction domain of discourse. J Biomed Semantics. BioMed Central. 2018;9:15.
- Loper E, Bird S. NLTK: the natural language toolkit. arXiv preprint cs/0205028. 2002.
- Srinivasa-Desikan B. Natural language processing and computational linguistics: a practical guide to text analysis with Python, Gensim, spaCy, and Keras. Birmingham: Packt Publishing Ltd; 2018.
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH The WEKA data mining software: an update. ACM SIGKDD explorations newsletter. ACM. 2009;11:10–8.
- Bechhofer S, Van Harmelen F, Hendler J, Horrocks I, McGuinness DL, Patel-Schneider PF, et al. OWL web ontology language reference. W3C recommendation. 2004;10.
- Noy NF, Sintek M, Decker S, Crubézy M, Fergerson RW, Musen MA Creating semantic web contents with protege-2000. IEEE Intell Syst. IEEE. 2001;16:60–71.
- Richardson M, Domingos P. Markov logic networks. Mach Learn. Springer. 2006;62:107–36.
- Martínez-Romero M, Vázquez-Naya JM, Rabunal JR, Pita-Fernández S, Macenlle R, Castro-Alvarino J, et al. Artificial intelligence techniques for colorectal cancer drug metabolism: ontologies and complex networks. Curr Drug Metab. Bentham Science Publishers. 2010;11:347–68.
- Vázquez-Naya JM, Martínez-Romero M, Porto-Pazos AB, Novoa F, Valladares-Ayerbes M, Pereira J, et al. Ontologies of drug discovery and design for neurology, cardiology and oncology. Curr Pharm Des. Bentham Science Publishers. 2010;16:2724–36.
- Chary M, Burns M, Boyer E. Tak: the computational toxicological machine. Clin Tox. 2019;57(6):482 [abstract EAPCCT].
- Madras BK. The growing problem of new psychoactive substances (NPS). Neuropharmacol New Psychoact Subst (NPS). Springer; 2016. p. 1–18.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.