2

Data, Analytics and Interoperability Between Systems (IoT) is Incongruous with the Economics of Technology: Evolution of Porous Pareto Partition (P3)

Shoumen Palit Austin Datta[1,2,3,]*, Tausifa Jan Saleem[4],
Molood Barati[5], María Victoria López López[6], Marie-Laure Furgala[7],
Diana C. Vanegas[8], Gérald Santucci[9], Pramod P. Khargonekar[10], and
Eric S. McLamore[11]

[1]*MIT Auto-ID Labs, Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
[2]*MDPnP Interoperability and Cybersecurity Labs, Biomedical Engineering Program, Department of Anesthesiology, Massachusetts General Hospital, Harvard Medical School, 65 Landsdowne Street, Cambridge, MA 02139, USA*
[3]*NSF Center for Robots and Sensors for Human Well-Being, Collaborative Robotics Lab, School of Engineering Technology, Purdue University, 193 Knoy Hall, West Lafayette, IN 47907, USA*
[4]*Department of Computer Science and Engineering, National Institute of Technology Srinagar, Jammu & Kashmir 190006, India*
[5]*School of Engineering, Computer and Mathematical Sciences Auckland University of Technology, Auckland 1010, New Zealand*
[6]*Facultad de Informática, Deparmento Arquitectura de Computadores y Automática, Universidad Complutense de Madrid, Calle Profesore Santesmases 9, 28040 Madrid, Spain*
[7]*Director, Institut Supérieur de Logistique Industrielle, KEDGE Business School, 680 Cours de la Libération, 33405 Talence, France*
[8]*Biosystems Engineering, Department of Environmental Engineering and Earth Sciences, Clemson University, Clemson, SC 29631, USA*
[9]*Former Head of the Unit, Knowledge Sharing, European Commission (EU) Directorate General for Communications Networks, Content and Technology (DG CONNECT); Former Head of the Unit Networked Enterprise & Radio Frequency Identification (RFID), European Commission; Former Chair of the Internet of Things (IoT) Expert Group, European Commission (EU); INTEROP-VLab, Bureau Nouvelle Région Aquitaine Europe, 21 rue Montoyer, 1000 Brussels, Belgium*
[10]*Vice Chancellor for Research, University of California, Irvine and Distinguished Professor of Electrical Engineering and Computer Science, University of California, Irvine, California 92697*
[11]*Department of Agricultural Sciences, Clemson University, Clemson, SC 29634, USA*

Opinions expressed in this essay (chapter) are due to the corresponding author and may not reflect the views of the institutions with which the author is affiliated. Listed coauthors are not responsible and may not endorse any/all comments and criticisms.

2.1   Context

Since 1999, the concept of the Internet of Things (IoT) was nurtured as a marketing term [2] which may have succinctly captured the idea of data about objects stored on the Internet [3] in the networked physical world. The idea evolved while transforming the use of radio frequency identification (RFID) where an alphanumeric unique identifier (64-bit EPC [4] or electronic product code) was stored on the chip (tag [5]) but the voluminous raw data were stored on the Internet, yet inextricably and uniquely linked via the EPC, in a manner resembling the structure of internet protocols [6] (64-bit IPv4 and 128-bit IPv6 [7]). IoT and, later, *cloud of data* [8] were metaphors for ubiquitous connectivity and concepts originating from ubiquitous computing, a term introduced by Mark Weiser [9] in 1998. The underlying importance of data from connected objects and processes usurped the term big data [10] and then twisted the sound bites to create the artificial myth of "Big Data" sponsored and accelerated by consulting companies. The global drive to get ahead of the "Big Data" tsunami, flooded both businesses and governments, big and small. The chatter about big data garnished with dollops of fake AI became parlor talk among fish mongers [11] and gold miners, inviting the sardonicism of doublespeak, which is peppered throughout this essay.

Much to the chagrin of the thinkers, the laissez-faire approach to IoT percolated by the tinkerers overshadowed hard facts. The "quick & dirty" anti-intellectual chaos adumbrated the artifact-fueled exploding frenzy for new revenue from "IoT Practice" which spawned greed in the consulting [12] world. The cacophony of IoT in the market [13] is a result of that unstoppable transmutation of disingenuous tabloid fodder to veritable truth, catalyzed by pseudo-science hacks, social gurus, and glib publicity campaigns to drum up draconian "dollar-sign-dangling" predictions [14] about "trillions of things connected to the internet" to feed mass hysteria, to bolster consumption. Few ventured to correct the facts and point out that *connectivity without discovery* is a diabolical tragedy of egregious errors. Even fewer recognized that the idea of IoT is *not a point* but an **ecosystem**, where collaboration adds value.

The corporate orchestration of the *digital by design* metaphor of IoT was warped solely to create demand for sales by falsely amplifying the lure of increasing performance, productivity, and profit, far beyond the potential digital transformation could deliver by embracing the rational principles of IoT (Figures 2.1–2.4).

Ubiquitous connectivity is associated with high cost of products (capex or capital expense) but extraction of "value" to generate return on investment (ROI) rests on the ability to implement SARA, a derivative of the PEAS paradigm (see Figures 2.7 and 2.8). SARA – **S**ense, **A**nalyze, **R**espond, **A**ctuate – is not a linear concept. Data and decisions necessary for SARA make the conceptual illustration more akin to The Sara Cycle, perhaps best illustrated by the analogy to the Krebs

By 2022, M2M will be a USD 1.2 trillion opportunity

Total revenue from machine-to-machine, 2011–2022
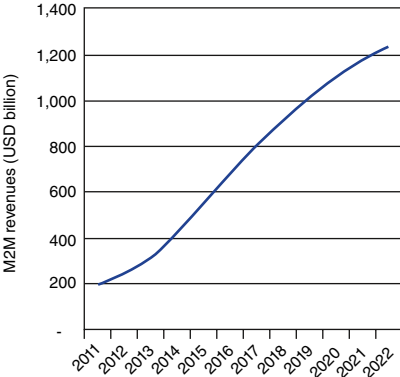
- **Total M2M revenue will grow from USD200 billion in 2011 to USD1.2 trillion in 2022, a CAGR of 18%**
- **Total revenue includes:**
  - device costs where connectivity is integral to the device
  - module costs where devices can optionally have connectivity enabled
  - monthly subscription, connectivity and traffic fees

Machina Research

Figure 2.1 From the annals [15] of the march of unreason: *Internet of things: $8.9 trillion market in 2020, 212 billion connected things*. It is blasphemous and heretical to suggest that this is a *research* [16] outcome.
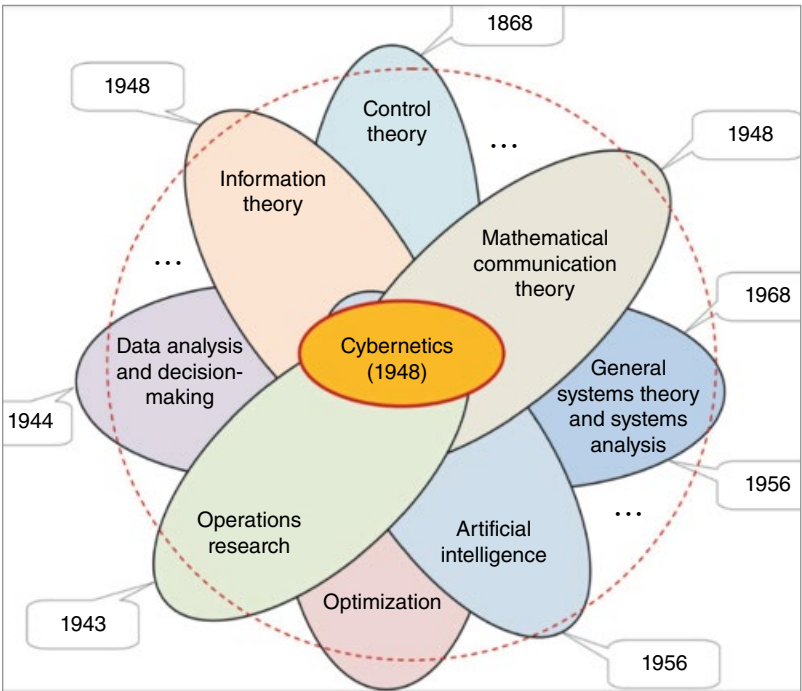
Figure 2.2 A Century of convergence the composition and structure of cybernetics [17]. *Source:* Novikov, D.A. Systems theory and systems analysis. Systems engineering. *Cybernetics.* vol. 47. Springer International Publishing. 2016, pp. 39–44. © 2016, Springer Nature.
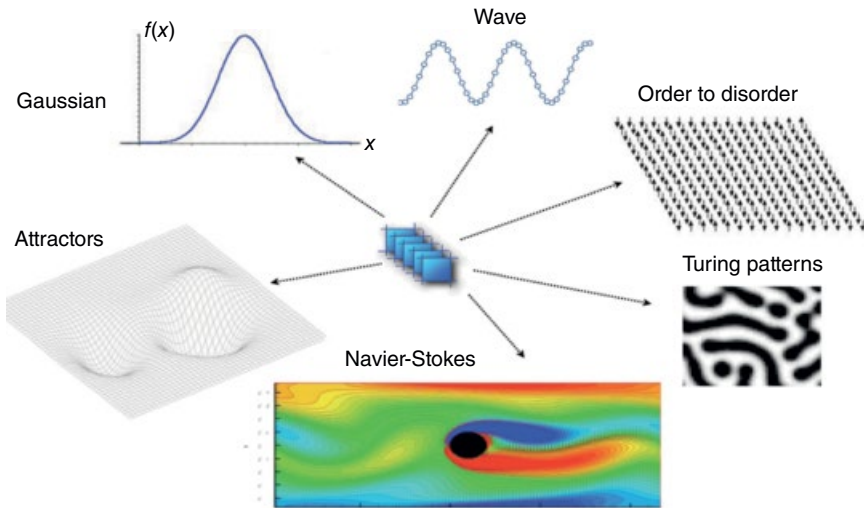
Figure 2.3 Only a few models may capture the behavior of a wide range of systems, underlies the idea of universality [18] (models illustrated in this figure: Gaussian distribution, wave motion, order to disorder transitions, Turing patterns, fluid flow described by Navier–Stokes equations, and attractor dynamics). *Source:* Based on Williams, L.P. (1989). "André-Marie Ampère." *Scientific American*, vol. 260, no. 1, pp. 90–97. © 1989, Scientific American.

[28] Cycle, an instance of bio-mimicry. Data and decisions constantly influence, optimize, reconfigure, and change the parameters associated with, *when* to sense, *what* to analyze, *how* to respond, and *where* to actuate or auto-actuate. Combining SARA with the metaphor of IoT by design may help to ask these questions, with precision and accuracy.

It is hardly necessary to overemphasize the value of the correct questions for each element of SARA in a matrix of connected objects, relevant entities which can be discovered, distributed nodes, related processes, and desired outcomes. Strategic inclusion of SARA guides key performance indicators (KPI). Lucidity and clarity of thoughtful integration of digital by design idea is key to reconfiguring operations management. Execution and embedding SARA is not a systems integration task but rather a fine-tuned *synergistic* integration based on the *weighted combination of dependencies* in the SARA matrix. Failure to grasp the role of data and semantics of queries, in the context of KPI may increase transaction costs, reduce the value proposition for customers, and obliterate ROI or profitability.

This essay meanders, not always aimlessly, around discussions involving data and decision. It also oscillates, albeit asynchronously, between a broad spectrum
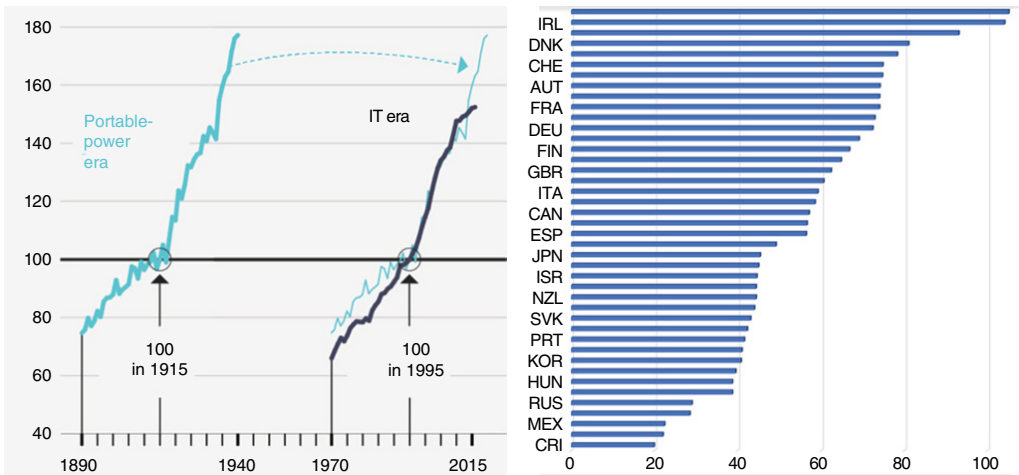
Figure 2.4 (Left) Labor-Productivity Index [19]: Has data failed to deliver? IT was billed as the bridge between *the haves* and *the have-nots*. General process technologies take ~25 years to reach market adoption [20]. *Source:* Syverson, C. (2018). Why hasn't technology sped up productivity?. *Chicago Booth Review.* © 2018, Chicago Booth Review. (Right) Labor Productivity [21] (OECD 2018) is yet another example how the arithmetic of productivity (ratio between volume of output vs input) is misguided, misdiagnosed, mismeasured, and misused as a metric of economic realities. Making Mexico (22.4) appear to be one-fifth as "productive" as Ireland (104.1) suggests formulaic manipulations [22] (GDP per hour worked, current prices, PPP).

of haphazard realities or "dots" which may be more about esoteric analysis rather than focusing on delivering real-world value. In part, this discussion questions the barriers to the rate of diffusion of technologies in underserved communities. Can implementing *simple* tools act as affordable catalysts? Can it lift the quality of life, in less affluent societies, by enabling meaningful use of data, perhaps small data, at the right time, at the lowest cost?

The extremely nonlinear business of delivering tools and technologies makes it imperative to consider the trinity of systems' integration, standards, and interoperability. We advocate that businesses may wish to gradually disengage with the product mindset (sensors, hardware, and software) and engage in the *ecosystem* necessary to deliver ***services*** to communities. The delivery of service to the enduser must be synergized. Hence, system integration may be a subset of synergistic integration. But, before we can view this "whole," it is better to understand the coalition of cyber (data) with the physical (parts). In many ways, this discussion is about cyberphysical systems (CPS) but not for lofty purposes, such as landing on Mars, but for simple living, on Earth.

2.2 Models in the Background

Because it may be difficult to grasp the whole, we tend to focus on the part, and parts, closest to our comfort zone, in our area of interest. This reductionist approach may be necessary *ab initio* but rarely yields a solution, *per se*. Reconstruction requires synthesis and synergy, the global glue which underlies mass adoption and diffusion, of tools, in an age of integration, which, itself, is a khichuri [29] of parts, some known (industrial age, information age, and systems age) and others, parts unknown.

Divide and conquer still remains a robust adage. It may be the philosophical foundation of reductionism. The latter has rewarded us with immense gains in knowledge and the wisdom as to why this *modus operandi* is *sine qua non*. For example, the pea plant (*Pisum sativum*) unleashed the cryptic principles of genetics [30] and unicellular bacteria shed light on normal physiological underpinnings of feedback control [31] common in genetic circuits as well as regulatory networks for maintenance and optimization of biological homeostasis, quintessential for health and healthcare in humans and animals. Cancer biology was transformed by Renato Dulbecco [32] by ***reducing*** the multifactorial complexity of human cancer research to focus on a ***single gene*** (the SV40 large T-antigen) from Papova viruses.

Biomimicry also inspired the creation of better machines and systems [17], using the principles and practice of control theory borrowed from science, strengthened by mathematics and successfully integrated with design and

manufacturing, by engineers. An early convergence [33] of control theory with communication may be found in the 1948 treatise "Cybernetics" by Norbert Wiener [34] (who may have borrowed [35] the word "cybernétique" proposed by the French physicist and mathematician André-Marie Ampère [18] to design the then nonexistent science of process control).

In other examples of "divide and conquer," the theoretical duo "Alice and Bob" is at the core [36] of cryptography [37] as well as the game theoretic [38] approach [39] to "prisoner's dilemma" which has influenced business strategies [40] and now it is spilling over to knowledge graph (KG) [41] databases. The simple concept of a lone travelling salesman proposed by Euler in 1759 appears to have evolved [42] as the bread and butter of most optimization engines, which, when considered together with data and information, continues to improve decision support systems (DSS) in manufacturing, retail, transportation, logistics [43], and omnipresent supply chain [44] networks, almost in every vertical which uses DSS.

The purpose of these disparate examples are to emphasize the notion that there are fundamental units of activity or models or set(s) of patterns or certain basic behavioral criteria (for lack of a better descriptive term) that underlie most actions and reactions. When taken apart or sufficiently reduced, we may observe these as isolated units or patterns or models of rudimentary entities. When combined, these simple models/units/patterns/elements can generate an almost unlimited variety of system behaviors observed on grand scales. When viewing the massive scale of systems from the "top," it may be quite counterintuitive to imagine that the observed manifestations are due to a few or a relatively small group of universal "truths" which we refer to as models, units, rules, logic, patterns, elements, or behaviors. To further illustrate this perspective, consider petals (flowers), pineapple (fruit), and pyramids. The variation between and within these three very different examples may boil down to Fibonacci [45] numbers, fractal [46] dimensions, and the Golden [47] Ratio [48] in some form, or the other. In another vein, the number, eight, seems to be central to atoms (octet) and an integral part of the Standard Model in physics (octonions [49]). Number 8 is revered by the Chinese due to its link with words synonymous with wealth and fortune (fa).

If one is still unconvinced and remain skeptical that small sets of underlying elements, generally, may be responsible, albeit in part, for the "big things" we consider diverse, then the "killer" example is that of nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), made up of only five subunits or molecules (adenine, guanine, cytosine, thymine, and uracil). DNA and RNA serve as the blueprint for all humans, animals, plants, bacteria, and viruses that may ever exist. The infinite diversity of multicellular [50] and unicellular organisms, whose creation is instructed by a *combination of these five molecules* in DNA and

RNA, may vastly exceed 5×10^{30} (5,000,000,000,000,000,000,000,000,000,000 [51]). The known exception to the DNA–RNA dogma may be the case for prions [52] which uses proteins [53] as the transmissible macromolecule.

Parallel examples can be drawn from physical sciences. Large-scale system behaviors can be reduced and mapped to simple models. Combination of these simple models, with widely different microscopic details, applies to, and generates, a large set of possible systems [54] and system of systems. Another example of "hidden complementarities" emerged from cryptic mathematical bridge embedded in natural sciences. It is now established that eigenvectors may be computed [55] using information about eigenvalues. Students are still taught that eigenvectors and eigenvalues are independent and must be calculated separately starting from rows and columns of the matrix. Mathematicians authored papers in related fields [56] yet none "connected the dots" between eigenvectors and eigenvalues. The insight that eigenvalues of the minor matrix encode hidden information may not be entirely new [57] but was neither understood nor articulated. The relationship of centuries-old mathematical objects [58] ultimately came from physicists. Nature inspires mathematical thinking because mathematics thrives when connected to nature. Grasping these connections enables humans to create tools to mimic nature (bio-mimicry).

2.3 Problem Space: Are We Asking the Correct Questions?

The lengthy and winding preface is presented to substantiate the opinion that there may be a disconnect between the volume of data we have generated as a result of the "information age" versus the lackluster gains in performance, as estimated by the productivity [59] index. We may have 2.7 zettabytes [20] (2.7 billion terabytes) of data, but some estimates claim as much as 33 zettabytes [60] of data, at hand (2018). It is projected to reach 175 zettabytes circa 2025.

The deluge of data as a result of "information technology" is far greater in magnitude than the diffusion of electricity [61] a century ago. Productivity increases due to the introduction of electricity and IT offers economic parallels [62] but based on the magnitude of change, the shortfall (in productivity) cannot be brushed aside by attributing the blame to mismeasurement explanations [63] for the sluggish [64] pace. Extrapolating measurements using the tools of classical productivity [65] to determine the impact of IT and influence of data is certainly fraught with problems [66], yet the incongruencies alone cannot explain the shrinkage. In socioeconomic terms, there is a growing chasm between IT and data/information versus productivity, improvement in quality of life, labor, compensation [67], and standard of living.

Despite trillions of dollars invested in data, digital transformation and other IT tools [68] (big data, AI, blockchain), the perforated ROI [69] increasingly points to massive [70] waste. One reason for this "waste" may be due to use of models of data where errors are aggregated under a generalized [71] form or variations [72] of the normal (homoskedastic) distribution. Heteroskedasticity was addressed [73] using ARCH [74] (autoregressive conditional heteroskedasticity [75]) and GARCH [76] models [77] (generalized ARCH). The use [78] of these proven techniques [79] for time series data (for example, sensor data showing water temperature in marine aquaponics [80] or cold chain [81] temperature log of vaccine package during transportation) in financial [82] econometrics [83] may be extended. Applications in predictive [84] modeling and forecasting [85] techniques may wish to adopt these econometric tools (GARCH) as a standard, whenever time series data are used (for example, supply chain [86] management, sensor data in health), but only ***if*** there is sufficient data (volume) to meet the statistical rigor necessary for successful error correction.

Perhaps, it is best to limit the postmortem analysis of IT failures, snake-oil sales of AI [87], and other debacles. Let us observe from this discussion that in the domain of data, and extraction of value from data to inform decisions and the tools necessary for ***meaningful*** transformation of data to inform decisions may benefit from <u>**re**</u>-viewing the processes and technologies with "new" eyes. We must ask, often, if we are pursuing the correct questions, if the tools are appropriate and rigorous. The productivity gap and reports of corporate waste are "sign-posts" on the road ahead, except that the signage is in the incorrect direction, with respect to the intended destination, that is, profit and performance.

2.4 Solutions Approach: The Elusive Quest to Build Bridges Between Data and Decisions

There are no novel proposed solutions in this essay, only new commentary about *approaches* to solutions. The violent discord between volume of data versus veracity of decisions appears to be one prominent reason why the productivity gap may widen to form a chasm. The "background" section discussed how the reductionist approach points to simple models or underlying units or key elements, which, when combined, in some form, by some rules or logic, may generate large-scale systems.

Data models [88] for DBMS are very different from ***models in data***. Pattern mining [89] from data [90] is a time-tested tool. What new features can we uncover or learn about data, from patterns? What simpler models or elements are cryptic in data? Are these the correct questions? *If* there are simpler models or patterns in some types of data, can we justify extrapolating these models and *patterns* as a

general feature of the data? The failure to accept and curate data which may be void of information is of critical importance. The contextual understanding of this issue appears to be uncommon and tools for semantic data curation are nonexistent. Although we have been mining for patterns and models (clustering, classification, categorization, and principal component analysis) for decades, why have not we found simpler models or patterns, yet? Are we using the wrong tools or wrong approaches or looking at wrong places? How rational are we in our search for these general/simple models in view of the fact that models of data from retail or manufacturing or health clinics *should* be quite different? Is model building by humans an irrational approach since humans are innate, irrational organisms endowed with sweeping bias?

Thus, the lowest common denominator of general models/patterns may not be an ingredient for building that experimental "thought" bridge. Increasing volume of data could help GARCH tools but it is a slippery slope in terms of data quality with respect to *informing* DSS and/or the veracity of decisions (output). Data models/patterns as denominators from grocery shopping or dry wall manufacturing or mental health clinics *are* different. In lieu of "universal" common denominators, we may create repertoires of domain-specific common denominators. A comparative analysis between common denominators of retail grocery shopping model from Boston vs. Beijing may reveal the spectrum of nutritional behaviors. If linked to eating habits, perhaps we can extrapolate its *influence* on health/mental health. As this suggestion reveals, we may be able to explore very tiny subsets of models.

Domain-specific denominator models (DSDM) are not new. It requires an infrastructure approach to data analytics which needs multitalented teams to explore almost every cross section and combination of very large volumes of data, from specific domains, to identify obvious correlations as well as unknown/nonobvious relationships. If there is any doubt about the quality of the raw data, then quality control may mandate data curation. The latter alone, makes the task exponentially complex. Curation may introduce reasonable doubt in evaluating any outcome because the possibility exists that curation algorithms and associated processes were error-prone or untrustworthy (post-curation jitters).

Another demerit for DSDM and the idea of denominator models, in general, may be rooted in the "apples vs oranges" dilemma. Denominator models that underlie science and engineering systems are guided by natural laws, deemed *rational*. The quest for denominator models in data (retail, finance, supply chain, health, and agriculture) are influenced, infected, and corrupted by irrational [91] human behavior. Rational models of irrational behavior [92] may coexist elsewhere but remains elusive for data science due to volatility and the vast *spectrum of irrationality* that may be introduced in data by human interference.

Perhaps, the concept of DSDM, ignoring its obvious caveats, may be applied to select domains for specific purposes, for example, healthcare, where deliberate human interference to introduce errors in data is a criminal offense. Case-specific

model building, and pattern recognition, may benefit from machine learning (ML) approaches. The latter fueled a plethora of false [93] claims but real success is still a *work in progress* because the bridge between data and decisions will be perpetually *under construction*. Productivity gap and corporate waste are indicators that existing approaches (see Figure 2.5) are flawed, failing, or have [94] failed. We need new roads. The boundary of our thought horizon "map" is in Figure 2.5. The tools are incremental variations [95] garnished with gobbledygook alphabet soup. Unable [96] to create any breakthrough, the return of seasonal "winters of AI" indicates the struggle to shed new light in this field since the grand edification [97] during the 1950s. Unable to cope with data challenges, hard facts [85], and difficult progress, the field offered a perfect segue for con artists and hustlers to inculcate falsehoods and deceive [98] the market. ML was substituted [99] by mindless drivel from ephemeral captains of industry and generated hype [100] from corporate [101] marketing machines as well as greedy academics.

2.5 Avoid This Space: The Deception Space

Data consumers have been led astray by vacuous buzz words manufactured mostly by consulting groups. Part of the productivity gap may be due to fake news, propaganda [94], and glib strategy from smug consultants to coerce large contracts with cryptic "billable hours" to help "monetize" false promises due to "big" data, fabricated [102] claims [103] of "intelligence" in artificial intelligence (AI) [104], and deliberately conniving misrepresentations [105] of "blockchain" as a panacea [106] for all problems [107] including basic food safety and security. Callous and myopic funding agencies invested billions in academic [108] industry partnerships to fuel banal R&D efforts orchestrated by corporate collusion [109] and perhaps [110] criminal [111] practices. Abominable predatory practices on display in Africa are disguised under the "smart cities" marketing campaign to mayors of ***African cities, which cannot even provide clean drinking water to its residents.*** Vultures from the industry [112] are selling mayors of African cities surveillance technology and AI in the name of cameras for smart city safety and security. These behemoths are cognizant as to how autocrats use data as an ammunition to plan and justify abuse of its citizens, through algorithms of repression.

2.6 Explore the Solution Space: Necessary to Ask Questions That May Not Have Answers, Yet

Uploading data from nodes along a variety of supply chains is an enormous undertaking given trillions of interconnected processes and billions of nodes with extraordinarily diverse categories of potential data streams, with different security

Figure 2.5 It appears that we have been mining for patterns and other simpler models (such as clustering, classification, categorization, regression, and principal component analysis). But, have we found a set(s) of simpler models or patterns, yet, to test the concept of domain-specific denominator models (DSDM)?

mandates, for example, (i) sensor data about heavy metal (mercury) contamination in water used for irrigation, (ii) near real-time respiratory rate (RR) of patient with chronic obstructive pulmonary disease (COPD) under remote monitoring telemedicine in rural nursing home, and (iii) automated check-out scan data from retail grocery store sales, of fast-moving consumer goods, contracted for replenishment (penalty for out of stock) under vendor managed inventory (VMI). The *e-tail* revolution is creative supply chain optimization and reducing retail information asymmetry.

Transforming data and data analytics to inform decision support for *small* cross-sections of examples cited, here, may be theoretically easy in "*power point*" diagrams which "connects" nodes and integrate decision feedback to optimize processes, *using pixels*. The reality may be different. Aggregating data from various nodes, sub-nodes, devices, and processes, on a platform, to enable collective evaluation of dependencies, which could influence outcomes/decisions, may be beneficial, or germane for certain domains, for example, healthcare [113] and clinical [114] environments where patient safety [115] must be of paramount importance.

Agreement on any *one* standard platform is unlikely to succeed. But an anastomosis of platforms is probably rational, if interoperable. An open platform of platforms with secure, selective, interoperable data exchange, between platforms, may be valuable. Synthesis and convergence of data acquisition and analytics begins to catalyze information flow, decision support, and ***meaningful*** [116] use of data [117]. This suggestion is a few decades old, but still far from practice. The drive to connect data was accelerated by the introduction of the concept of the IoT [118]. Platform [119] efforts [120] are addressing data [121] upload from devices and sensors [122] but nowhere near a turn-key implementation (Figure 2.6).

2.7   Solution Economy: Will We Ever Get There?

There are not any silver bullets and one shoe does not fit [124] all. If we focus on the data to decision process, alone, in any vertical or domain, the variations of analysis and analytics may be astronomical. Initial investments necessary for these endeavors almost guarantee that the extracted value from data (and relevant information) may not be democratized or made functionally available to those who cannot pay the high cost. In principle, the outcome from data to decisions, when appropriate, may be sufficiently distributed and democratized to provide value for communities under economic constraints. Any ***meaningful*** solution, therefore, is not a scientific or engineering outcome, alone, but must be combined with the economics of technology [125] which must be a catalyst for implementation and adoption by the masses, if transaction costs [126] can be sustained by the community of users, in less affluent geographies.
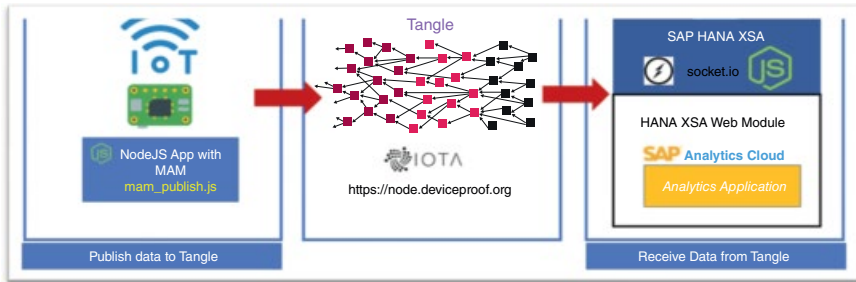
Figure 2.6 National and international consortiums, in partnership with large and small software companies, are addressing data acquisition and aggregation. In this example, Tangle appears to be a data aggregation platform (example shows data from a temperature sensor) which can serve data analytics engines to extract information (if data contain information). Replication of "Tangle" for various verticals (retail, health, and logistics) and the ability to use open-data distribution services [123] may facilitate interoperability between data "holding" services like Tangle. When coupled with supply chain track and trace systems, a retail store (Target, Tesco, Metro, and Ahold) can use Tangle data to inform a customer that the One-Touch blood glucose testing strip (healthcare product manufactured by J&J) will arrive at the store (third-party logistics provider and distribution transportation service) on Monday by 2 a.m. and placed on store shelf by 530 a.m. (retail store replenishment planning) or delivered to the customer on Tuesday before 9 a.m. (online fulfillment services). Any data can be uploaded/downloaded from Tangle.

The economic principle for impoverished environments may be rooted in micro-finance [127] and micro-payments [128] with *low* transaction costs (*the downside*: misinformation [129] can be propagated and disseminated at *low* cost, too). By eliminating classical "product sales," the focus shifts to delivery of "service" which is a *package* of the product plus other resources (retail mobile banking, infrastructure, telecommunications, cybersecurity [130], security [131], and customer service). Users pay (pennies) *only when they use the **service***. Pay-a-penny-per-use or pay-a-price-per-unit [132] (PAPPU) is a *metaphor* for economic instruments which may lower the barrier to entry into markets with billions of users.

The economic incentive for democratization of data is the potential to unleash and create new markets for data, information, and decision support, for billions of new consumers (users). The **reward** in the lucrative service economy model depends on harvesting the economies of scale where each user (market of billions) may pay one or more "pennies" (micro-payment for pay-per-use services). The **risk** in the service economy is the collection of that "penny" (per use) at the last step of the *seamless* service delivery process, if the user is satisfied with the quality of service (QoS) metrics. The plethora of partners necessary to create and sustain the ecosystem to deliver the *seamless* service is a herculean task. Sharing a fraction

of that "penny" with the partners in the ecosystem is not a trivial challenge. If the QoS delivery metrics suffer due to poor performance of any one partner (component), the end-user "penny" may be unpaid if the QoS metric fails to reach a predetermined value (time, duration, speed, rate, and volume). The inability of one provider (weakest link) in the service supply chain can be financially detrimental to all other supply chain partners due to loss of that penny, *albeit, only for that transaction* (unless the partner has a chronic problem, then, it must be excluded from the ecosystem and the entire value network [133]). Delivery of service is a real-time convergence of operations management which includes (but is not limited to) multiple value chains which must integrate [134] the physical supply chain and the financial supply chain with the service supply chain and customer relationship management (brand expectation).

Determining the cost of execution, to deploy the example in Figure 2.6, may be one way to study feasibility. Simulating models to explore financial engineering of "what if" scenarios, may project the potential for adoption of services in the context of various economies of scale and PAPPU models. The reward for unchaining the economics of technology is in adoption, by the next billion users.

2.8 Is This Faux Naïveté in Its Purest Distillate?

Decision scientists must build a compass to help extract value from data. One compass will not suffice to guide domain-specificity. Existing tools may limp along with *snail-ish* advances (Figure 2.5) yet it may remain inaccessible to the masses because the tools may not be feasible for mass adoption. The struggle to transform data into information is still in quest of a Renaissance.

The path from *data-informed* to *information-informed* to *knowledge-informed* decision remains amorphous. Transforming information to knowledge is in the realm of unknown unknowns. Making sense of data is handicapped due to (i) an apparently insurmountable semantic barrier, (ii) scarcity of tools to facilitate location-aware and context-aware discovery of data at the edge or point of use, and (iii) lack of standards and interoperability between objects, platforms, and devices for data and analytics sharing.

Users in less affluent nations may not want to idle away while the architects of Renaissance are still in short supply. In the near term, it is necessary that we continue to work on dissemination of data which can deliver at least some value, sooner, rather than later. Decision support based on sensor data analytics may provide economic benefits [135] and incentives, if we can share the digital dividends with the masses, for example, in health [26] and agriculture, including every facet of food, required daily, globally.

Tangle, a tool [136] to share sensor data using masked authenticated messaging [137] (MAM) may offer hope. Can nano-payments for sensor data address some of the feasibility challenges [138] and pave the way for human-centric economy of things [139] using IoT as a design metaphor? SNAPS [140] is a tiny step in that general direction: distributing low-cost tools to enable data-informed decision support (DIDAS) for less complex problems. Assuming the Pareto principle holds true, perhaps 80% of the problems may be addressed, and even resolved, with simple tools to deliver solutions as a ***service***, at the right-time, at the point of use.

2.9 Reality Check: Data Fusion

The inflated view of the sensor-based economy [141] is carefully [142] crafted [143] to create [144] new markets [145] and momentum [146] for sales [147] of sensors and data services [148] aimed to amplify the IoT [149] hype to fortify the deception game. It is promoting the desired effect by spawning mass hysteria and skillfully obfuscating the hard facts which then paves the ground for hordes of consultants to act as "trusted advisors" to make sense of this "revolution" which is supposedly going to change the future of work, life, and living. One glaring outcome of delusional [150] propaganda [151] is the near-trillion dollar [48] waste related to investment in technology with a failed ROI. Trillions of sensors and devices that *could* connect to the Internet (basis for the cosmic scale of IoT) is due to the scale of unique identification [152] made possible by adopting a 128-bit structure in the internet protocol. The unique address spaces in IPv6 [153] is ***29 orders of magnitude higher*** than IPv4 if one compares [154] 4.3×10^9 address spaces for the 64-bit IPv4 versus 3.4×10^{38} unique address spaces for the 128-bit IPv6. New possibilities [155] and applications [156] may arise due to the flexibility of IPv6 to directly connect to the Internet (rather than sub-nesting under/via gateway nodes).

The difference between promise and perils in deploying the concept of IoT as a design metaphor is rooted in grasping the difference between connectivity, discovery, and actionable insight. Just because something is *connected* does not mean value emerges, automatically, without a ***connected ecosystem***. If a visitor's tablet can discover the printer in an office and use it to print a meeting agenda, then we have extracted some type of value between the connectivity of the tablet and the printer, which were able to "discover" each other, and that discovery enabled the gain in efficiency (printing the agenda). In its basic form, this is an example of very simple data fusion which leads to an actionable output and provides "information" for the meeting attendees in terms of the printed agenda. Connecting trillions of entities to the Internet is futile unless discovery and data fusion enable semantically meaningful extraction of data to move up the DIKW [157] value chain [158] where data precede information, knowledge, and wisdom.

The PEAS paradigm resembles OODA (Figure 2.8) because "observations" refer to scanning (sensing) the environment and "orientation" informs the image of the environment by encapsulating both descriptive and predictive analytics ("decide" includes prescriptive analytics). The integration of data fusion and analytics with ABS (Agent-Based System) is critical in the era of IoT. The networked society faces a deluge [159] of data, yet the human ability to deal with data, analytics, and synthesis of information may be inefficient. How can devices discover data and facilitate processes without intervention by humans? Automated on/off action taken by a domestic thermostat and HVAC based on temperature sensors may be quite primitive when considering autonomous objects in air (UAV), land, and water.

Raw sensor data unless discovered and combined with "perceptions" from the environment, may be context-deprived and over/under utilized, which lowers the value of the data with respect to the desired goals. The perception from the environment is not unique but a "learning" task for the system. It may reuse the experience (learning), when relevant and appropriate, at a different instance (Figure 2.7). Can this "learning" become mobility-enabled and "teach" other devices, for example, by transmitting a *tutor* virion to another computer or drive or system? Can this device communicate in natural language (NL) and/or respond/understand the semantics in human queries?

Taken together, unleashing the value of data may require coordination of ABS in every facet of our interaction with machines, objects, and processes which may benefit from feedback. ABS is an old [160] concept [161] but resistant to succinct definition [162] because agent activity must remain agile and adapt to the operating objective (PEAS) and problem context (OODA). Equation-based models
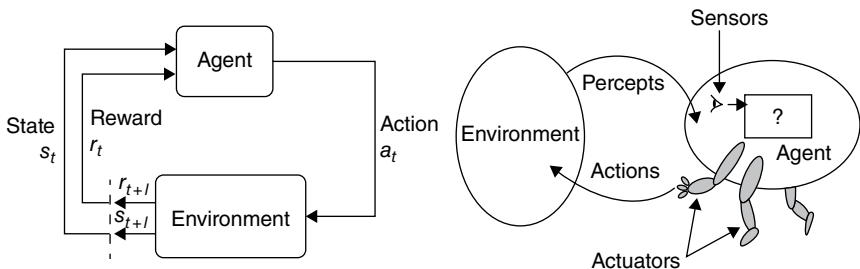

Figure 2.7 SNAPS is one layer in a proposed analytics platform (layer cake) which consists of a portfolio of tools aligned with the concept of PEAS [23], a mnemonic borrowed from agent-based systems (ABS) (**R**), to address systems performance through convergence of percepts, environment, actuators, and sensors. *Source:* Modified from Russell, Stuart and Norvig, Peter (2010) *Artificial Intelligence: A Modern Approach*. 3rd ed, Prentice Hall. (**L**) Reinforcement Learning [24] (Figure 2.5), a ML technique, compared with PEAS. *Source:* Modified from Sutton, Richard S., and Andrew G. Barto, 2018. *Reinforcement Learning: An Introduction*. 2nd edition, MIT Press.
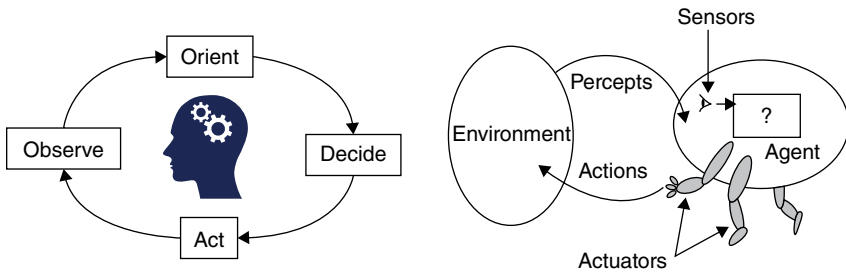
Figure 2.8 PEAS, a mnemonic borrowed from agent-based systems (**R**), addresses systems' performance through convergence of percepts, environment, actuators, and sensors. The OODA [25] loop (**L**) and PEAS [26] contribute to advance DIKW (data, information, knowledge, and wisdom), which begins with data fusion [27]. *Sources*: Angerman, W.S. (2004). Coming full circle with boyd's ooda loop ideas: an analysis of innovation diffusion and evolution (Thesis); Kamenetz, A. (2011). *Esther duflo bribes India's poor to health. Fast Company* (8 August); Castanedo, F. (2013). *A Review of Data Fusion Techniques.* The Scientific World Journal.

(EBM) create rigid, hard-coded software. ABS design induces agility, may enable "drag & drop" variant configuration to adjust (on-demand) to volatility, uncertainty, and ambiguity, inherent in most environments. In the context of democratization of data and benefits for the masses, agents can be highly personalized and "belong" to people, for example, personal agents, as discussed [163] elsewhere, with respect to cybersecurity. A similar modus operandi can be adopted for other use-cases where data fusion [164] can be dynamic and composable (composed when necessary, depends on context) not only for use-cases but also for individual user-specific case/application (healthcare treatment plan) (Table 2.1).

The role of software agents to "discover" and then determine which data and/or data fusion may be meaningful or relevant (user-specific), is an old idea, still waiting to be effectively applied. Connecting data must be contextual. The *established* contextual relationship must be discovered and "understood" by agents or group of agents. Another old idea is to *pre-establish* the context based on KGs (Knowledge Graphs). The thinking that W3C standard resource description framework [166] (RDF) triples are the solution for KGs is incomplete. This myopia, is, in part, one reason why the semantic web [167] failed to flourish. The brilliant idea of representing subject-predicate-object (SPO) as a relational RDF graph is certainly useful and applicable in many instances but the approach *bites the dust* when the reductionist 1:1 granular relationship fails to represent reality. The latter is painfully obvious, especially in medicine and healthcare, where the rigidity of the RDF standard structure and RDF schema may be an anathema. The "force-fitting" of RDF to healthcare applications [168] oversimplifies scenarios to the point where it may, inadvertently, introduce errors, simply due to exclusion, which may prove to be fatal.

Table 2.1 Generally, agents are computational entities (software) designed to perform specific tasks, autonomously.

Agent characteristics	Definition
Autonomy	Operates without the direct intervention of humans or others
Sociability	Interacts with other agents, that is, communicates with external environment such as sensors, fusion systems, and human operators
Reactivity	Perceives its environment and responds in a timely fashion
Pro-activity	Exhibits goal-directed behavior by taking the initiative
Learnability	Learns from the environment over time to adjust knowledge and beliefs
Mobility	Moves with code to a node where data reside
Anthromorphicity	Externally behaves like human

Agents embedded in devices (sensors) may have logic capabilities to perform artificial reasoning tasks (ART) and/or optimization [165] in multi-agent systems (MAS).
Source: Based on Rafferty, E.R.S., et al. (2019). Seeking the optimal schedule for chickenpox vaccination in Canada: using an agent-based model to explore the impact of dose timing, coverage and waning of immunity on disease outcomes. *Vaccine* (November 2019).

One proposal suggests adapting [169] RDF by creating relationships between sets/subsets (rather than points and vertices as in classical SPO) using the set theoretic [170] approach. It is easy to grasp why "set" of symptoms and potential set of causes may make more sense in medicine and healthcare. The overlapping (Venn diagram) subset of relationships may be indicative of likely causes for symptoms. Generic symptoms, for example, fever, can be due to a plethora of causes and why a rigid 1:1 relationship in RDF could turn lethal in healthcare applications. The finer granularity of RDF is a disadvantage, yet it is key to merging attribute lists about an entity sourced from different data sources. The latter enables better search and discovery across diverse domains, the hallmark of globalization of enterprise systems.

An even older idea [171] which is recently [172] enjoying scientific [173] as well as public attention [174] is the labeled property graph (LPG). It is suitable for use-cases which may be focused on providing stores for single applications and single organizations, such as, DSDM. LPG proponents are less committed to standardization, interoperability, and sharing. It is in contrast to the W3C ethos and RDF which favors standardization, interoperability, and sharing, which makes it useful in discovery using graph pattern searches. Optimization of local (domain specific)
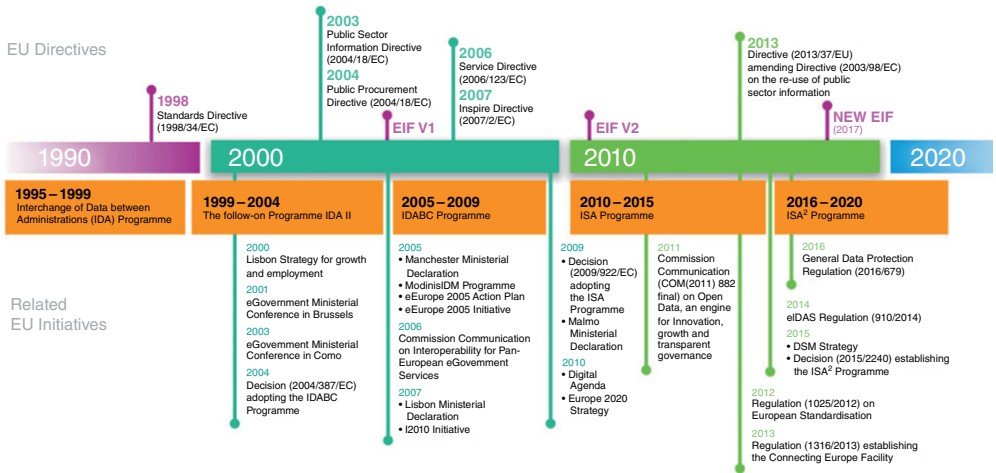
searches [175] using graph-traversal algorithms [176] are better suited for property graph (PG) databases. KG networks embedded between sets/subsets may give rise to amorphous "linked" clouds, which could be industry [177] specific and may be domain specific [178] as well as user specific. Imagine if data from each patient could be used by an ***automated knowledge graph engine*** to create precision, patient-specific, personalized KGs. Extracting relationships and contextualizing the relevance of symptoms may improve the accuracy of diagnosis. When viewing KGs in a population study (epidemiology), it may be easier to detect outlier events or cases that did not fit the expected patterns.

Therefore, DSDM may be represented as domain-/user-specific KG networks. Agents may be invaluable in working within this environment to discover relationships and contexts (specificity reduces search space), as well as discover data sources, and perhaps, based on embedded logic, decide whether the features or attributes calls for data fusion.

For any agent-based approach to succeed, it is critical that the agent framework and standards are interoperable with the KG network and the data domains where the agent is searching. The opposing tendencies of RDF vs LPG in terms of standardization, interoperability, and sharing may limit agent-mediated "cross-investigation" of domains, discovery, and data. Therefore, it begs to question the expectation that one agent must perform in all domains. Perhaps, the success of agent search and discovery depends on semantically annotated structured data. The latter depends on ontological structure. W3C proposed [179] OWL standard web ontology language [180] and recent variations (VOWL [181]) may contribute to interoperability. The old idea of Internationalized Resource Identifier (IRI) as a complement to the Uniform Resource Identifier (URI) [182] to identify resources (to facilitate discovery) is a valid principle but yet to be adopted in practice. The plethora of old ideas (referred here) suggests that the value of these ideas may have to be revisited. We need new "blood" and new "eyes" to reimagine new ways to address interoperability. However, in reality, today, on top of this wobbly incompatible infrastructure, we are layering the "snake oil AI" and unleashing an incorrigible torrent of half-truths.

Timeline 2.1   Long march of graph-related [183] specifications. Is standardization a sufficient solution?

1998
Standards Directive
(1998/34/EC)

2003
Public Sector
Information Directive
(2004/18/EC)
2004
Public Procurement
Directive (2004/18/EC)

2006
Service Directive
(2006/123/EC)
2007
Inspire Directive
(2007/2/EC)

EIF V1

EIF V2

2013
Directive (2013/37/EU)
amending Directive (2003/98/EC)
on the re-use of public
sector information

NEW EIF
(2017)

1990	2000		2010	2020

1995 – 1999
Interchange of Data between
Administrations (IDA) Programme

1999 – 2004
The follow-on Programme IDA II

2005 – 2009
IDABC Programme

2010 – 2015
ISA Programme

2016 – 2020
ISA² Programme

Related
EU Initiatives

2000
Lisbon Strategy for growth
and employment
2001
eGovernment Ministerial
Conference in Brussels
2003
eGovernment Ministerial
Conference in Como
2004
Decision (2004/387/EC)
adopting the IDABC
Programme

2005
• Manchester Ministerial
  Declaration
• ModinisIDM Programme
• eEurope 2005 Action Plan
• eEurope 2005 Initiative
2006
Commission Communication
on Interoperability for Pan-
European eGovernment
Services
2007
• Lisbon Ministerial
  Declaration
• I2010 Initiative

2009
• Decision
  (2009/922/EC)
  adopting
  the ISA
  Programme
• Malmo
  Ministerial
  Declaration
2010
• Digital
  Agenda
• Europe 2020
  Strategy

2011
Commission
Communication
(COM(2011) 882
final) on Open
Data, an engine
for Innovation,
growth and
transparent
governance

2016
General Data Protection
Regulation (2016/679)

2014
eIDAS Regulation (910/2014)
2015
• DSM Strategy
• Decision (2015/2240) establishing
  the ISA² Programme

2012
Regulation (1025/2012) on
European Standardisation
2013
Regulation (1316/2013) establishing
the Connecting Europe Facility

Timeline 2.2 EU's elusive [184] quest for interoperability: is 30 years not enough? *Source:* Carbone, L., et al. (2017) State of play of interoperability: report 2016.

2.10 "Double A" Perspective of Data and Tools vs. The Hypothetical Porous Pareto (80/20) Partition

Africa (>1 billion) and Asia (>4 billion), if combined, may soon represent 80% of the world's population. Global corporations view this "80%" world as a "market" which promises new markets, new customers, and new wave of consumerism. It has little to do with lifting the lives of people. Discussion about the physics and mathematics of data, therefore, is *a tempest in a tea cup*. For ~7 billion people, the trials and tribulations of data and data analytics, we have discussed here, can be dismissed with an eye roll. It is useless for pressing daily applications for ~7 billion of the ~7.8 billion people in the world.

Thus far, what we have discussed, on one hand, may be an exploration of the tessellated facets in our search for meaning, and on the other hand, it is a discussion which may find parallels with the "six blind men and the elephant" syndrome [185] apparently divorced from complementarity [186] or synergy. It is as if the "commerce" from 20% of the global population, relevant or not, is thrust upon the remainder of the world market. In 80% of the cases for 80% of the global population, the daily decisions about FEWSH (bare necessities of life: food, energy, water, sanitation, healthcare) do not require AI, ML algorithms, or optimization of "state space" for hundreds of variables (Figures 2.9 and 2.10).

In 80% of the cases for 80% of the global population, the daily decisions about FEWSH require data, ***small data***, data in near real time, and data that impact and enhance the user experience. In that context, the clamor about *democratization of data* is tantamount to chest-thumping. The data in these use-cases may be related to a subset of FEWSH (food, farm, agriculture, water, healthcare). If the tools are there to acquire these data, then the data are available. Therefore, is democratization really an issue? Is it a politically correct word that the 20% world prefers to use as a hand-waving advocacy of problems that are divorced from reality on the ground? Is "democratization" a "theme" song for advocacy groups in OECD nations who are displaying the symptoms of the *six blind men and elephant* syndrome?

Table 2.2 offers a glimpse of one problem in healthcare. The tools to acquire the data are in short supply. Measuring the risk of osteoporosis is a prerequisite for prevention and treatment, if affordable. Arm-chair "scenarios" of medical IoT will want to connect DEXA (DXA) scan data with sales of milk and exposure to sunlight as a "wellness" indicator. From the "double A" perspective, it may be a futile "power point" exercise because milk *may not be available* for the age group [200] generally at high risk of osteoporosis in the AA nations. In most parts of Africa and Asia, there is an opulence of sunlight.

Just because there is an "IoT" scenario, does not mean it is worthwhile or valid for users in "double A" nations. Just because there are data, does not mean there is information. Can we reduce incidence rates of osteoporosis simply by adding
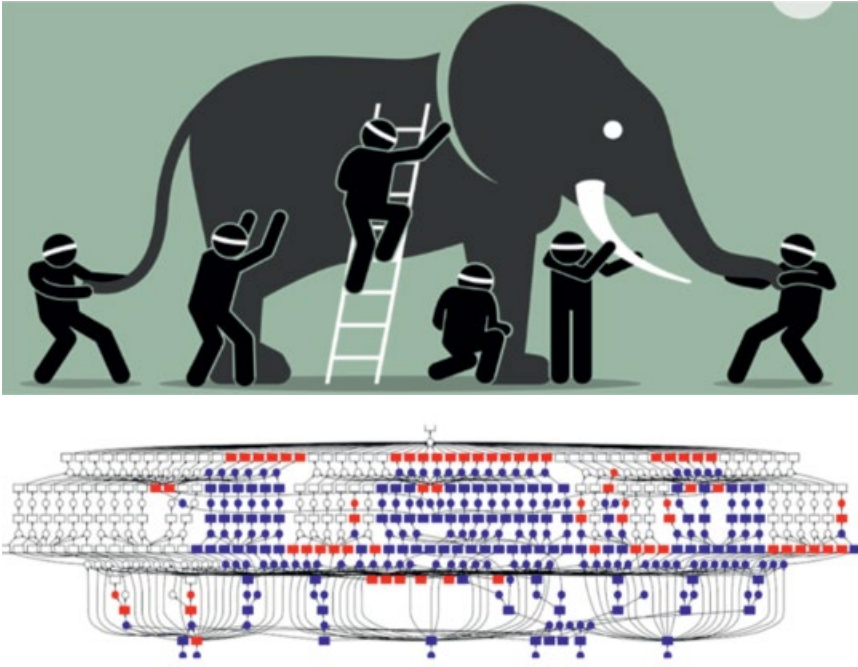
Figure 2.9 Blind men and elephant. Each man guesses his own part of the elephant but blinded by hype [187] they cannot perceive the "whole" elephant. A metaphor for focus on parts, which occludes the system. Cartoon (bottom): penchant for decision trees by *power-point* rather than search for low hanging fruits. *Source:* Modified from Morrison, A. (2019). Is data science/machine learning/AI overhyped right now?

more DXA machines per capita? A recent (2013) study using seven national electronic healthcare records (EHR) databases revealed that Denmark (14.2 DXA units per million) showed age- and sex-standardized incidence rates (IRs) of hip/femur fractures 2× higher than those observed in the United Kingdom (8.2 DXA units/million), Netherlands (10.7 DXA units/million), and Spain (8.4 DXA units/million), while Germany (21.1 DXA units/million) yielded IRs in the middle range (Table 2.3).

2.11 Conundrums

On one extreme we have presented sophisticated ideas for making sense of data. On the other hand, we doubt whether the toothless call for *democratization of data* from the affluent 20% of the world can help to lift the lives of people on the other side of the porous Pareto partition (80% of the world). It is not a true "Pareto"
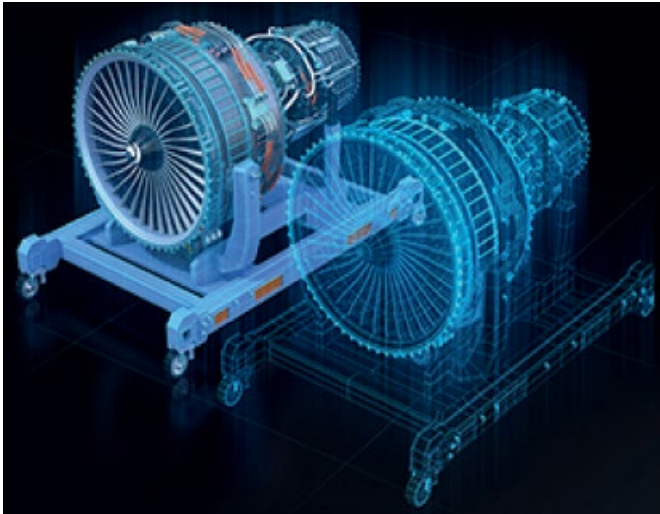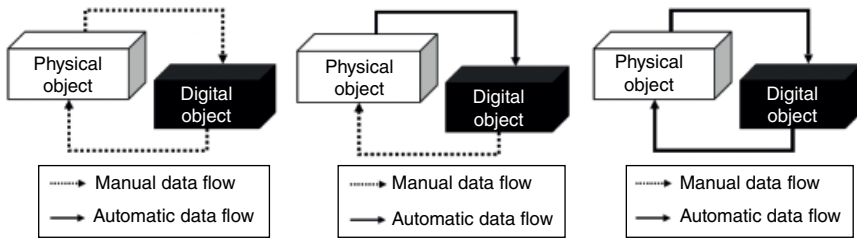
Figure 2.10 Digital Duplicate (left), Digital Shadow [188] (center), Digital Proxy (right), and Digital Twin (bottom) are variations of digital models of physical objects, integrated with data flow. But, do we know if it is ***meaningful*** for data-related needs for 80% of the world? It is unlikely to be solved by Digital Twins [189] or flamboyant gimmicks peddled by fake pundits on the pages [190] of *Forbes*. However, the R&D related to these tools may trickle through the "pores" from the 20% side of the partition to the other side (80%) of the *porous* Pareto [191] partition and occasionally [192] may be helpful. *Source:* Modified from Fuller, A., et al. (2019). Digital twin: enabling technology, challenges and open research. October 2019.

scenario, but the 80/20 nature of this problem evokes the Pareto principle as an analogy, hence, Pareto partition. The R&D outcome of the 20% may contribute certain elements to the 80% side. Thus, the "partition" is a metaphorical porous membrane, with bidirectional porosity.

But operations must be contextual. For example, is it necessary to deal with data and data models in this scholastic [203] manner (Figure 2.11) for all problems? The 80/20 global partition may be prominent in agriculture, healthcare, and energy. In case of the latter, what is the value of smart metering or load balancing

Table 2.2 Availability [193] of DEXA (dual energy X-ray absorptiometry) scan machines to measure bone mineral density (BMD), a fair prognosticator for osteoporosis.

The number and provision of central DXA units available in the EU27 (Data on reimbursement and waiting time [10])

	DXA units/million	Waiting time (d)	Cost (€)	Reimbursement
Austria	28.7	14	30 [11]	Yes
Belgium	53	14	34 [12]	Partial
Bulgaria	1.2	0	59 [9]	None
Cyprus	23.9	20	75 [9]	Yes (depending on income)
Czech Republic	5.2	40[*]	32 [9]	Yes
Denmark	14.6	30	187 [13]	Yes
Estonia	8.9	14	14 [14]	Yes
Finland	16.8	1	146 [28]	Yes
France	29.1	14	41 [9]	Yes (conditional)
Germany	21.1	0	36 [9]	Yes
Greece	37.5	11[*]	115 [9]	Yes
Hungary	6.0	15[*]	7 [15]	Yes
Ireland	10.0	140[*a]	99 [16]	Yes (conditional)
Italy	18.6	83[*]	81 [9]	Yes (conditional)
Latvia	4.9	10[*]	18 [29]	Yes
Lithuania	3.4	6[*]	28 [30]	No
Luxembourg	2.0	30	59 [31]	Yes
Malta	9.7	105[*]	184 [32]	Yes
Netherlands	10.7	14[*]	84 [17]	Yes
Poland	4.3	1	10 [9]	Yes (conditional)
Portugal	26.9	8	5 [33]	Yes
Romania	2.4	7	5 [34]	Yes
Slovakia	10.7	18[*]	32 [9]	Yes
Slovenia	27.1	11[*]	29 [35]	Yes (conditional)
Spain	8.4	105[*]	109 [9]	Yes
Sweden	10.0	60	152 [18]	Yes
UK	8.2	11[*]	51 [36]	Yes[a]

The European standard [194] is 11 (DEXA) DXA units/million. In an updated estimate, the poorest country in EU27 offers four machines/million [195] whereas Bulgaria's neighbor Greece boasts 37.5 DXA units/million. In comparison, Indonesia [196] has only 0.13, India [197] 0.18, and Morocco [198] 0.6 DXA units/million. For the health of the people in these nations, how can democratization of data lower their risk of osteoporosis? Are we asking the correct questions? Are we pursuing the wrong reasons? Are we arm-chair analysts helping the (BMD) medical device industry [199] accelerate their sales campaigns to AA nations? Can data provide relevant answers? *, average of range; a, data; d, days.
Source: Hernlund, E., et al. "Osteoporosis in the European Union: medical management, epidemiology and economic Burden: a report prepared in collaboration with the international osteoporosis foundation (IOF) and the European federation of pharmaceutical industry associations (EFPIA)." *Archives of Osteoporosis*, vol. 8, no. 1–2, 2013, p. 136. © 2013, Springer Nature.

Table 2.3 Plague of unethical profitability makes US pharmaceutical [201] business model in healthcare an abomination which is inappropriate for mimicry in any part of the world.

Drug	Prescribed for	UK price	US price	Price rise
NEXIUM per 20 mg tablet	Acid reflux	£0.66	£7.40	**1120%**
ACTIMMUNE, 12 vials	**Genetic diseases, osteopetrosis**	**£5,400**	**£42,990**	**800%**
DARAPRIM per tab	HIV, cancer, malaria patients	£2.30	£619	**26,900%**
NASONEX, 50 mg	**Nasal allergies**	**£7.68**	**£224**	**2900%**
CINRYZE, 2 vials	HAE, genetic disorder	£1,336	£3,645	**200%**
HARVONI per tab	**Hepatitis C virus**	**£464**	**£928**	**£270%**
SOVALDI per 400 mg tab	Hep C in children under 12	£416	£855	**200%**
DIAZEPAM, per tab	**Anxiety, relaxation, muscle spasms**	**£0.02**	**£3.05**	**15,200%**
OVEX, 100 mg tablet	Threadworm parasite	£2.54	£300	**11,800%**
LIPITOR, per 10 mg tab	**Statin**	**£0.46**	**£4.50**	**980%**
VIAGRA, per 25 mg tab	Male impotency	£4	£61	**1500%**
ZOCOR per 10 mg tab	**Statin**	**£0.64**	**£4.20**	**656%**
CYMBALTA per 30 mg capsule	Anti-depressant	£0.80	£9.48	**1200%**
EPIPEN, 300 mg	**Allergies**	**£52.90**	**£523**	**1000%**
HUMALOG INSULIN	Diabetes	£16.61	£215.30	**1300%**
HIP REPLACEMENT OPERATION		**£7,313**	**£26k-£37k**	**350%**
KNEE REPLACEMENT OPERATION		£6,315	£24,801	**390%**
CATARACT OPERATION		**£803**	**£5,780**	**720%**

Source: Dr James Nolan [202].

algorithms when there is not enough energy, at an affordable cost, to supply the basic tenets of economic growth? How many farmers in "80%" world can afford to use drone-on-demand [208] systems? Why should people from the majority sector (80%) need useless marketing tools [209] when daily healthcare for the less fortunate can solve a myriad of problems with *just-in-time little bits of data*, for example, daily blood glucose level from a diabetic (versus the *always-on* real-time
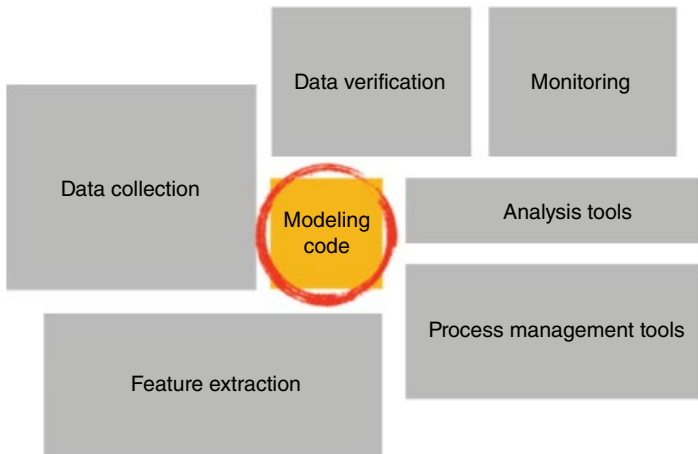
Figure 2.11 From monitoring an event to using the data to inform a decision, there are a plethora of steps [204] in the standard operating procedure (SOP) for the "20%" deploying data to drive decisions. However, irrespective of socioeconomic issues, in future, all aspects of feature selection and feature engineering may emerge as a pivotal or rate-limiting step in dealing with diverse data sources. In this context, automated feature extraction and other feature-related steps may be a very significant step. *In combination, automated feature engineering and automated knowledge graph engines may usher new dimensions in data and data analytics, if automated data curation could improve data quality* (Figure 2.12) [207]. *Source:* Mendez KM, Broadhurst DI, Reinke SN. The application of artificial neural networks in metabolomics: a historical perspective. *Metabolomics* 2019; 15(11):142. doi: 10.1007/s11306-019-1608-0.
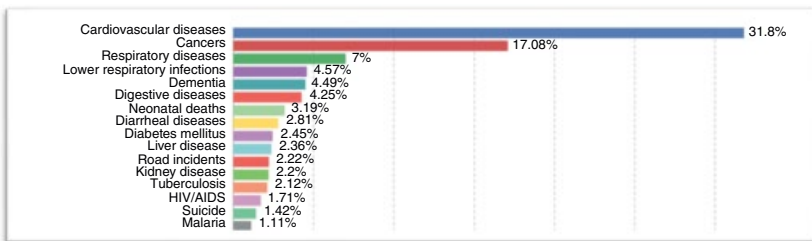
Figure 2.12 Share of deaths, by cause (2017) percent of total deaths [205]. Data refers to specific cause of death, which is distinguished from risk [206] factors for death (water and air pollution, diet, sanitation). *Source:* Ritchie, H. (2018). What do people die from? Licensed under CC BY 4.0

monitoring of blood glucose) or monitoring individuals for silent myocardial ischemia [210], a leading contributor to death. This is the debate where the economics of technology and its ***relevance to the community*** are crucial issues which may ***enable*** adoption or ***disable*** the dissemination of technology, which

could have contributed to economic growth, workforce development, and sustainable job creation.

Worldwide, Africa [212] accounts for 9 out of every 10 child deaths due to malaria, 9 out of every 10 child deaths due to AIDS, and for half of the world's child deaths due to diarrheal disease and pneumonia. More than one billion children are severely deprived of at least one of the essential goods and services they require to survive, grow, and develop [213] – these include nutrition, water, sanitation facilities, access to basic healthcare services, adequate shelter, education, and information. As a result, almost 9.2 million children under five die every year and 3.3 million babies are stillborn. Most of the 25,000 children under five that die each day are concentrated in the world's poorest countries in sub-Saharan Africa and South Asia. There, the child mortality rate is 29 times greater than in industrialized countries: 175 deaths per 1000 children compared with 6 per 1000 in industrialized countries.

These facts (paragraph above) and Figure 2.13 offer a vastly contrasting view to that of data tools and democratization of data as essential for lifting the lives of the people living on the majority side of the porous Pareto partition. Simple forms of small amount of data, sufficiently informing ordinary tasks, may be suitable for delivery of global public goods and services to the majority of the 80% world. It is absolutely ludicrous to think that "big" data, AI/ML, blockchain, or smarmy publicity [214] stunts may help, in this context. What we need is the concept [215] of "bit dribbling" perhaps coupled with PAPPU systems to help people improve their quality of life without the constant quest for charity.

Technology may play a central role to reach the billions who need services but not in the form of business [216] which is stable in the West and copied by the thoughtless Eastern schools, especially in India. Technical tools will generate data. The ability to use that data, judiciously, may be key to the value of data, for impoverished nations. Coupling social need with technical catalysts must be optimized in the context of the community and not according to *Wired* or *MIT Tech Review* or *HBR*. Advanced R&D is the bread and butter of progress, but the application of advanced tools must be contextual to the services that the community can sustain. Just because auto maker Koenigsegg claims the *Agera* model was built with a "less is more" philosophy does not mean it is a pragmatic standard of transportation suitable for Calcutta, India. In the realm of systems engineering courses and education, dynamic optimization (DO, Figure 2.14) illustrates a similar perspective. The principle is worth teaching, worldwide, but the practice must be relevant to the case. Do we all need DO in everyday life and living? Is it necessary for all types of edge analytics to process data using convolutional neural networks (CNN) on a mobile device or phone?

The conundrum of ***not*** applying the tools we think we have mastered is counterintuitive to the problem-solving ethos in the 20% world. We are ever ready to
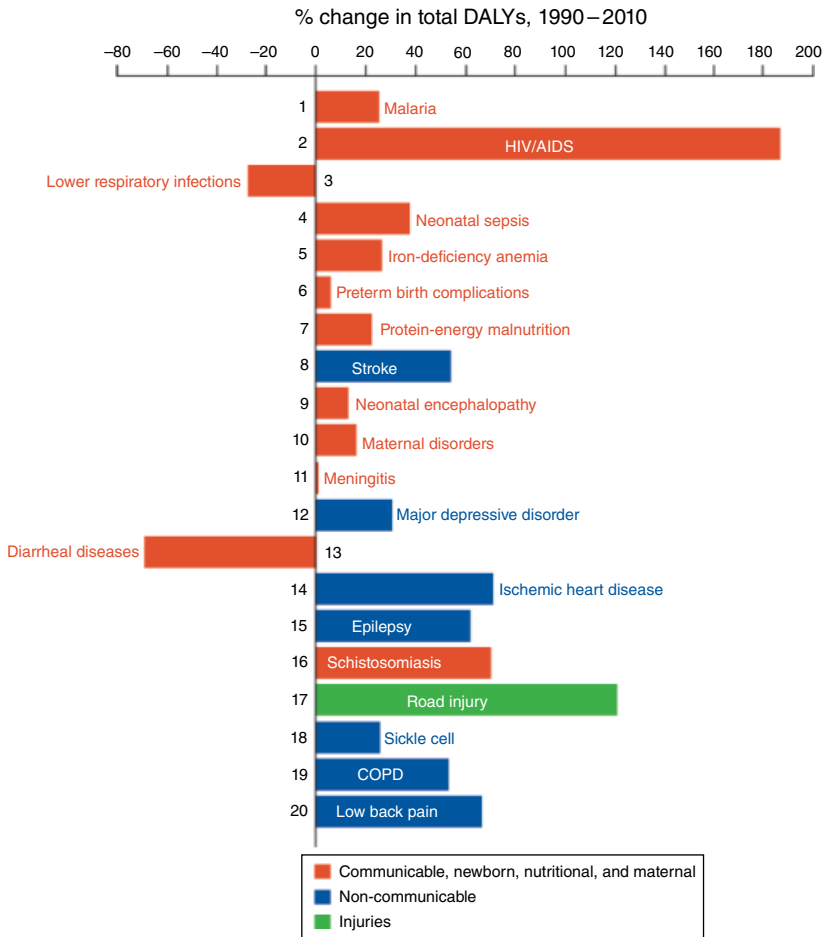
% change in total DALYs, 1990−2010

Figure 2.13 Shifts [211] in leading causes of DALYs for females, Ghana (1990−2010). The leading 20 causes of DALYs are ranked from top to bottom in order of the number of DALYs they contributed in 2010. DALYs (Disability-adjusted life years): The sum of years lost due to premature death (YLLs, Years of life lost due to premature mortality) and years lived with disability (YLDs, Years of life lived with any short-term or long-term health loss causing disability). DALYs are also defined as years of healthy life lost. *Source:* Institute for Health Metrics and Evaluation, Human Development Network, The World Bank. *The Global Burden of Disease: Generating Evidence, Guiding Policy − Sub-Saharan Africa Regional Edition.* Seattle, WA: IHME, 2013. © 2013, Institute for Health Metrics and Evaluation.

use the latest and greatest gadgets from the bleeding edge to derive and drive the best possible perfection and performance. The quagmire of lies aside, we do have real tools which offer notable advantages. But, the volume of the 80% of the world and the economic handicap in these communities must be assimilated in order to
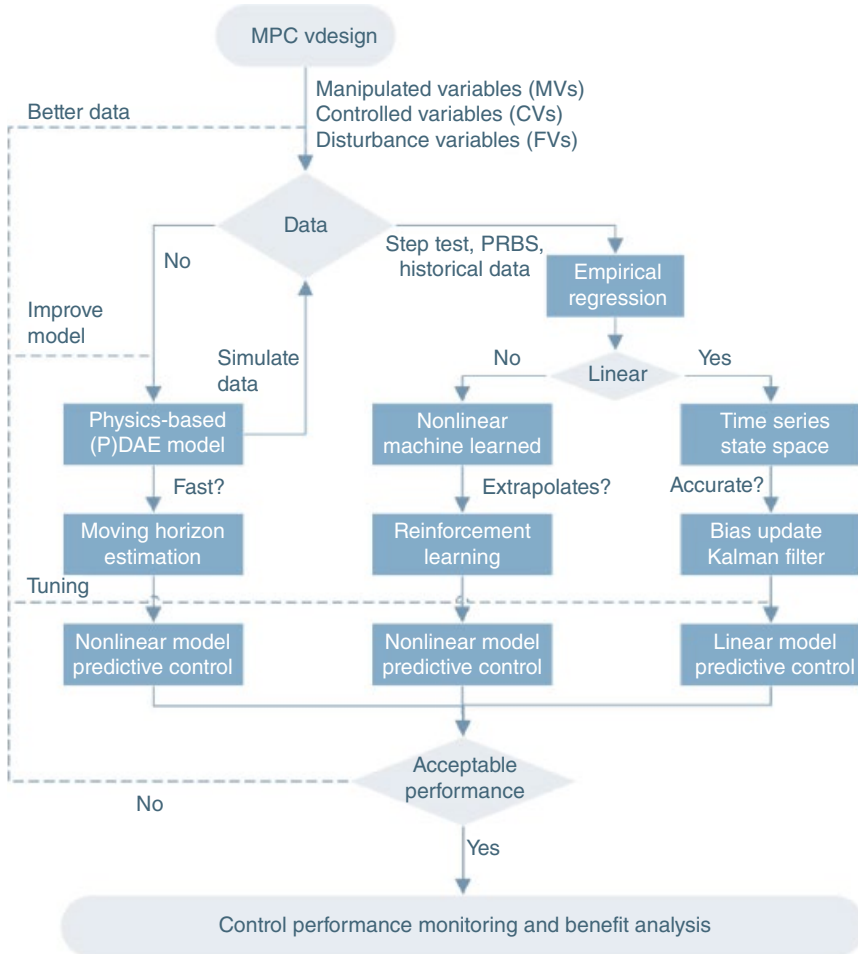
Figure 2.14 Dynamic optimization [217] is a central component of systems engineering where applications of numerical methods for solution of time-varying systems are used to improve performance and precision of engineering design and real-time control applications, which may have a broad spectrum of use, for example, from optimizing the artificial pancreas to fuel cells. Principles of DO may be taught [218] worldwide but DO, systems level data science [219], and Bayesian [220] statistics are excellent tools yet, often, less useful for 80% of the tasks for 80% of the world on the other side of the porous Pareto partition. *Source:* Modified from Hedengren, J.D. Drilling automation and downhole monitoring with physics based models.

change our thinking. The acronym KPI is for "performance" which is euphemistic for profit in the affluent world. It may not be in the best interest of the people. For 80% of the world, perhaps KPI should stand for "key people indicator" and ascertain whether a tool or the service improves the life of people.

Improving lives, however, is relative to the life you aim to improve, a life with disabilities [221] versus life with social [222] void are active domains in robotics. Robotics is useful but the robot propaganda, written mostly by hacks [223] and driven by media [224] sales, is a sign of the times. Essential robotics and robots for tasks that are dangerous, dirty, and dull (repetitive) are a welcome relief, for all involved.

The idea of the automated robotic factory was popularized by Philip K. Dick's fiction "Autofac" published [225] in 1955 (*Galaxy* magazine). The "lights-out" automated manufacturing facility FANUC [226] (factor automated numerical control) has been in operation since 2001, in Japan, but it is an exception. Even though "lights-out" robotics made significant strides in heavy industry, it is far from the Orwellian scenarios promoted through chicanery [227] and buffoonery [228] by discombobulating the masses. The promise of robotics must be balanced with the degree of ***trust*** [229] ***in automated execution*** and the system must be ***acutely aware*** of the perils due to cyberthreats. Cybersecurity is quintessential for automation.

In other instances where human life is at risk (for example, transportation, manufacturing, and mining), the trust in automated action (robot) is as good as the planning for "what ifs" ***when*** the auto execution goes awry. But, that is a deterministic perspective where what could go wrong is anticipated, albeit with some degree of uncertainty. However, if a mobile robot crashes with a holonic manufacturing podium, it may generate a cascade of events where the outcome may be nondeterministic. The critical question in such a scenario is the extent to which a nondeterministic outcome can be tolerated and the *acceptable* cost of risk despite the "open-ended" uncertainty. Few can even approach to answer this/these questions [230] because it verges on the domain of unknown unknowns.

But, that may not deter simulation aficionados from pursuing stochastic (*what if*) models to capture distribution of randomness in nondeterministic outcomes. Heuristics approaches may surface to suggest contingency measures. This is "video gaming" of automation [231] which could turn deadly in reality. The *executive* robot may be suited for "3D" tasks (dull, dirty, and dangerous) but unsuitable for relinquishing human oversight and control if lives are at risk. However, even worse are evil acts perpetrated by humans to bury [232] and ignore [233] the failure of automation, in the pursuit of profit.

Robotic tools in the 20% world are engaged in sophisticated activity which may be subjected to high oversight. In general, the 80% world is not a customer for such implementations in terms of mass consumption. Automation replacing or

reorganizing jobs is not a new event (for example, auto industry) because technology [234] shifts the cycle of jobs and with it, the economy. Rapid changes in skill sets and the volatility of job categories influence other domains, namely, K-16 education, training, skills development, (capital, labor market, and employment) and communication (hopefully, the truthful variety). The rate of change in certain ecosystems is dreadfully slow (for example, education system) whereas the evolution of the job market may resemble the rapid pace of bacterial growth, albeit slower than viral growth rates. The diffusion of robotics will take time and only if the building blocks of automation can be popularized, globally, in a manner that Lego blocks may have inspired young minds to compose, create, and construct.

The 80% world can benefit from robotics, for example, by reducing global disease [235] burden in emerging economies. In India, children are still used to clean sewers (flexible enough to reach cramped spaces, similar to chimney sweepers [236] in seventeenth-eighteenth century UK). Can robotic tools replace the children? Robotics can improve lives and public health [237] rather than fear-mongering and flagrant deception [238] how robots will replace human jobs. Similar to enantiomeric profiling of chiral drugs [239] and opioids excreted in sewer water [240], the post-pandemic world must monitor wastewater for pathogens [241] as a surveillance strategy (prevention tool) or an early warning system prior to the onslaught of detectable clinical [242] symptoms in the general population.

2.12 Stigma of Partition vs. Astigmatism of Vision

The "partition" suggestion does not disguise the reality of the "ours" vs. "theirs" view of a divided world. It is unfortunate but necessary to serve as a constant thorn in our conscience and sow discomfort. The "partition" thinking originates from the corporate pursuit of developing a smörgåsbord of bleeding edge tools and then coerce 80% of the world to buy such products and services ("next billion users"). To add insult to injury, corporations from the 80% world are salivating to acquire rights to these products and bring it to their market (for example, Tata (TCS), TechMahindra, Wipro, Infosys, and other "body-shops" in India [243]). There exists a *nano-cosm* of people in the 80% world who could be a part of the 20% world. Because they are an influential minority and hold the financial power in the 80% world, they are aligning their astigmatic vision, greed, and "profit" objectives with the 20% world.

This mismatch may be at the heart of this global dilemma and creates the necessity to consider the porous Pareto partition in terms of people and service for the end-user. People in AA nations are not buying facial recognition software systems. The abuse [244] is perpetrated by governments. People in the 80% world are not seeking quantum computing to process exabytes of data. People are seeking

simple information, for example, for their health (data for blood cholesterol level) or from their farm (data about concentration of heavy metal contaminants in irrigation water used for fresh produce, such as, tomatoes). These services help people, the end-user, the consumer. This discussion is about what science, engineering, and technology businesses can do for people where the KPI is user-centricity and human-centric [245] well-being, a fact which is immensely clear in the post-pandemic world.

This mismatch between the business to consumer (B2C) services versus the business to business (B2B) services is not new. The 80% world is always looking to the 20% world when planning strategic moves for climbing "up" the supply chain. The fact that the tools from the 80% of the world may not fit in the 20% world is obvious from "frugal innovation" calls [246] by others. Yet, imagination, invention [247], and innovation [248] from R&D in the 20% world are often helpful in the lives of the 80% world. What may be often lost is the *translation* of the advances from the 20% world, for people-centric applications, in the 80% world. This discussion is not singing the praises about the investment in research that only the 20% of the world can afford to push forward because we know [249] the facts. The world is indebted for the strides made possible due to entrepreneurial innovation in such havens such as Massachusetts and California. This discussion is about exposing the lies [250] but not slowing the leaps of vision from the 20% world which may help create low-cost [251] tools to serve the 80% world. The world needs R&D from the 20% world, hence, "porous" partition may facilitate the flow of innovation. Not "as is" but with contextual modifications to better serve communities in the 80% world, at a self-sustaining cost (for example, the PAPPU model, pay-a-penny-per-use or pay-a-price-per-unit).

Mental health is one problem where "porosity" is most welcome because most of the world are affected by generic [252] as well as specific issues, which contributes to economic [253] drain. Inherited bipolar and unipolar disorders [254] do not discriminate on the basis of race, color, religion, or national origin. The neurochemical, neuroendocrine, and autonomic abnormalities associated with these disorders need biomedical research to elucidate the neurobiological basis of these diseases. The latter is not feasible for the 80% world. Harvesting data [255] from external symptoms and pattern analysis [256] may offer a low-cost substitute, to inform the nature of treatment required. People in the 80% world may find it useful.

However, this discussion is not a *to-do* list. It is not a roadmap. It may be a compass, oscillating asynchronously from esoteric thoughts to bare necessities. We are immersed in this duality. One cannot exist without the other. The role of the "partition" is to help focus on the issues that are unique to the environment and community that we wish to serve. It is not a partition of R&D or people or products but a partition for ***delivery of service***.

The idea of democratization of data is a bit buzzy but gimmicks are key to marketing. August institutions, including MIT, are complicit in sponsoring potentially puffy pieces to keep the hype [257] alive. But, the fact remains that enabling data to inform decision is a bedrock of measurement, central to all, irrespective of economic status. The porous Pareto partition is a catalyst to focus on services for the 80% world, where less could be more and serves our sense of égalité.

Dribbling bits of data to inform a person that her RR (Respiratory Rate) is fluctuating, too often, may be a preventative measure (think of the proverb *a stitch in time saves nine*). Informing the person that her RR data are not copasetic, may reduce future morbidity due to COPD (Chronic Obstructive Pulmonary Disease). Providing data and information may be **without** impact on the quality of life in the absence of follow-up (clinic). In terms of data and information, alone, by enabling something simple and even mundane, the people-centric application of technology and data, preferably at the edge [258] (point of use), may help to do more with less. Unbeknownst to us, we are attempting to use the pillars of science, engineering, data, information, and knowledge to build bridges which may serve as a platform to provide service to billions of users. Rather than *gilding the lily*, we are offering a "bare bones" bridge which serves a rudimentary purpose and still may exclude a few. The volume and demand for such low-cost services [**PAPPU** services] may be, eventually, profitable for the business ecosystem.

Supporting a sustainable effort, to lift the quality of life, will depend on the extent of the product **ecosystem** and many other "things" in addition to technical and sensor data as well as the *cohesion of the service supply chain*. Socioeconomic data [259] and related factors are equally significant. Core elements are education of women [260] and trust [261] in women, followed by civic honesty [262], social value [263] as well as inculcating the practice of ethical profitability in social business and entrepreneurial innovation to accelerate the pace of creating pragmatic tools and solutions for remediable [264] injustices.

2.13   The Illusion of Data, Delusion of Big Data, and the Absence of Intelligence in AI

Neither data nor AI [265] is a panacea. Acquisition of data and analysis of data are not a guarantee that there is information in the data or that the information is actionable in terms of delivering value for the user, at an affordable cost. The COVID-19 pandemic has made it clear that the global public goods that define "life-blood" are food, energy, water, sanitation, and healthcare (FEWSH). The 80% world needs contextual, advanced, and affordable array of tools and technologies to leapfrog the conventional practices of FEWSH in the 20% world to vastly improve their crisis response systems.

In this context, energy is one rate-limiting entity and in a "tie" with food and water, in terms of human existence and life. The "hand-me-downs" from the 20% world of energy may not be sustainable. Perhaps the Sahara Desert may be a source of energy for creating a global "battery" field, an idea [266] triggered by an 1877 [267] proposal, in a different context (it was, too, subjected to misrepresentation [268] and mockery [269]). Whether this is a "good" idea or not is *not exclusively* a matter of technological feasibility of implementation or transaction cost of service delivery. The question is, if it is *good for the people*. Global public goods are a matter of context for the community as well as the continent. Exploring the cleavage between entrepreneurial engineering innovation and complexities of social egalitarianism requires willingness to recognize, and adapt, among many *different conceptions of a sense of the future*. It may not be the future deemed appropriate by the 20% world experts. The future is asynchronous and nonlinear (Figures 2.15 and 2.16).

It may be a nonbinary future with multiple paths and unequal connectivity between amorphous nexus of networks representing nonlinear choices, aspirations, and outcomes. "Good" decisions are relative to **that mix** which defies definition yet works as a catalyst for economic rejuvenation. Even this **type** of "good" will (must) change with time and culture because no one version of good can fit all the world [271]. A binary *outcome*, with exceptions, must not be confused with binary *decision-making* because a plethora of nonbinary factors can influence the outcome, which may *appear* as a binary output.

An oversimplified and cherubic example of the latter may resonate with residents of the Boston area. The choice between Mike's [272] and Modern [273], famed confectioners located almost opposite each other on Hanover Street in Boston's North End, is far from binary. The filling in the cannoli, taste of java, and the length of the queue are factored in the decision-making process, which generally presents itself cloaked in a binary-esque outcome. The choice masquerades a slew of nonbinary experiences.

Thus, creativity, imagination, and knowledge (from science, engineering, technology, medicine, and mathematics) may need to connect a few or many "dots" to inform solution development and delivery. It is true that "porosity" may contribute to solutions in the 80% world, and perhaps, less is more, but it will be remiss to leave the reader with the impression that invention/innovation may have to take a second place in the 80% world. In some cases, we must seek *out-of-the-world* or counterintuitive ideas [274] and blend it with incisive insight which may be nontraditional. Far-reaching *convergence* of bio, nano, info, and eco [275] is not an alternative but an imperative to stitch practical solutions, to satisfy, survive, and surpass the criteria dictated by the economics of technology and technology policy [276], which may be necessary to transform grand visions [270] into reality (and uncover new [277] tools, in the process).
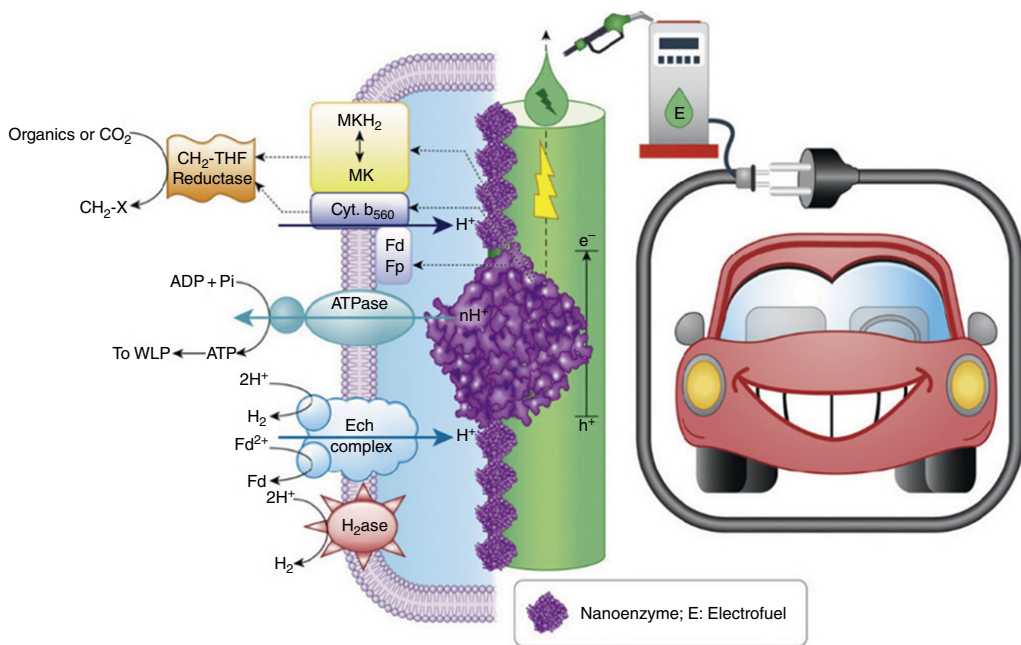
Figure 2.15 The trinity of imagination, invention, and innovation is central for the 80% world to leapfrog the dead weight of old technology and conventional wisdom from the 20% world. *Nanoenzyme–microbe interaction for clean and **affordable** bio-electrofuel production (Figure 1 from Singh et al. Ref.* [270]). *Source:* Singh, L et al. (2020). Bioelectrofuel synthesis by nanoenzymes: novel alternatives to conventional enzymes. *Trends in Biotechnology*, 38 5, p. S0167779919303129 doi:10.1016/j.tibtech.2019.12.017. © 2020, Elsevier.
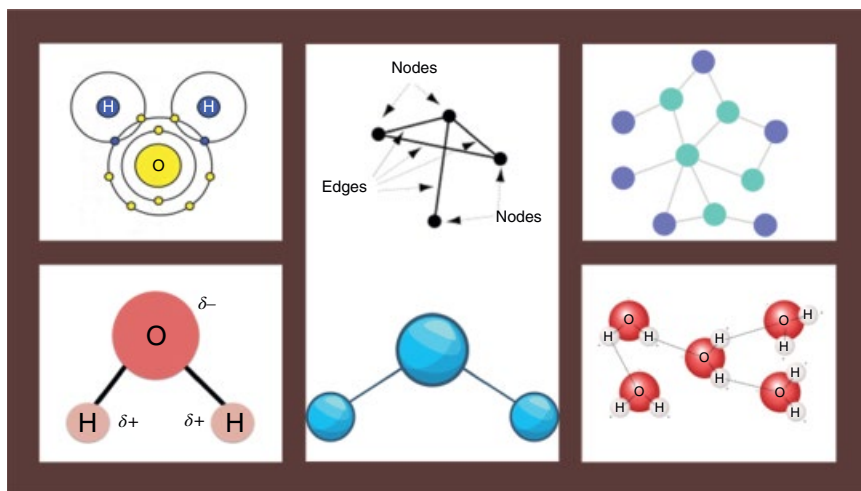
Figure 2.16 Training a neural network to recognize molecules relies on the fact that every molecule may be represented as a *graph* (or a collage of connected graphs, eliciting the idea of a *knowledge graph*). The water molecule may be viewed as a graph with oxygen (O) as the node (vertex). Bonds between oxygen and hydrogen (O–H) serve as the "side" or edge. Most molecules (within reason) may be transformed to a molecular graph and is at the heart of MPNN training to recognize different types of molecules. Then, the *trained* neural network, MPNN, is used to search for similar or *dissimilar* molecules in a repository.

But "grand visions" are often manufactured [278] as incremental mediocrity. Patents for using quantum [279] computing [280] are as absurd as the misuse of the term "cognitive" and the accompanying belligerence [281] in marketing. Those who throw around the term "cognition" may not have consulted a credible expert [282] or explored its meaning/definition {*cognition [n] mental action or process of acquiring knowledge and understanding through thought, experience, and the senses*}. Neither neurology nor modern computational neuroscience comprehends the *combined* electrochemical, cellular, and molecular nature of what it may mean to "acquire knowledge" by animals or humans. Any model, equation, algorithm, or hand-waving "AI" is simply false [283] because it is far beyond our grasp, at this point, to claim anything more than a vague impression of what "*acquire knowledge*" may mean. The other key words in the definition (*thought, experience*) are at depths we do not even dare to know how to measure. Deciphering "processes" based on functional nuclear magnetic resonance imaging (fMRI "activation" maps of real-time blood flow) is mindless drivel due to constraints in spatiotemporal resolution and limited ability of fMRI [284] to reliably detect functional activation. At the current state of instrumentation, resolution is inversely

proportional to the ability to detect functional activation. Optimizing both is essential before fMRI data may be even considered precise.

The absurdity cryptic in the claims about cognition, learning, experience, and thought is neither colored by the author's cognitive dissonance nor a figment of our uninformed imagination. Table 2.4 [285] captures the duration of so-called "deep learning" training over "days" on a tensor processing unit [286] (TPU) scale with vast amounts of data (GB) which generated undifferentiated [287] rubbish. By comparing row 1 vs. 10 (bottom), the scores of the relevant match (#1, 0.892) between learning (saved query) vs. challenge (new query) is unimpressively different (#10, 0.765). According to Google BERT, after several days of "deep learning" ***"Blah blah blah blah"*** it was challenged with the query ***"Does this integrate with gmail?"*** which generated value of 0.765, suggesting ***76.5% similarity between the two.***

The fanfare of GPT-3 [288] and ballyhooed *context-awareness* of ELMo [289], BERT [290], and its cousin ALBERT [291], due to permutations and combinations of including masked language model (MLM) and next sentence prediction (NSP), is utterly devoid of intelligence. Except for fanatics feigning ignorance any rational observer may not be incorrect in thinking that throwing data (please see "Data" row in the upper part of Table 2.4) or using generic high volume of data for training (BERT claims to use "Wikipedia" in column 1, row 4) is ineffective. To be effective, these tools (ELMo, BERT, ALBERT, DILBERT, ROBERTA, and GPT-3) must use training data *relevant to the context* of the target (search). Context-awareness in the absence of data curation is as fake as claiming that a marble bowl is made of gold. Nevertheless, these advances in search techniques are immense strides [292] but the tools still are not "intelligent" but dumb as doorknobs. The doorknob does not turn unless one turns it or actuates it, manually or mechanically. Wikipedia as an experimental ***control*** is a plausible idea. Using curated data for training (ANN) may improve accuracy of search and better guide ***informed users*** to extract notions of connections and relationships with BERT-esque tools as ***supplementary aids*** (Tshitoyan et al. [293] provides supporting evidence, Table 2.5). These tools are of limited value for non-mission-critical applications, for example, recommendations (movies, books, and restaurants), weather for entertainment (IBM's Weather Channel), and fault-tolerant uses (open garage door, on/off sprinklers). Nonessential human-centric uses (congestion routing, temperature control, and voice message to email) may qualify if the outcome is almost correct in 80% of the cases. Actual use with humans-in-the-loop (healthcare, emergency response, and security) may be scuppered if based on any credible risk versus reward analytics, except for offering *nonbinding and non-executable* suggestions or alerts for human decision makers.

The paramount significance of curated contextual data in training any model (including artificial neural networks, ANNs) cannot be overemphasized.

Table 2.4 Google BERT thinks *"blah blah blah"* is **76.5% similar** to *"Does this integrate with gmail?"*

	BERT	RoBERTa	DistilBERT	XLNet
Size (millions)	**Base:** 110 **Large:** 340	**Base:** 110 **Large:** 340	**Base:** 66	**Base:** ~110 **Large:** ~340
Training time	**Base:** 8×V100×12 days[*] **Large:** 64 TPU Chips×4 days (or 280×V100×1 days[*])	**Large:** 1024×V100×1 day; 4–5 times more than BERT	**Base:** 8×V100×3.5 days; four times less than BERT	**Large:** 512 TPU Chips×2.5 days; five times more than BERT
Performance	Outperforms state-of-the-art in October 2018	2–20% improvement over BERT	3% degradation from BERT	2–15% improvement over BERT
Data	16 GB BERT data (Books Corpus + Wikipedia). 3.3 Billion words	160 GB (16 GB BERT data + 144 GB additional)	16 GB BERT data. 3.3 Billion words.	**Base:** 16 GB BERT data **Large:** 113 GB (16 GB BERT data + 97 GB additional). 33 Billion words.
Method	BERT (Bidirectional transformer with MLM and NSP)	BERT without NSP[**]	BERT distillation	Bidirectional transformer with permutation-based modeling

(*Continued*)

Table 2.4 (Continued)

	Saved query	New query	BERT score	USE score	ELMO score	XLNet score
1	How much will this cost?	Is this expensive?	0.892	0.803	0.742	0.720
2	Where is your data stored?	How secure is your product	0.890	0.625	0.765	0.705
3	What temperature is it today?	What time is the game on?	0.880	0.677	0.746	0.730
4	How do I change my password	I cannot find the settings page	0.868	0.671	0.717	0.706
5	Can I sign up for a free trial	Do I need a credit card to get started?	0.868	0.736	0.753	0.759
6	Where can I view my settings	Does this integrate with gmail?	0.866	0.620	0.717	0.691
7	I really do not like this product	I really like this product	0.865	0.747	0.864	0.870
8	What is the Capital of Ireland?	What time is the film in the cinema	0.865	0.578	0.663	0.778
9	Hello, is there anyone there?	What time is the game on?	0.832	0.594	0.680	0.723
10	Blah blah blah blah	Does this integrate with gmail?	0.765	0.519	0.585	0.668

The laughable outcome is not at all surprising despite the hordes of brilliant scientists working to create tools (RNet, XLNet, ELMo, BERT, ALBERT, DILBERT, and ROBERTA) over the past 20 years because it is gnarly to capture *semantics* of language which has evolved over the past 200,000 years.
Sources: Suleiman Khan (2019). BERT, RoBERTa, DistilBERT, XLNet — which one to use. © 2019, Medium; FloydHub Blog (2019). When not to choose the best NLP model.

Table 2.5 (from Tshitoyan *et al.* Extended Data Table 2.4) **–** Analogy scores (%) for materials science versus "grammar" from different sources.

Text corpus	Materials	Grammar	All	Corpus size
Wikipedia	2.6	72.8	51.0	2.81B words
Wikipedia elements	2.7	72.1	41.4	1.08B words
Wikipedia materials	2.2	72.8	41.3	781M words
All abstracts	43.3	58.3	51.0	643M words
Relevant abstracts	48.9	54.9	52.0	290M words
Pre-trained model	10.4	47.1	30.8	640k papers

Training using Wikipedia for – metals – is grammatically rich (>72%) but content poor even when using select Wikipedia for materials (2.2% analogy). The smallest corpus (290 M words) used for training using continuous bag of words (CBOW) offers the best performance (48.9%) on materials-related analogies when **_curated_** for "relevant" abstracts. The best performance for grammar may turn out to be profitable by enabling ELMo, BERT, ALBERT, ROBERTA, and DILBERT (DistilBERT) to be the voices of artificial trainers for the standardized twaddle marketed with impunity and known as *Test of English as a Foreign Language* (TOEFL). The contextual enrichment in this table is similar to the example of enrichment shown in Figure 2.17, suggesting a need for relevance and curation.
Source: Vahe Tshitoyan, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A. Persson, Gerbrand Ceder and Anubhav Jain (2019) Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571,** 95–98. © 2019, Springer Nature.

Individuals and institutions in possession of less than lofty ideals may revert to trickery in an attempt to sow doubt or discombobulate or disqualify the type of outcomes, for example, presented in Table 2.5. It is the age-old deception due to over-fitting [295] which can be also applied to ANN during training and the "fit" may be driven to precision using tools such as recursive feature addition [296] (RFA). For readers seeking a simpler analogy may wish to revisit what we discussed as the "force-fitting" of RDF to healthcare applications (Ref. [168]). The erudition necessary to train ANN with curated data is not easily gleaned from a cursory review. Extensive perusal of scholastic research [297] begins to reveal the minutiae with respect to the nature of the domain-specific data and the context of data curation (see sections 2.3 and 2.4 in Nandy et al., Ref. [297]) that forms the bulk of the *preparatory* work based on rigor and strength of broad-spectrum [298] knowledge. In an earlier section, we referred to *domain-specific* models in a "macro" sense whereas the domain specificity of this example (Nandy et al.) is at the molecular (atomic and/or subatomic) scale.

(Right) Data show large ANN error (−4.91 eV) with respect to DFT for a quintet [Mn(HNNH)6]3+ transition metal complex. The quintet [Mn(HNNH)6]3+

complex highest occupied molecular orbital (HOMO) level is underestimated by 4.9 eV, which is almost **double** the mean absolute error (MAE). This ANN was specifically trained using ΔE_g data models on a set of 64 octahedral homoleptic complexes (OH64). The discrepancy (ANN error) is significant because frontier molecular orbital energetics provide essential insight into chemical reactivity and dictate optical and electronic properties. Small errors could make an immense difference in terms of chemistry of the transition metal complex. In this illustration, the metals are shown as spheres and coordinating atoms as sticks (C atoms, gray; N atoms, blue; H atoms, white). If your healthcare diagnosis and treatment was based on such an ANN outcome, would you trust, accept, and abide by the direction of the treatment suggested by such results? If this outcome is based on data from your electronic health records (EHR) which is known to be erroneous, would you trust poor data quality to inform a poorly performing ANN engine to design your healthcare?

The third piece of evidence that also dispels the marketing myths of AI in favor of viewing through the lens of **artificial reasoning** tools (**ART**, referring to ANN, CNN, RNN, DL, RL), is another variety of neural network [300] with credible capabilities. Message passing neural network [301] (MPNN) for molecules [302] is a tool [303] to unleash data [304] for human-centric applications in health and medicine. This example centers on uncovering and repurposing a previously known molecule as an antibiotic [305] using a plethora of tools including MPNN and collectively referred to as deep learning (DL). Stokes *et al.* and the two other papers (Tshitoyan, Nandy) emphasize data curation and learning, without mentioning the term AI or "artificial intelligence" in the scientific papers. Unfortunately, the marketing and MIT news item [306], as expected, did not shy away from fake sensationalism to bolster the false appeal of AI.

The *learning* that generated the antibiotic (renamed Halicin) is nauseatingly detailed and the *training* (MPNN) was excruciatingly structured, optimized, and re-optimized (using hyperparameter [307] optimization). The old idea of ensembling [308] was applied to improve outcomes *in silico* but predictions were **biologically** tested through rigorous experiments. Even after repeated steps to minimize errors, the authors remain cognizant of the pitfalls: *"It is important to emphasize that machine learning is imperfect. Therefore, the success of deep neural network model-guided antibiotic discovery rests heavily on coupling these approaches to appropriate experimental designs"* (Stokes et al., page 698).

A curated set of 2335 molecules were used as the training set for new antibiotic molecules. The 2335 training dataset included a FDA library of 1760 molecules **pre-selected** (curated) based on their ability to inhibit microbial (*E. coli* BW25113) growth. In other words, molecules with structure and function **known** to possess antimicrobial activity. Training MPNN with this dataset enables the neural network to *learn* the structures in order to select similar (or dissimilar) structures

from a larger library of structures. The expectation is that when a "challenge" library is presented to the MPNN, the degree of similarity or dissimilarity, in terms of the output from the MPNN, can be *tuned* by modifying selection parameters. For example, using prediction scores (PS) to categorize molecules from a larger library (in this case, the ZINC database with ~1.5 billion molecules). By selecting higher PS value (>0.7, >0.8, >0.9), the outcome is "enriched" and a subset of molecules (in this case, 107,347,223, reductionism at work) is further subjected to other selection criteria, for example, nearest neighbor analysis (Tanimoto score). Finally, potential molecules (in this case, 23) are biologically screened (microbial assay) to identify the "new" antibiotic candidate(s). One such candidate is Halicin (Stokes et al.), previously identified as the c-Jun N-terminal kinase inhibitor SU3327 and rediscovered as a broad-spectrum antibiotic, renamed Halicin, but still the same molecule as SU3227, albeit repurposed, based on function.

In combination, these three examples offer preliminary evidence that ART such as ANN, MPNN, DL, RL (reinforcement learning), etc., are **excellent** tools. However, ART and related tools are not intelligent, they do not self-operate and the outcome is solely due to the skill and sophistication of the human operators. The steps must be designed with cautious intellectual strategy embracing the **breadth** of diverse knowledge, often dismissed by many institutions. Execution demands **depth** of erudition and incisive foresight to weigh the pros and cons of the criteria used to assess the *quality of curated data* prior to training neural networks with such data.

It is essential to learn the *meaning of context* in order to sufficiently inform the "artificial" part of ART. Models and patterns are like chicken playing tic-tac-toe [309] without context and semantics. Human *knowledge* to equip ART is almost impossible to transfer because we do not have a clue how to abstract **continuous** knowledge and use **discrete** processes to **build** it into an artificial system. Hence, ART may not "possess" an internal model of the external world. The immense variability in terms of features and which features may be *relevant* in which environment makes it difficult to model a state by claiming that feature selection will address all relevant and discrete contexts that the item or object may experience. Even if feature engineering was automated to levels of precision and continuity that could encapsulate all possible permutations and combinations of the behavior of an entity or object, the model may be still inadequate in the hands of different users due to inherent bias. It is not trivial but may not be impossible to model behavior and optimize for some features within a narrow cross section in a retail environment (for example, who may shop at Whole Foods, who may return to Andronico's versus Mollie Stone's).

If reason could inform common sense, then one may prefer ART over AI and champion the value of reasoning in ML techniques using neural networks. Neural

networks and ML tools are amplifying, modifying, and regurgitating whatever humans have programmed into the tool. It cannot **learn** beyond the range of data or information provided, until humans decide to change, adjust, or add/subtract parameters/attributes which will influence the **learning** and the output from ART. The obstreperous zeal to move away from the misnomer of modern [310] AI [311] and adopt ART as a generic term may be a *back to the future* moment for rule-based [312] expert systems [313] and principles [314] but coupled with new ML tools [315]. Marketing panders to creators of unstructured data but zettabytes of data anarchy occasionally may offer value irrespective of the clamor for general AI or ambient AI or intuitive AI or cognitive AI. Isn't it possible to deliver value using ART?

Does the acronym matter? AI is a false trigger for technology transitions [316] but it is cheaper [317] and cheaper [318] to promote. It is profitable for conference organizers, narcissistic speakers, greedy social gurus, and other forms of eejits. Irresponsible computation may be draining the energy [319] economy, yet the marketing world is oblivious of the socioeconomic incongruencies in terms of thermodynamics [320] of computation, which is absent from daily discussions. Perpetrating the myth of intelligence in AI may be a moral anathema. ART lacks the cachet and panache, but promotes the rational idea of *learning* tools, which are, and will be, helpful for society.

The state of artificial learning is analogous to receiving a map of the world on a postage stamp and expect the bearer of the map (stamp) to arrive at 77 Massachusetts Avenue, Cambridge, MA, using that postage-stamp-sized map as the only guide. Neurologists shudder [321] at AI, the public are ignorant of the evidence of sham (Tables 2.4–2.6 and Figure 2.17) while marketing accelerates the "show" over substance. Sensationalism amplifies attention and siphon funds away from real-world issues, making it harder for elements of FEWSH [322] to move forward, for the 80% world. The task ahead is to be creative, more than expected, and avoid the oxymoronic implementation of *innovation as usual.* Dynamic combinations [323] and cross-pollination of counterintuitive connections may be worth exploring [324] to find many *different* ways to lift billions of boats, not just a few yachts. Future needs égalitarian resistance to our "default" state of society where we allow, acquiesce, and accept greed [325] as a price and penalty we pay for progress (Figures 2.18).

2.14   In Service of Society

The tapestry of this discussion touches upon models of data, fake propaganda about AI, use of ART, and the lack of managed sanitation services for at least a billion people in the world ravaged by a pandemic. It presents a tortuously

Table 2.6 Even after extensive training using precision data enriched for features using RFA, it is not surprising when gross errors are found in the outcome (analysis).
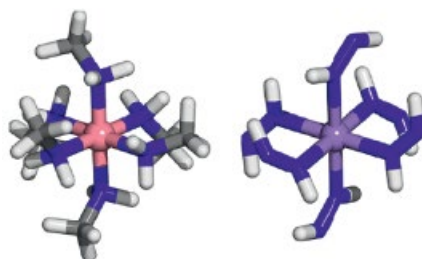
	$[Co(NH_2CH_3)_6]^{3+}$	$[Mn(HNNH)_6]^{3+}$
DFT	$-20.00\,eV$	$-18.64\,eV$
ANN	$-19.91\,eV$	$-23.55\,eV$

Figure 7 (from page 13981 in Nandy et al.) is one example how artificial neural networks (ANN) used in machine learning (ML) exercises and analytics generate erroneous results. ΔE_g data (LEFT) shows ANN error (0.09 eV) with respect to DFT (density functional theory [299]) in a singlet [Co(NH2CH3)6]3+ transition metal complex.
Source: Aditya Nandy, Chenru Duan, Jon Paul Janet, Stefan Gugler, and Heather J. Kulik (2018). "Strategies and software for machine learning accelerated discovery in transition metal chemistry." *Industrial and Engineering Chemistry Research*, volume 57, number 42, pages 13973–13986. © 2018, American Chemical Society.

complex series of challenges each with its own bewildering breadth (Figures 2.20 and 2.21).

More than ever, it is thorny to grasp how to balance efforts to continue creativity of thought considering the grim fact that more than 15 million people are infected by SARS-CoV-2 virus and deaths may soon exceed a million [340] people. The pandemic [341] may continue for another few years (2021–2025) and fluctuate in severity (acuity) due to antigenic drift [342] naturally caused by mutations. Lofty, esoteric ideas from any affluent [343] oasis poses moral and ethical dilemma when billions are facing a mirage in their effort to obtain the essentials for survival (FEWSH).

Yet, the pandemic has made molecular epidemiology the third most important job in the world (medical professionals and essential workers are first and second, respectively). Epidemiology is at the heart of public health ***data***. Without metrics to measure performance, we may be forced to rely merely on anecdotes. The plural of anecdotes is not evidence.
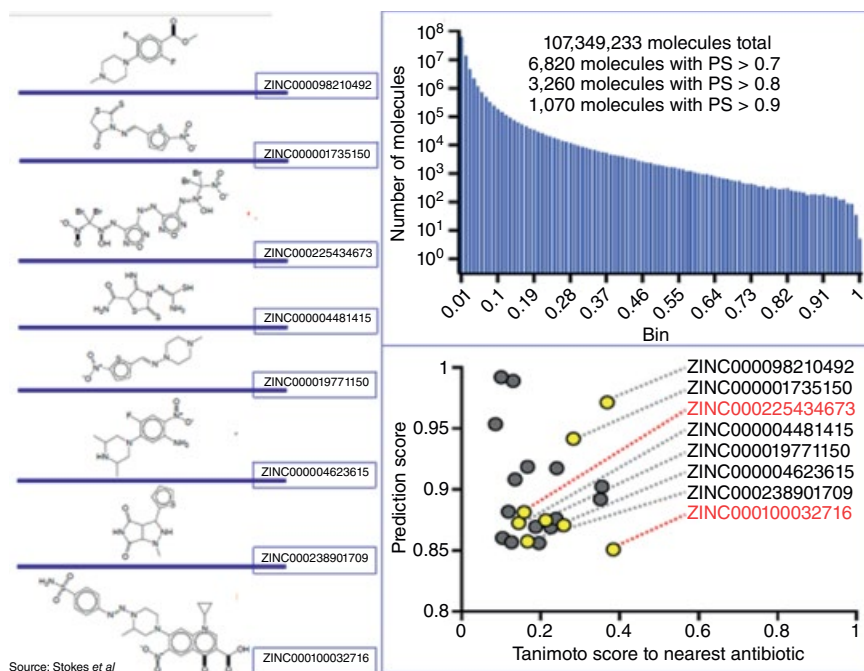
Figure 2.17 Potential candidates (eight molecules) from ZINC database (structures on the left) were scored using nearest neighbor (NN) analysis (yellow circles, bottom right). NN is based on principles derived a thousand years ago [294] (circa 1030). Data are curated at successive steps by enriching for **context** (selecting higher prediction scores, PS, top right) in a manner similar to Table 2.5 (Wikipedia vs relevant abstracts). *Source: Based on Marcello Pelillo (2014) Alhazen and the nearest neighbor rule. Pattern Recognition Letters* 38 (2014) 34–37.

2.15   Data Science in Service of Society: Knowledge and Performance from PEAS

We continue exploring the many facets of data but begin to ascend the pyramid how data may optimize the output or performance, an enormously complicated topic with vast gaps of knowledge.

The idea that tools must perform adequately is obvious and simple to grasp. Yet, performance as a scientific and engineering metric is a product of a fabric of dependencies and parameters which must be combined to deliver service, hopefully, useful for the community, for example, public health. In the digital domain, data is the common denominator, for example, data is the central force in epidemiology. The use of data to extract information must be a part of the discussion
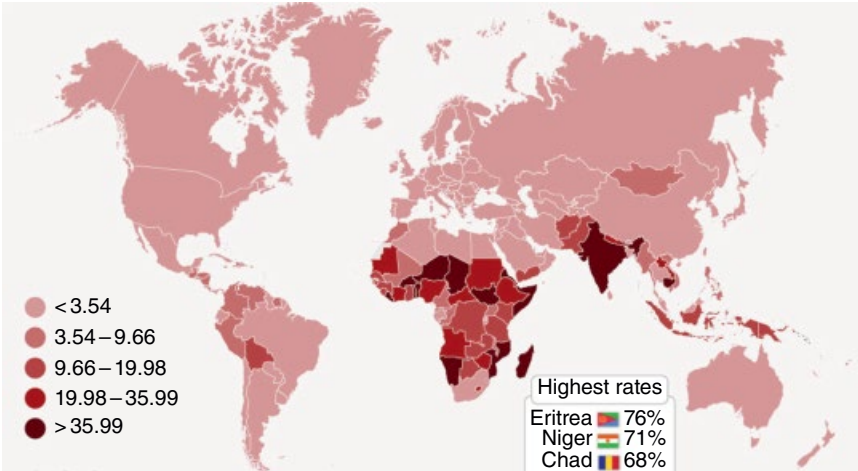
Figure 2.18 Is GINI coefficient [326] a nonstationary end-goal for economic redistribution through ethical social entrepreneurial innovation? A billion [327] people [328] defecate outdoors (figure shows percent of population who are forced to defecate outdoors). The pay-a-penny-per-unit (PAPPU) model could rake in billions if managed sanitation services were developed as a business. If a billion people paid one penny (US) per use per day for their "leased" sanitation service (at home), then the global gross earning for pay-per-use sanitation may be US$3.65 billion annually, an indication of earnings potential and wealth from the business of the poor. The primary assumption is that the individual will choose to pay one penny per day even if their income is only $2 per day (lowest per capita average earnings). This business model depends on availability of many different domains [329] of infrastructure necessary to offer home sanitation as an e-commerce [330] service. The ROI will be realized gradually because earnings will not be US$3.65 billion in the first year. Is the inclination to invest, and wait longer for a ROI, too much to expect from global organizations which could help facilitate delivery of global public goods? [331] *Source:* Andrew, G. (2015). *Open Defecation Around the World.* The World Bank.
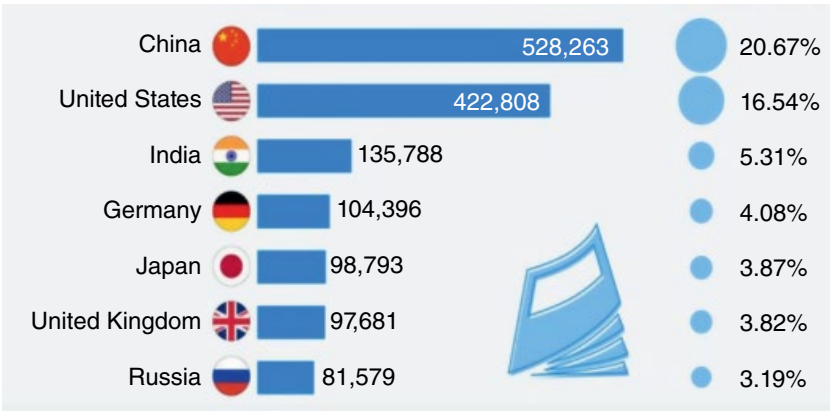

Figure 2.19 Publications in peer-reviewed journals (2018). Face-saving feel-good false positives? Propaganda [332] (does **not** exclude scientists [333]) masks facts [334] and analyses, by ignoring the quality of publications, citations, and investment in R&D (% GDP). Dubious research [335] output tarnishes the image and publications. Are we pointing fingers at a quarter (25.98% = 20.67% + 5.31%) of the global share?

Figure 2.20 GINI coefficient gone awry? Vast slums, adjacent to high-rise residential buildings in a section of Mumbai (Photograph by Prashant Waydande [336]), contradicts the notion that India may be an emerging leader in credible scientific research [337] (Figure 2.19). Are these a few symptoms stemming from grave gender bias, discrimination [338] against females and inequity of women in science and society?
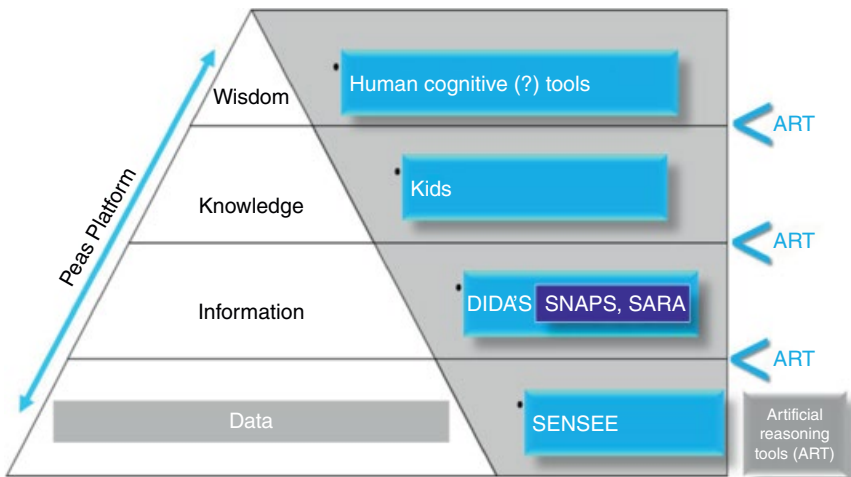

Figure 2.21 Combined subset of P, E, A, S is also a "performance" platform to aggregate tools. Data enable performance (using data from sensor search engines, SENSEE), other databases, logic functions (for example: if this, then that [339] type of operations) and/or ART to deliver near real-time output (mobile app) for end-users (sensor analytics point solution, SNAPS). At the next level, "point" data may auto-actuate to reach SARA (sense, analyze, response, actuate), a part of being data-informed (DIDAS).

with respect to science in the service of society. These tools must be ubiquitous, accessible, composable, and modifiable, on demand.

Information depends on raw data but data may not, always, contain information. Even though data are driving the granularity, the ***curation*** of the contextual from the granular (data) to extract relevant meaningful information is essential for decision systems including digital transformation. A plethora of real and fake tools and technologies are peddled on this pilgrimage from data to information. On this route, one often finds the snake-oil of AI [283] which is a fake tool and a marketing gimmick. But, under the bonnet of fake AI lies the very valuable techniques from ML and key statistical tools which are the "bread and butter" for many instances, applications, and information-informed outcomes.

We return (please rereview Figures 2.7 and 2.8) to the systems world to revisit PEAS, a mnemonic borrowed from agent systems which consists of percepts (P), environment (E), actuators (A), and sensors (S). Performance is the informed outcome from the PEAS process which is a superset of SARA [344], which we mentioned in the beginning of this essay (see Section 2.1). Performance is an overarching goal of PEAS because solutions evolving from science and engineering in the service of society are not only "point solutions" or "tokens" but a fabric or tapestry, which must integrate, analyze, synthesize, and synergize data and information from various inter-dependent domains to arrive at recommendations or actions that can deliver value to nonexpert users. Point solutions (SNAPS) are useful but "fabrics" may better suit complex cases. PEAS must combine data/information from subsystems (P, E, A, S) for data- (informed) DIDAS and in future, knowledge-informed decision support (KIDS). Data may be inextricably linked to sub-domains and outcomes may be influenced by interrelationships which are predominantly nonlinear. For specific instances we may need to select cross sections of data domains pertinent to specific applications. Once we determine which segments of data are required, the process of "discovery" is key to link the (contextual) data. Logic tools, agents, and ART (in combination) may extract the information. Metrics must evaluate the (semantically viable) outcome, if we wish to certify it as rigorous, reproducible, and credible for data and/or information-informed decision support (DIDAS).

Digital transformation depends on making sense of these relationships and the granularity of nearly noise-free data to support the claim that the relationships in question indeed have dependencies (as opposed to correlation without causation). The ability to discover and connect these interrelated data domains is challenging and yet is the heart of "knowledge" systems. Only those who do not dread the fatiguing climb of the steep path from data to information may reap the harvest of what it is to be information-informed. Ascension to knowledge is difficult by orders of magnitude due to the necessity to discover the relationships/dependencies and then validate the connectivity with credible (curated) data which may be globally distributed in unstructured or structured (databases) forms.

One "connectivity" tool from graph theory is referred to as KG which is a connected graph of data and associated metadata. KG represents real-world entities, facts, concepts, and events as well as relationships between them, yielding a comprehensive representation of relationships between data (healthcare data, financial data, and company data). The metadata about the data may catalyze feature integration, access to models, information assets, and data stores. Interoperability between KG tools is quintessential for data discovery and growth of knowledge networks. Although "knowledge" is the term of choice, the outcome of KGs are still far from knowledge but efficient in establishing connectivity between entities. The term "connectivity graph" is most appropriate. But in the data science parlance it is exaggerated to convey the impression that graph-theoretic tools may approach the "knowledge" level in the DIKW pyramid. It does not. KGs are excellent tools for better connected DIDAS but not KIDS.

The cartoon in Figure 2.22 is a hypothetical example of how KGs and graph networks may serve as a backend for point of care (circle in square) applications, for example, a patient suffering from coronavirus infection who is treated by one or more medical professionals. The physician or nurse may create an *ad hoc* medical profile of the patient to prepare a treatment plan by combining data and/or information, for example, [1] stored data (medical history) from EHR about patient's blood group [2], patient's blood analysis results from pathology lab to determine antibodies (mAbs) in serum [3], available drugs (remdesivir, dexamethasone) for treatment, and [4] whether the ventilator (device) is available, if necessary. Accessing the strands of data in real time at the point of care (on a secure mobile device) is typical of medical decision support. The data/information at the point of care must be relevant to the context of the patient (age, height, weight, symptoms, existing conditions, and allergies). Relationships between data and their dependencies with respect to this ***specific*** patient must be "understood" by the data discovery process (the importance of semantic metadata) when the medical professional queries the system (technical metadata). Logic, limits, rates, flows, policies, exclusions, etc., which determines "where/when/how/if this then that" must be a part of this system. Data discovery may be a combination of standard data (for example, the range of values for normal blood count, such as, platelets, hemoglobin, etc.) as well as data discovered specifically for the patient's treatment plan.

KGs are useful for relationship mapping and mining. KG maps are useful for discovery of data assets using agents, algorithms, and search engines in knowledge graph networks (KGN) which can be accessed and triggered by external queries. KG is just one element of digital transformation. It is ***not a panacea*** or a general solution. KGN requires pre-created KG and generating KG requires deep understanding of the nodes that the graph will connect. Individuals creating KGs for domain-specific use must know computational aspects as well domain
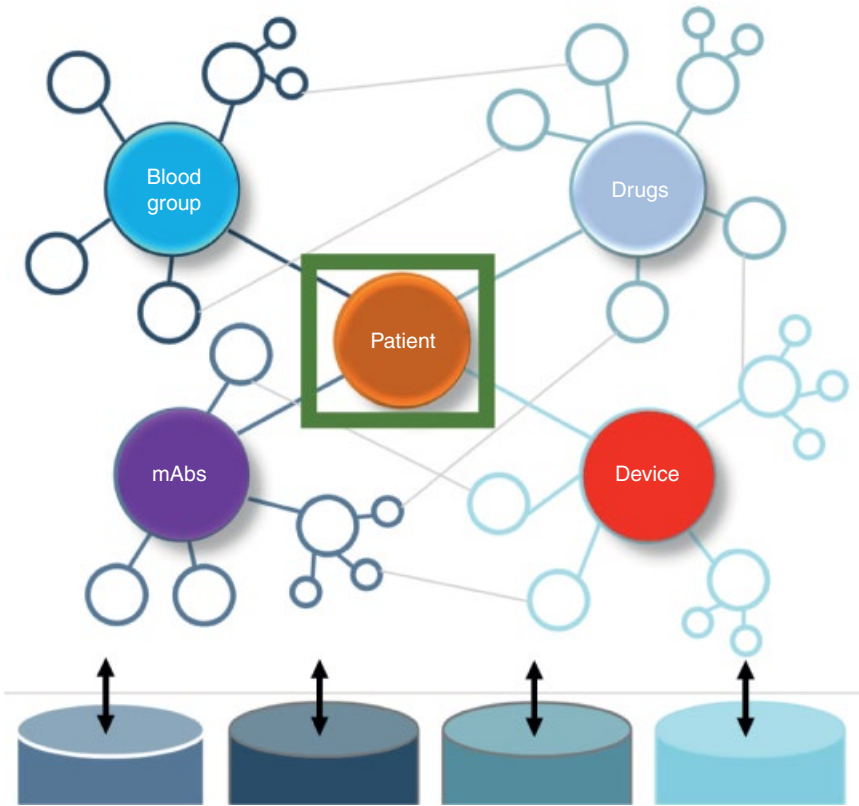
Figure 2.22 Knowledge graphs (KG) may play an increasingly important role if standards may endow KGs with the ability to access, connect, and catalyze data fusion using relevant data (sourced using data discovery agents/tools) followed by reasonable convergence, curation, and analytics of contextual data from distributed databases, based on query or problem specificity. The ability for search and discovery tools to semantically understand the query (language) is one caveat in using KG tools as a backend layer for applications at the point of use. Agents and algorithms for search and discovery of graph networks to access data and information assets may be handicapped by the lack of standards and interoperability between standards in this developing field. Global standards may not be easy to formalize but domain-specific standards developed through agreement between associations may be one mechanism for enabling the dissemination of graph tools for use at the edge (customer, user) in select domains, for example, in a medical subfield (otolaryngology) or water resource optimization (management of reusing wastewater for irrigation) or prevention of food waste (predictive perishability, and shelf-life).
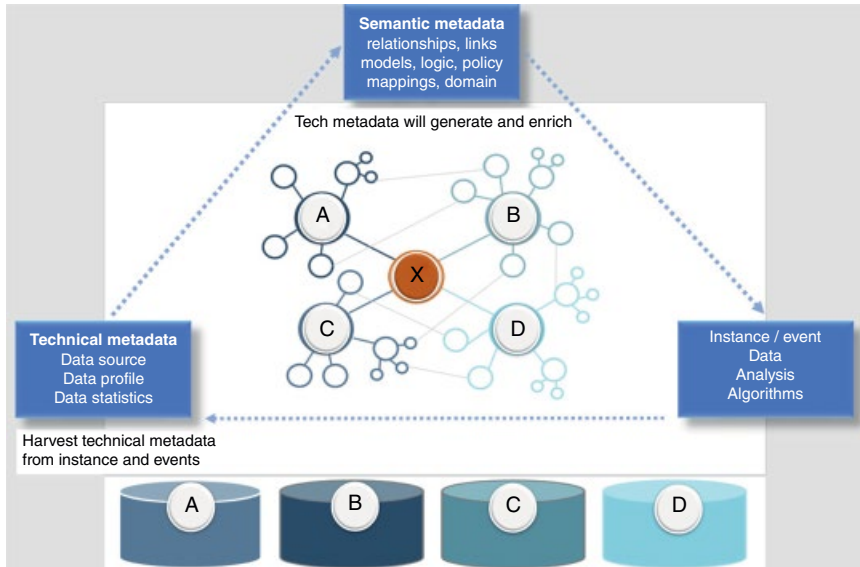
Figure 2.23 Generalized approach to use of knowledge graph networks for application-specific use which assumes interoperability between standards, tools, data, and information (but systems may be incompatible).

knowledge. Introducing bias by connecting select domains may introduce errors and cause harm. On-demand *ad hoc* tools to automate creation of KG and KGNs may suffer from semantic ambiguity of NL-driven bots.

KG abstractions in the PEAS context suggests an overlap between two strategies, both grounded in a multifactorial approach to decision support based on links, relationships, and dependencies. The difference between cartoons in Figures 2.22 and 2.23 versus illustration in Figure 2.24 is one of domain specificity (healthcare, Figure 2.22 or generic, Figure 2.23) compared to the big picture of PEAS (Figure 2.24). In the context of domains, KGs are connecting sub-domains and even granular data/information assets which lie far below the surface (for example, A, B, C, D, Figure 2.23).

Taken together, meaningful use of data and selected types of digital transformation tools are applicable to many domains which can help society. The financial instruments based on the paradigm of PAPPU may be key for ethical profitability if we wish to democratize the benefits and dividends from digital transformation for billions of users. Social business ideas to provide services to help reduce food waste and provide access to primary healthcare may use some version of PAPPU to collect micro-revenue. Suggestions (see items 00, 01, 02, 03, and 04 [345]) about entrepreneurial innovation related to FEWSH may be essential as global public
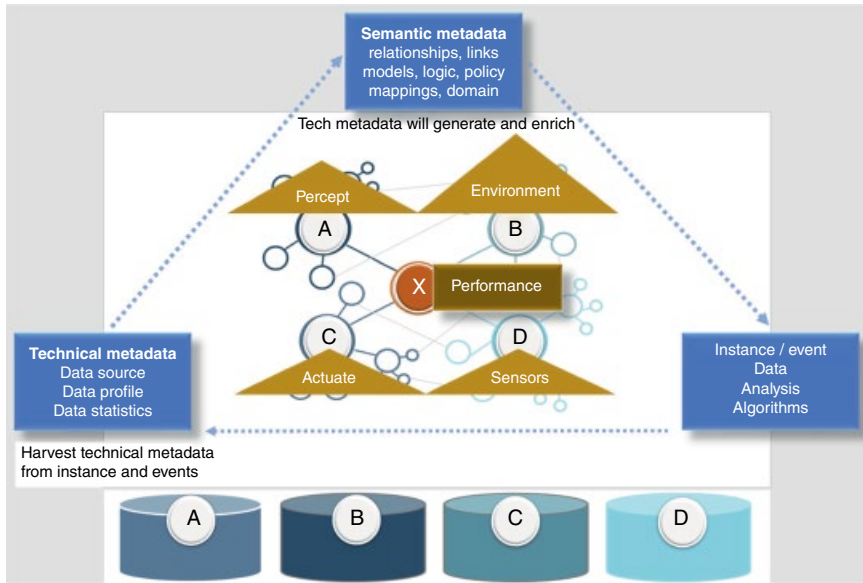
Figure 2.24 Connectivity permeates KG which create relationships between nodes (granular data and information). PEAS is a view from the top of MANY layers of knowledge graphs which may be broadly grouped under percept, environment, actuate, and sensors, representing trillions of use-cases where specific elements within these groups may be involved/weighted/combined in delivering decisions.

goods but they must work "in concert" to deliver the performance – survival in the post-pandemic world and improving the quality of life. Performance as the outcome of the PEAS paradigm (Figure 2.24) works by connecting ***ideas of domains*** (percept, environment, actuate, and sensors) using KGs. Comparing Figures 2.21–2.24 may help to visualize the big picture as an ***abstraction of connectivity*** between PEAS and KG.

These system of systems and multilayer convergence suggested in Figure 2.25 embrace a dynamic broad-spectrum of options and opportunities which may be beyond the grasp of end-users, for decades. It is useful to sketch the "big picture" but implementation of these systems and extracting the synergies are beyond civilian reach. The complexity in the cartoon may be analogous to the problems faced by the "omics" tools in medicine and healthcare (metabolomics, genomics, and proteomics). Physiological changes prior to and during a disease state are almost never a binary outcome. The network view in Figure 2.25 is even more intensified in humans and animals because changes in physiology and metabolism are almost always a network effect and genome-wide associations are only too common. Dissecting data to extract useful information may not be achieved by
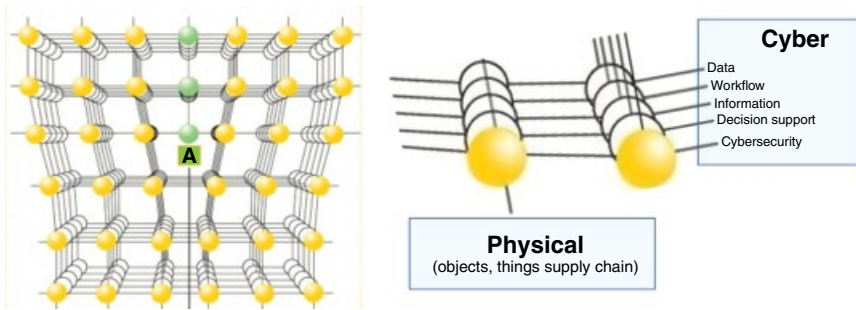
Figure 2.25 PEAS and knowledge graph (KG) networks are elements of the DIKW paradigm which must work in conjunction with cyber-physical systems (CPS), a superset of IoT activities. Actuation (A) could induce commencement of activities limited to a few "point" outcomes ("green" circles in cartoon) or potentially trigger network effects ("yellow" knobs illustrate control elements which may be in a farm, irrigation system, field sensors, or oil/gas pipeline) to modify rate, flow, or simply turn on/off. The actuation may be automatic (if this, then, that) or machine-to-machine (M2M) or some variation of machines and humans in the loop. The ***wire frame*** layer connecting the lattice represents ecosystems affected directly or indirectly by actuation. The elements may be "cyber" (workflow, cybersecurity, data exchange, information arbitrage, and decision support) as well as "physical" (involving objects – for example – flow of water, detection of molecules and physical in terms of the supply chain of products, and/or services influenced by actuation – for example – recall select batches of lettuce if pathogen is detected by sensor). Actuation is a "performance" outcome, which may engage with and affect multiple nodes, data, and IoT domains.

focusing on one or two aspects. The sluggish pace of progress in harvesting the value of genomics in precision medicine is one example. The herculean task of protein profiling [346] is still a work in progress. Hence, field applications may still need point solutions (SNAPS).

2.16 Temporary Conclusion

Digital transformation is a buzz word unless contextual data are meaningfully used to deliver information of value. Graph structures (KG) are an infrastructure of relationships to connect relevant granular data to improve total performance. The choice of tools from graph theory is based on their mathematical credibility. The combination of KG with PEAS may offer the rigor we seek for measurable progress of data-informed digital transformation to help the 80% and the 20% world. We are still in the early stages and changes will accompany digital transformation (for at least another century) not because it is ***due to*** digital transformation but because change is a gift of periodicity. Focus, applications, and our demands
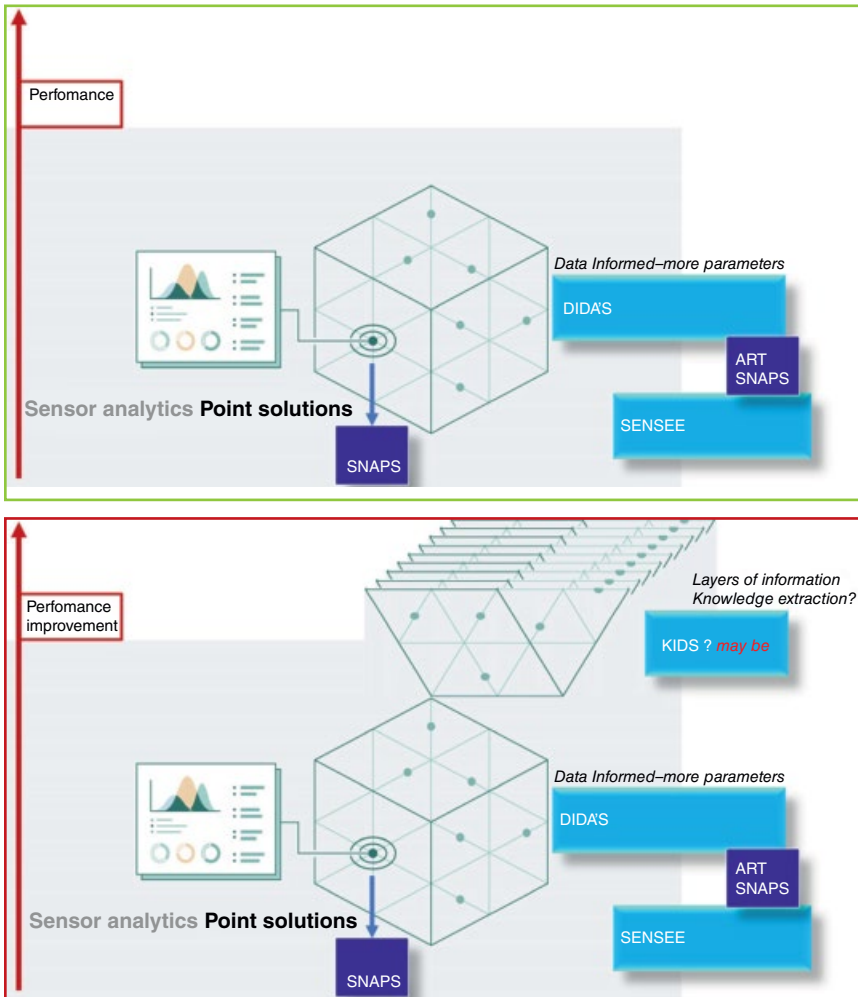
Figure 2.26 Complexity _**seems**_ erudite (red border) but less useful in the real world. Probably 80% of users may benefit from point solutions (SNAPS) generating _**usable information**_ from a cross section of "small" data (time series) which may assist in addressing local point of use problems in near real time.

change. Analyzing, predicting, and understanding these changes may be challenging. Tools in the process of change must evolve, adapt, and serve what is best for our society, industry, and commerce, at that time. With the progress of time, events evolve and unleash their influence on the global ecosystem.

If data cannot deliver performance, then the value of such data is negligible in a world where the "outcome economy" is no longer about the product but the value
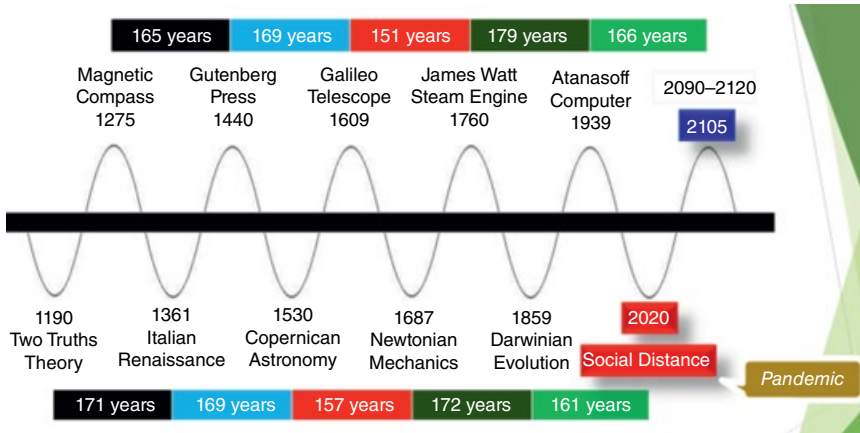
Figure 2.27 A sinusoidal wave illustrates the periodicity of change. We are currently immersed in a social coronary [347] due to the coronavirus pandemic. The Great Transformation [348] of 2021 is expected, soon.

of the meaningful service (which may include the product). The trainer or sneaker is now a "service" where the consumer pays for duration of the service (life cycle) provided by the trainer or sneaker. It may seem tad obsequious but the service economy is here to stay, especially for the 80% world. The PAPPU model may unlock markets of billions who may have nonzero amounts of disposable income. Wealth of the poor may fuel ethical profitability for social business by tapping into markets of next billion users. Current shades of "western" business ethos, roles, and models are unsuitable as pillars for the future of service economies. Expect resistance from *western* business behemoths if they fail to control the 80% world.

Nevertheless, data is only one tiny element in the forthcoming global transformation, which will test our knowledge and our greed. Compassion without knowledge is ineffective. Knowledge without compassion is inhuman [349]. In our pursuit of life, we must think about our duties and responsibilities. In our pursuit of ideas, we must strive to think about the correct questions. Wrong questions will generate wrong answers [350]. Humanity needs compassionate dreamers [351] and an ethical sense of higher purpose guided by humility. Leaders with a higher moral fabric are few and far between. Humanity seeks those who can rise above their personal greed and narcissistic needs. To believe in greater good for the greatest number is that pursuit of "what life expects of me" [352] which often presents itself as a braided lifeline of chance, choice, and character.

Les savants des autres nations, à qui nous avons donné l'exemple, ont cru avec raison qu'ils écriraient encore mieux dans leur langue que dans la nôtre. L'Angleterre nous a donc imités; l'Allemagne, où le latin semblait s'être réfugié, commence insensiblement à en perdre l'usage; je ne doute pas qu'elle ne soit bientôt suivie par les Suédois, les Danois et les Russes. Ainsi, avant la fin du dix-huitième siècle, un philosophe qui voudra s'instruire à fond des découvertes de ses prédécesseurs, sera contraint île charger sa mémoire de sept à huit langues différentes, et. après avoir consume à les apprendre le temps le plus précieux de sa vie, il mourra avant de commencer à s'instruire [353].	The scholars of other nations, to whom we have provided an example, believed with reason that they would write even better in their language than in ours. England has thus imitated us; Germany, where Latin seems to have taken refuge, begins insensibly to lose the use of it: I do not doubt that it will soon be followed by the Swedes, the Danes, and the Russians. Thus, before the end of the eighteenth century, a philosopher who would like to instruct himself about his predecessor's discoveries will be required to load his memory with 7–8 different languages; and after having consumed the most precious time of his life in acquiring them, he will die before having begun to instruct himself [354].

Acknowledgements

References

1 https://en.wikipedia.org/wiki/Pareto_principle.

2 https://www.postscapes.com/iot-history/.

3 Sarma, S., Brock, D. and Ashton, K. (1999). The networked physical world - proposals for engineering the next generation of computing, commerce, and automatic-identification. MIT Auto-ID Center White Paper. MIT-AUTOID-WH001, 1999. https://autoid.mit.edu/publications-0 https://pdfs.semanticscholar.org/88b4/a255082d91b3c88261976c85a24f2f92c5c3.pdf.

4 Sarma, S., Brock, D., and Engels, D. (2001). Radio frequency identification and the electronic product code. *IEEE Micro* 21 (6): 50–54. https://doi.org/10.1109/40.977758.

5 Sarma, S. (2001). Towards the 5 cents tag. Auto-ID Center. MIT-AUTOID-WH-006, 2001.

6 https://www.rfc-editor.org/rfc/pdfrfc/rfc791.txt.pdf.

7 Deering, S. and Hinden, R. (2017). Internet protocol, version 6 (IPv6) specification. *Internet Engineering Task Force (IETF)* https://www.rfc-editor.org/rfc/pdfrfc/rfc8200.txt.pdf.

8 https://www.bizjournals.com/boston/blog/mass-high-tech/2012/11/mit-aims-to-harness-cloud-data-on-consumer.html.

9 Mühlhäuser, M. and Gurevych, I. (2008). *Introduction to Ubiquitous Computing.* IGI Global https://pdfs.semanticscholar.org/ab0e/b44c7c81a1af3fc2d23fa03f8f04f9e4ca2d.pdf.

10 https://www.bigdataframework.org/short-history-of-big-data/.

11 Kourantidou, M. (2019). Artificial intelligence makes fishing more sustainable by tracking illegal activity. https://theconversation.com/artificial-intelligence-makes-fishing-more-sustainable-by-tracking-illegal-activity-115883.

12 https://www.postscapes.com/iot-consulting-research-companies/.

13 Which market view? The imperfect categories are consumer and industrial IoT. Definitions proposed by domestic and foreign governments (US DHS, US FTC, EU ENISA) are (not surprisingly) neither systematic nor in sync. According to DHS, IoT is defined as "connection of systems and devices (with primarily physical purposes e.g. sensing, heating/cooling, lighting, motor actuation, transportation) to information networks (internet) via interoperable protocols, often built into embedded systems." (p. 2 in "Strategic Principles for Securing the Internet of Things (IoT)", Version 1.0, 15 November 2016. https://www.dhs.gov/sites/default/files/publications/Strategic_Principles_for_Securing_the_Internet_of_Things-2016-1115-FINAL_v2-dg11.pdf). ENISA defines IoT as "a cyber-physical ecosystem of inter-connected sensors and actuators, which enable intelligent decision making." (p. 18 in "Baseline Security Recommendations for IoT" www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot/at_download/fullReport). One version of IoT is in H.R.1668 "Internet of Things Cybersecurity Improvement Act of 2019" (https://www.congress.gov/116/bills/hr1668/BILLS-116hr1668ih.pdf). H.R. 4792 "U.S. Cyber Shield Act of 2019" mentions "internet-connected products" and stipulates that the term 'covered product' means a consumer-facing physical object that can (a) connect to the internet or other network; and (b) (i) collect, send, or receive data; or (ii) control the actions of a physical object or system (https://www.congress.gov/116/bills/hr4792/BILLS-116hr4792ih.pdf).

14 Cisco CEO at CES 2014. Internet of Things is a $19 trillion opportunity. https://www.washingtonpost.com/business/on-it/cisco-ceo-at-ces-2014-internet-of-things-is-a-19-trillion-opportunity/2014/01/08/8d456fba-789b-11e3-8963-b4b654bcc9b2_story.html.

15 www.zdnet.com/article/internet-of-things-8-9-trillion-market-in-2020-212-billion-connected-things/.

16 https://machinaresearch.com/static/media/uploads/machina_research_press_
release_-_m2m_global_forecast_&_analysis_2012-22_dec13.pdf.

17 Novikov, D.A. (2016). Systems theory and systems analysis. Systems engineering.
Cybernetics 47. Springer International Publishing: 39–44. https://doi.org/10.1007/
978-3-319-27397-6_4, www.researchgate.net/publication/300131568_Systems_
Theory_and_Systems_Analysis_Systems_Engineering.

18 Williams, L.P. (1989). André-Marie Ampère. *Scientific American* 260 (1): 90–97.
https://www.jstor.org/stable/24987112.

19 Syverson, C. (2018). Why hasn't technology sped up productivity?. *Chicago Booth
Review*. https://review.chicagobooth.edu/economics/2018/article/why-hasn-t-
technology-sped-productivity.

20 Wassen, O. (2019). Big data facts - how much data is out there? *NodeGraph*
https://www.nodegraph.se/big-data-facts/.

21 OECD (2019). *OECD Compendium of Productivity Indicators 2019*. OECD
https://doi.org/10.1787/b2774f97-en.

22 Huff, D. (1954). *How to Lie with Statistics*. Norton http://faculty.neu.edu.cn/
cc/zhangyf/papers/How-to-Lie-with-Statistics.pdf.

23 Artificial Intelligence - Intelligent Agents (2017). https://courses.edx.org/asset-v1:
ColumbiaX+CSMM.101x+1T2017+type@asset+block@AI_edx_intelligent_
agents_new__1_.pdf.

24 Sutton, R.S. and Barto, A.G. (2018). *Reinforcement Learning: An Introduction*,
2nde. MIT Press https://web.stanford.edu/class/psych209/Readings/
SuttonBartoIPRLBook2ndEd.pdf.

25 Angerman, W.S. (2004). Coming full circle with boyd's ooda loop ideas: an
analysis of innovation diffusion and evolution (Thesis). https://apps.dtic.mil/dtic/
tr/fulltext/u2/a425228.pdf.

26 Kamenetz, A. (2011). *Esther duflo bribes India's poor to health*. *Fast Company*
(8 August). https://www.fastcompany.com/1768537/esther-duflo-bribes-indias-
poor-health.

27 Castanedo, F. (2013). *A Review of Data Fusion Techniques*. The Scientific World
Journal. doi:https://doi.org/10.1155/2013/704504. http://downloads.hindawi.
com/journals/tswj/2013/704504.pdf.

28 https://www.the-scientist.com/uncategorized/nature-rejects-krebss-paper-
1937-43452.

29 https://www.npr.org/sections/thesalt/2017/07/20/527945413/
khichuri-an-ancient-indian-comfort-dish-with-a-global-influence.

30 Mendel, G. 1866. Versuche über Plflanzenhybriden. Verhandlungen des
naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865, Abhandlungen,
pp. 3–47. http://www.esp.org/foundations/genetics/classical/gm-65.pdf.

31 Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of
proteins. *Journal of Molecular Biology* 3 (3): 318–356. https://doi.org/10.1016/
S0022-2836(61)80072-7.

32 Dulbecco, R. (1967). The induction of cancer by viruses. *Scientific American* 216 (4): 28–37. https://doi.org/10.1038/scientificamerican0467-28, http://calteches. library.caltech.edu/230/1/cancer.pdf.

33 http://web.mit.edu/esd.83/www/notebook/Cybernetics.PDF.

34 Weiner, N. (1961). *Cybernetics, or Control and Communication in the Animal and the Machine*, 2nde. MIT Press https://doi.org/10.1037/13140-000, https://uberty. org/wp-content/uploads/2015/07/Norbert_Wiener_Cybernetics.pdf.

35 https://www.sissa.it/fa/workshop_old/DCS2003/reading_mat/zuazuaDivS EMA.pdf.

36 https://www.americanscientist.org/article/alice-and-bob-in-cipherspace.

37 Rivest, R.L. et al. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM* 21 (2): 120–126. https://doi.org/10.1145/359340.359342, https://dl.acm.org/citation. cfm?id=359342.

38 Nash, J.F. (1950). Equilibrium points in N-person games. *Proceedings of the National Academy of Sciences of the United States of America* 36 (1): 48–49. https://doi.org/10.1073/pnas.36.1.48, www.pnas.org/content/pnas/36/1/48. full.pdf.

39 Axelrod, R.M. (2006). *The Evolution of Cooperation*. Basic Books http://www. eleutera.org/wp-content/uploads/2015/07/The-Evolution-of-Cooperation.pdf.

40 Kreps, D.M. et al. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory* 27 (2): 245–252. https://doi.org/10.101 6/0022-0531(82)90029-1.

41 https://en.wikipedia.org/wiki/Graph_database#/media/File:GraphDatabase_ PropertyGraph.png.

42 Little, J.D.C. et al. (1963). An algorithm for the traveling salesman problem. *Operations Research* 11 (6): 972–989. https://doi.org/10.1287/opre.11.6.972, https://dspace.mit.edu/bitstream/handle/1721.1/46828/ algorithmfortrav00litt.pdf.

43 Schrader, C.R. (1997). *United States Army Logistics, 1775-1992: An Anthology*, vol. 1. https://history.army.mil/html/books/068/68-1/cmhPub_68-1.pdf.

44 Datta, S. et al. (2004). Adaptive value networks. In: *Evolution of Supply Chain Management: Symbiosis of Adaptive Value Networks and ICT* (eds. Y.S. Chang et al.), 3–67. Springer US https://doi.org/10.1007/0-306-48696-2_1, https://link. springer.com/chapter/10.1007/0-306-48696-2_1.

45 www.cs.cmu.edu/afs/cs/academic/class/15251/Site/current/Materials/Lectures/ Lecture13/lecture13.pdf.

46 https://kluge.in-chemnitz.de/documents/fractal/node2.html.

47 https://trove.nla.gov.au/work/21161891?q&versionId=25229969.

48 Sterling, M.J. (2015). Mathematics and Art. https://www.bradley.edu/dotAsset/ d62c7fce-87ed-4b61-9ce7-23a78b70144d.pdf.

49 Carmody, K. (1988). Circular and hyperbolic quaternions, octonions, and sedenions. *Applied Mathematics and Computation* 28 (1): 47–72. 10.1016/0096-3003(88)90133-6. 10.1016/0096-3003(88)90133-6.

50 Pennisi, E. et al. (2018). The momentous transition to multicellular life may not have been so hard after all. *Science* https://www.sciencemag.org/news/2018/06/momentous-transition-multicellular-life-may-not-have-been-so-hard-after-all.

51 How Many Bacteria Live on Earth*? Sciencing*. https://sciencing.com/how-many-bacteria-live-earth-4674401.html.

52 Prusiner, S.B. (1982). Novel proteinaceous infectious particles cause scrapie. *Science* 216 (4542): 136–144. https://doi.org/10.1126/science.6801762, https://science.sciencemag.org/content/216/4542/136.long.

53 Pattison, I.H. and Jones, K.M. (1967). The possible nature of the transmissible agent of scrapie. *Veterinary Record* 80 (1): 2–9. https://doi.org/10.1136/vr.80.1.2, https://veterinaryrecord.bmj.com/content/80/1/2.

54 *1.4 Universality*. New England Complex Systems Institute. https://necsi.edu/14-universality.

55 Denton, P.B. et al. (2019). Eigenvectors from Eigenvalues. *ArXiv* 1908: 03795. https://arxiv.org/pdf/1908.03795.pdf.

56 Tao, T. and Van, V. (2010). Random matrices: Universality of local eigenvalue statistics. *ArXiv* 0906: 0510. https://arxiv.org/pdf/0906.0510.pdf.

57 Wu, L. et al. (2013). A spectral approach to detecting subtle anomalies in graphs. *Journal of Intelligent Information Systems* 41 (2): 313–337. https://doi.org/10.1007/s10844-013-0246-7, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.499.3228&rep=rep1&type=pdf.

58 Denton, P.B. et al. (2019). Eigenvalues: the rosetta stone for neutrino oscillations in matter. *ArXiv* 1907: 02534. https://arxiv.org/pdf/1907.02534.pdf.

59 Committee on Information Technology, Automation, and the U.S. Workforce et al. (2017). *Information Technology and the U.S. Workforce: Where Are We and Where Do We Go from Here?* National Academies Press https://doi.org/10.17226/24649. National Academies of Sciences, Engineering, and Medicine 2017. doi: 10.17226/24649 and https://www.nap.edu/download/24649.

60 www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf.

61 David, P.A. (1990). The dynamo and the computer: an historical perspective on the modern productivity paradox. *The American Economic Review* 80 (2): 355–361. www.jstor.org/stable/2006600. https://pdfs.semanticscholar.org/dff7/9b2f28cbb79da91becaab803667f30394233.pdf?_ga=2.215911511.1520557695.1570682035-1238830782.1562127126.

62 Syverson, C. (2013). Will history repeat itself? Comments on is the information technology revolution over? *Intl Productivity Monitor* 25: 37–40. www.csls.ca/ipm/25/IPM-25-Syverson.pdf.

63 Syverson, C. (2016). Challenges to mismeasurement explanations for the U.S. Productivity slowdown. Working Paper, 21974, National Bureau of Economic Research, February 2016. doi:https://doi.org/10.3386/w21974. https://www.nber.org/papers/w21974.pdf.

64 Gordon, R. (2016). The rise and fall of American growth: U.S. Standard of living since the Civil War. https://press.princeton.edu/books/hardcover/9780691147727/the-rise-and-fall-of-american-growth.

65 Solow, R.M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics* 70 (1): 65. https://doi.org/10.2307/1884513.

66 Information technology and the U.S. workforce: where are we and where do we go from here?. doi:https://doi.org/10.17226/24649 https://www.nap.edu/read/24649/chapter/5#55.

67 Bivens, J. and Mishel, L. (2015). *Understanding the Historic Divergence Between Productivity and a Typical Worker's Pay: Why It Matters and Why It's Real*. Economic Policy Institute https://www.epi.org/files/2015/understanding-productivity-pay-divergence-final.pdf.

68 Blum, A. et al. (2020). *Foundations of Data Science*, 1ste. Cambridge University Press https://www.cs.cornell.edu/jeh/book.pdf.

69 David, J. (2015). Study: nearly 70 percent of tech spending is wasted. *Vox* (31 October). https://www.vox.com/2015/10/31/11620222/study-nearly-70-percent-of-tech-spending-is-wasted.

70 Zobell, S. Why digitaltransformations fail: closing the $900 billion hole in enterprise strategy. *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2018/03/13/why-digital-transformations-fail-closing-the-900-billion-hole-in-enterprise-strategy/.

71 Giller, G.L. (2005). *A Generalized Error Distribution. SSRN Electronic Journal* https://doi.org/10.2139/ssrn.2265027, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.542.879&rep=rep1&type=pdf.

72 McDonald, J.B. and Yexiao, J.X. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics* 66 (1–2): 133–152. https://doi.org/10.1016/0304-4076(94)01612-4, https://www.sciencedirect.com/science/article/abs/pii/0304407694016124.

73 Granger, C.W.J. (1983). Co-integrated variables and error-correcting models. UCSD Discussion Paper 83-13.

74 Engle, R.F. (1976). Interpreting spectral analyses in terms of time-domain models. *Annals of Economic and Social Measurement* 5 (1): 89–109. www.nber.org/chapters/c10429.pdf.

75 Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 (4): 987. https://doi.org/10.2307/1912773, http://www.econ.uiuc.edu/~econ508/Papers/engle82.pdf.

76 Granger, C.W.J. et al. (2001). *Essays in Econometrics: Collected Papers of Clive W.J. Granger*. Cambridge University Press https://dl.acm.org/citation.cfm?id=781849.

77 Granger, C.W.J. (2004). Time series analysis, cointegration, and applications. https://escholarship.org/uc/item/2nb9f668.

78 Engle, R. (2001). GARCH 101: the use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives* 15 (4): 157–168. https://doi.org/10.1257/jep.15.4.157, http://www.cmat.edu.uy/~mordecki/hk/engle.pdf.

79 https://www.nobelprize.org/prizes/economic-sciences/2003/summary/.

80 Fronte, B, Galliano, G. and Bibbiani, C. (2016). From freshwater to marine aquaponic: new opportunities for marine fish species production. Conference VIVUS, 20 and 21 April 2016, Biotechnical Centre Naklo, Strahinj 99, Naklo, Slovenia. https://pdfs.semanticscholar.org/f8b2/fad3fe3f6f92a7b1c8276ad57876d8fd0a70.pdf.

81 Matthias, D.M. et al. (2007). Freezing temperatures in the vaccine cold chain: a systematic literature review. *Vaccine* 25 (20): 3980–3986. https://doi.org/10.1016/j.vaccine.2007.02.052, https://www.nist.gov/sites/default/files/documents/2017/05/09/FreezingReviewArticle-Vaccine.pdf.

82 Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering*, 477–504. Springer Texts in Statistics https://doi.org/10.1007/978-1-4419-7787-8_18, https://faculty.washington.edu/ezivot/econ589/ch18-garch.pdf.

83 Duan, J.C. (2001). GARCH model and its application. http://jupiter.math.nctu.edu.tw/~weng/seminar/GarchApplication.pdf.

84 Diebold, F.X. (2019). Econometric data science: a predictive modeling approach. http://www.ssc.upenn.edu/~fdiebold/Textbooks.html. https://www.sas.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf.

85 Datta, S. and Granger, C. (2006). Potential to improve forecasting accuracy: advances in supply chain management. https://dspace.mit.edu/handle/1721.1/41905.

86 Datta, S. et al. (2007). Management of supply chain: an alternative modelling technique for forecasting. *Journal of the Operational Research Society* 58 (11): 1459–1469. https://doi.org/10.1057/palgrave.jors.2602419, https://dspace.mit.edu/handle/1721.1/41906.

87 Narayanan, A. (2019). How to recognize AI snake oil. https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf.

88 Li, L. et al. (2020). A survey of uncertain data management. *Frontiers of Computer Science* 14 (1): 162–190. https://doi.org/10.1007/s11704-017-7063-z, https://link.springer.com/article/10.1007/s11704-017-7063-z.

89 Agrawal, R., Imielinski, T., Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data - SIGMOD '93*, ACM Press, 1993, pp. 207–16. doi:https://doi.org/10.1145/170035.170072.

90 Coenen, F. (2011). Data mining: past, present and future. *The Knowledge Engineering Review* 26 (1): 25–29. https://doi.org/10.1017/S0269888910000378, https://www.researchgate.net/publication/220254364_Data_mining_Past_present_and_future.

91 Kahneman, D. (2012). *Thinking, Fast and Slow*. Penguin Books http://sysengr.engr.arizona.edu/OLLI/lousyDecisionMaking/KahnemanThinkingFast&Slow.pdf.

92 Akerlof, G. and Yellen, J. (1987). Rational models of irrational behavior. *American Economic Review* 77 (2): 137–142. https://notendur.hi.is/ajonsson/kennsla2003/Akerlof_Yellen.pdf.

93 https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care.

94 Minds and machines postponed / canceled - post regarding GE digital layoffs. https://www.thelayoff.com/t/VtbKYAf.

95 http://bit.ly/MRC-BERT-HAN.

96 Rumelhart, D.E. et al. (1986). Learning representations by back-propagating errors. *Nature* 323 (6088): 533–536. https://doi.org/10.1038/323533a0, https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf.

97 McCarthy, J., Minsky, M.L., Rochester, N. and Shannon, C.E. (1955). A proposal for the dartmouth summer research project on artificial intelligence (31 August 1955). http://www-formal.stanford.edu/jmc/history/dartmouth.pdf.

98 Levy, F. (2018). Computers and populism: artificial intelligence, jobs, and politics in the near term. *Oxford Review of Economic Policy* 34 (3): 393–417. https://doi.org/10.1093/oxrep/gry004, https://www.russellsage.org/sites/default/files/gry004.pdf.

99 GE minds + machines (15 November 2016). www.youtube.com/watch?v=OYn9ZtpWCUw.

100 Why robots won't take over the world. https://phys.org/news/2018-04-robots-wont-world.html.

101 Freedman, D.H. What will it take for IBM's Watson technology to stop being a dud in health care?. *MIT Tech Review*. www.technologyreview.com/s/607965/a-reality-check-for-ibms-ai-ambitions/.

102 Salmon, F. (2018). IBM's Watson was supposed to change the way we treat cancer. here's what happened instead. *Slate Magazine* (18 August). https://slate.com/business/2018/08/ibms-watson-how-the-ai-project-to-improve-cancer-treatment-went-wrong.html.

103 Perez, C.E. (2017). Why we should be deeply suspicious of backpropagation. *Medium* (13 October). https://medium.com/intuitionmachine/the-deeply-suspicious-nature-of-backpropagation-9bed5e2b085e.

104 Datta, S. (2016). *Intelligence in Artificial Intelligence*. https://arxiv.org/abs/1610.07862.

105 There's no good reason to trust Blockchain technology. *Wired*. https://www.
wired.com/story/theres-no-good-reason-to-trust-blockchain-technology/.

106 Stinchcombe, K. (2018). Blockchain is not only crappy technology but a bad
vision for the future. *Medium* (9 April). https://medium.com/@kaistinchcombe/
decentralized-and-trustless-crypto-paradise-is-actually-a-medieval-hellhole-
c1ca122efdec.

107 RuuviLab - IOTA masked authentication messaging. *RuuviLab*. https://lab.
ruuvi.com/iota.

108 Andreas, K., Fonts, A. and Prenafeta-Boldὺ, F.X. (2019). The rise of blockchain
tech in agriculture and food supply chains. https://arxiv.org/ftp/arxiv/
papers/1908/1908.07391.pdf.

109 Press, G. Big data is dead. Long live big data AI. *Forbes*. www.forbes.com/sites/
gilpress/2019/07/01/big-data-is-dead-long-live-big-data-ai/#5a262cf71b05.

110 Walsh, M.W., and Emily, F. (2019). McKinsey faces criminal inquiry over
bankruptcy case conduct. *The New York Times* (8 November). https://www.
nytimes.com/2019/11/08/business/mckinsey-criminal-investigation-
bankruptcy.html.

111 Black, E. (2001). *IBM and the Holocaust: The Strategic Alliance between Nazi
Germany and America's Most Powerful Corporation*, 1ste. Crown Publishers
http://posoh.ru/book/htm/ibm.pdf.

112 Allison, S. Huawei's pitch to African mayors: our cameras will make you safe.
The M&G Online. https://mg.co.za/
article/2019-11-15-00-our-cameras-will-make-you-safe.

113 2007 Joint Workshop on High Confidence Medical Devices, Software, and
Systems and Medical Device Plug-and-Play Interoperability: HCMDSS//MD PnP
2007: Improving Patient Safety through Medical Device Interoperability and
High Confidence Software. Proceedings: 25–27 June 2007, Cambridge,
MA. https://nam.edu/wp-content/uploads/2018/02/3.1-Goldman-
Jan-2018-002.pdf.

114 Hatcliff, J. et al. (2011). Medical application platforms – rationale, architectural
principles, and certification challenges. www.nitrd.gov/nitrdgroups/
images/8/8b/MedicalDeviceInnovationCPS.pdf.

115 Makary, M.A. and Michael, D. (2016). Medical error—the third leading cause of
death in the US. *BMJ* 353 https://doi.org/10.1136/bmj.i2139, https://www.bmj.
com/content/353/bmj.i2139.

116 Slight, S.P. et al. (2015). Meaningful use of electronic health records: experiences
from the field and future opportunities. *JMIR Medical Informatics* 3 (3): e30.
https://doi.org/10.2196/medinform.4457, https://medinform.jmir.
org/2015/3/e30/.

117 Meaningful use. *CDC*. (10 September 2019). https://www.cdc.gov/
ehrmeaningfuluse/introduction.html.

118 MIT AUTO-ID LABORATORY. https://autoid.mit.edu/about-lab.

119 Shabandri, B. and Piyush, M. (2019). Enhancing IoT security and privacy using distributed ledgers with IOTA and the tangle. *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*, IEEE, 2019, pp. 1069–75. doi:https://doi.org/10.1109/SPIN.2019.8711591 https://assets.ctfassets.net/r1dr6 vzfxhev/2t4uxvsIqk0EUau6g2sw0g/45eae33637ca92f85dd9f4a3a218e1ec/ iota1_4_3.pdf.

120 The Coordicide. https://files.iota.org/papers/Coordicide_WP.pdf.

121 https://www.iota.org/.

122 Send IoT data to the IOTA tangle with SAP HANA XSA and analytics cloud. https://blogs.sap.com/2019/10/08/ send-iot-data-to-the-iota-tangle-with-sap-hana-xsa-and-analytics-cloud/.

123 https://www.dds-foundation.org/.

124 Shepard, M, Baicker, K., and Skinner, J.S. (2019). *Does one medicare fit all? The economics of uniform health insurance benefits* in tax policy and the economy. *Moffitt* 34 https://doi.org/10.3386/w26472, https://www.nber.org/papers/ w26472.pdf.

125 http://bit.ly/Economics-of-Technology.

126 http://bit.ly/COASE5PAPERS.

127 Merelli, A. The inventor of microfinance has an idea for fixing capitalism. *Quartz.* https://qz.com/1089266/ the-inventor-of-microfinance-has-an-idea-for-fixing-capitalism/.

128 Georgescu, C. (2012). Simulating micropayments in local area networks. *Procedia - Social and Behavioral Sciences* 62: 30–34. https://doi.org/10.1016/j. sbspro.2012.09.007.

129 Spence, M. (2011). The next convergence: the future of economic growth in a multispeed world. *Farrar, Straus and Giroux.* http://pubdocs.worldbank.org/ en/515861447787792966/DEC-Lecture-Series-Michael-Spence- Presentation.pdf.

130 Tabassi, E. et al. (2019). *A Taxonomy and Terminology of Adversarial Machine Learning. NIST IR 8269-draft*. National Institute of Standards and Technology https://doi.org/10.6028/NIST.IR.8269-draft, https://nvlpubs.nist.gov/nistpubs/ ir/2019/NIST.IR.8269-draft.pdf.

131 Duddu, V. (2018). A survey of adversarial machine learning in cyber warfare. *Defence Science Journal* 68 (4): 356. https://doi.org/10.14429/dsj.68.12371.

132 Morgan, V., Casso-Hartman, L., Bahamon-Pinzon, D. et al. (2020). Sensor-as-a- service: convergence of sensor analytic point solutions (SNAPS) and Pay-A- Penny-Per-Use (PAPPU) paradigm as a catalyst for democratization of healthcare in underserved communities. *Diagnostics* 10: 22; doi:https://doi. org/10.3390/diagnostics10010022. MIT Library SNAPS -. https://dspace.mit.edu/ handle/1721.1/111021.https://dspace.mit.edu/handle/1721.1/123983

133 Datta, S. et al. (2004). Adaptive value networks: convergence of emerging tools, technologies and standards as catalytic drivers. In: *Evolution of Supply Chain Management: Symbiosis of Adaptive Value Networks and ICT* (eds. Y.S.,.C. et al.). Kluwer Academic Publishers https://dspace.mit.edu/handle/1721.1/41908.

134 Silvestro, R. and Paola, L. (2014). Integrating financial and physical supply chains: the role of banks in enabling supply chain integration. *International Journal of Operations & Production Management* 34 (3): 298–324. https://doi.org/10.1108/IJOPM-04-2012-0131, https://www.emerald.com/insight/content/doi/10.1108/IJOPM-04-2012-0131/full/html.

135 Banerjee, A.V. and Duflo, E. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, 1ste. PublicAffairs https://warwick.ac.uk/about/london/study/warwick-summer-school/courses/macroeconomics/poor_economics.pdf.

136 Lie, R. (2019). Robertlie/Dht11-Raspi3. 2018. *GitHub*. https://github.com/robertlie/dht11-raspi3.

137 Handy, P. (2018). Introducing masked authenticated messaging. *Medium* (9 April). https://blog.iota.org/introducing-masked-authenticated-messaging-e55c1822d50e.

138 Kazimirova, E.D. (2017). Human-centric internet of things. Problems and challenges. https://www.researchgate.net/publication/319059870_Human-Centric_Internet_of_Things_Problems_and_Challenges.

139 Calderon, M.A. et al. (2016). A more human-centric internet of things with temporal and spatial context. *Procedia Computer Science* 83: 553–559. https://doi.org/10.1016/j.procs.2016.04.263, https://www.sciencedirect.com/science/article/pii/S1877050916302964.

140 McLamore, E.S., Datta, S.P.A., Morgan, V. et al. (2019). SNAPS: sensor analytics point solutions for detection and decision support. *Sensors* 19 (22): 4935. https://www.mdpi.com/1424-8220/19/22/4935/pdf.

141 The sensor-based economy. *Wired* (January 2017). https://www.wired.com/brandlab/2017/01/sensor-based-economy/.

142 Pont, S. (ed.) (2013). *Digital State: How the Internet Is Changing Everything*. Kogan Page.

143 Leading the IoT: gartner insights on how to lead in a connected world. https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf.

144 Evans, D. (2011). The internet of things: how the next evolution of the internet is changing everything. https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf.

145 Puglia, D. (2014). Are enterprises ready for billions of devices to join the Internet?. *Wired* (December 2014). https://www.wired.com/insights/2014/12/enterprises-billions-of-devices-internet/.

146 Gartner Says 5.8 Billion enterprise and automotive IoT endpoints will be in use in 2020. https://www.gartner.com/en/newsroom/press-releases/2019-08-29-gartner-says-5-8-billion-enterprise-and-automotive-io.

147 www.cisco.com/c/dam/en/us/products/collateral/se/internet-of-things/at-a-glance-c45-731471.pdf.

148 The Internet of things: sizing up the opportunity. *McKinsey*. https://www.mckinsey.com/industries/semiconductors/our-insights/the-internet-of-things-sizing-up-the-opportunity.

149 IoT overview handbook: 2019 background primer on the topics & technologies driving the Internet of things. *Postscapes*. https://www.postscapes.com/iot/.

150 Fleming, M., Clarke, W., Das, S., Phongthiengtham, P., and Reddy, P. (2019). The future of work: how new technologies are transforming tasks (31 October 2019). https://mitibmwatsonailab.mit.edu/research/publications/paper/download/The-Future-of-Work-How-New-Technologies-Are-Transforming-Tasks.pdf.

151 Andrew, N.G. Why AI is the new electricity. https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity.

152 Cheekiralla, S. and Engels, D.W. (2006). An IPv6-based identification scheme. *2006 IEEE International Conference on Communications* 1: 281–286. https://doi.org/10.1109/ICC.2006.254741, https://ieeexplore.ieee.org/document/4024131.

153 Stallings, W. (1996). IPv6: the new Internet protocol. *IEEE Communications Magazine* 34 (7): 96–108. https://doi.org/10.1109/35.526895, https://ieeexplore.ieee.org/document/526895.

154 Difference between IPv4 and IPv6. *Tech Differences* (4 August 2017). https://techdifferences.com/difference-between-ipv4-and-ipv6.html

155 Datta, S.P.A. (2012). An unified theory of relativistic identification of information in the systems age: proposed convergence of unique identification with syntax and semantics through internet protocol version 6 (IPv6). *International Journal of Advanced Logistics* 1 (1): 66–82. https://doi.org/10.1080/2287108X.2012.11006070, https://dspace.mit.edu/handle/1721.1/41902.

156 Datta, S. (2011). Mobile eVote as an IPv6 App. (23 May 2011). https://shoumendatta.wordpress.com/2011/05/23/mobile-e-vote-ipv6-app/. https://dspace.mit.edu/bitstream/handle/1721.1/41902/IPv6%20Apps%20and%20SaaS.pdf?sequence=11&isAllowed=y.

157 https://bentley.umich.edu/elecrec/d/duderstadt/Speeches/JJDS6/jjd1341.pdf.

158 Figueroa, A. (2019). Data demystified — DIKW model. *Medium* (24 May 2019). https://towardsdatascience.com/rootstrap-dikw-model-32cef9ae6dfb.

159 https://www.wur.nl/en/Education-Programmes/wageningen-academy-1/What-we-offer-you/Courses/show-1/Course-Towards-Data-driven-Agri-Food-Business-1.htm.

160 CS 540 lecture notes: intelligent agents. http://pages.cs.wisc.edu/~dyer/cs540/notes/agents.html.

161 Maes, P. (1997). Intelligent Software. *Proceedings of the 2nd International Conference on Intelligent User Interfaces - IUI'97*, ACM Press, pp. 41–43. doi:https://doi.org/10.1145/238218.238283.

162 Paolucci, M. and Sacile, R. (2005). *Agent-Based Manufacturing and Control Systems: New Agile Manufacturing Solutions for Achieving Peak Performance*. CRC Press http://jmvidal.cse.sc.edu/library/paolucci05a.pdf.

163 Datta, S. (2017). Cybersecurity: agents based approach?. https://dspace.mit.edu/handle/1721.1/107988.

164 White, F.E. (1997). Data fusion group. https://apps.dtic.mil/dtic/tr/fulltext/u2/a394662.pdf.

165 Rafferty, E.R.S., et al. (2019). Seeking the optimal schedule for chickenpox vaccination in Canada: using an agent-based model to explore the impact of dose timing, coverage and waning of immunity on disease outcomes. *Vaccine* (November 2019). doi:https://doi.org/10.1016/j.vaccine.2019.10.065.

166 Lassila, O. and Swick, R.R. (1999). Resource description framework (RDF) model and syntax specification. *W3C working draft* (February 1999). www.w3.org/TR/REC-rdf-syntax/.

167 Berners-Lee, T., Hendler, J., and Lassila, O. (2001). *The Semantic Web*. Scientific American.

168 Shi, L. et al. (2017). Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services. *BioMed Research International* 2017: 1–12. https://doi.org/10.1155/2017/2858423, http://downloads.hindawi.com/journals/bmri/2017/2858423.pdf.

169 Chakraborty, A., Munshi, S. and Mukhopadhyay, D. (2013). Searching and establishment of S-P-O relationships for linked RDF graphs: an adaptive approach. 2013 Int Conf on Cloud & Ubiquitous Comp & Emerging Tech. https://arxiv.org/ftp/arxiv/papers/1311/1311.7200.pdf.

170 Chakraborty, A., Munshi, S. and Mukhopadhyay, D. (2013). A proposal for the characterization of multidimensional inter-relationships of RDF graphs based on set theoretic approach. https://arxiv.org/ftp/arxiv/papers/1312/1312.0001.pdf.

171 Buneman, P. (1974). A characterisation of rigid circuit graphs. *Discrete Mathematics* 9 (3): 205–212. https://doi.org/10.1016/0012-365X(74)90002-8, http://homepages.inf.ed.ac.uk/opb/homepagefiles/phylogeny-scans/rigidcircuitgraphs.pdf.

172 Sharma, C. and Roopak, S. (2019). A schema-first formalism for labeled property graph databases: enabling structured data loading and analytics. *Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies - BDCAT'19*, ACM Press, 2019, pp. 71–80 doi:https://doi.org/10.1145/3365109.3368782.

173 Wang, D.-W. et al. (2019). CK-modes clustering algorithm based on node cohesion in labeled property graph. *Journal of Computer Science and Technology* 34 (5): 1152–1166. https://doi.org/10.1007/s11390-019-1966-0.

174 https://aibusiness.com/ending-the-rdf-vs-property-graph-debate-with-rdf/.

175 Tarjan, R. (1972). Depth-first search and linear graph algorithms. *SIAM Journal on Computing* 1 (2): 146–160. https://doi.org/10.1137/0201010.

176 Fleischer, R, and Gerhard, T. (2003). Experimental studies of graph traversal algorithms. In: *Experimental and Efficient Algorithms* (ed K. Jansen, et al), vol. 2647, Berlin Heidelberg: Springer, pp. 120–133 doi:https://doi.org/10.1007/3-540-44867-5_10.

177 Noy, N. et al. (2019). Industry-scale knowledge graphs: lessons and challenges. *Communications of the ACM* 62 (8): 36–43. https://doi.org/10.1145/3331166, https://queue.acm.org/detail.cfm?id=3332266.

178 https://tech.ebayinc.com/engineering/akutan-a-distributed-knowledge-graph-store/.

179 https://www.w3.org/TR/owl-guide/.

180 Sengupta, K. and Hitzler, P. (2014). Web Ontology Language (OWL). In: *Encyclopedia of Social Network Analysis & Mining* (eds. R. Alhajj and J. Rokne). NY: Springer https://doi.org/10.1007/978-1-4614-6170-8.

181 http://vowl.visualdataweb.org/v2/.

182 https://tools.ietf.org/html/rfc3987.

183 www.eads-iw.net/web/nfigay.

184 Carbone, L., et al. (2017) State of play of interoperability: report 2016. *Open WorldCat*. http://dx.publications.europa.eu/10.2799/969314 https://ec.europa.eu/isa2/sites/isa/files/docs/publications/report_2016_rev9_single_pages.pdf.

185 Blind men and the elephant. www.allaboutphilosophy.org/blind-men-and-the-elephant.htm.

186 Bohr, N. (1950). On the notions of causality and complementarity. *Science* 111 (2873): 51–54. https://doi.org/10.1126/science.111.2873.51.

187 Morrison, A. (2019). Is data science/machine learning/AI overhyped right now?. www.quora.com/Is-data-science-machine-learning-AI-overhyped-right-now/answer/Alan-Morrison.

188 Fuller, A., et al. (2019). Digital twin: enabling technology, challenges and open research. October 2019. http://arxiv.org/abs/1911.01276. https://arxiv.org/ftp/arxiv/papers/1911/1911.01276.pdf.

189 https://newsstand.joomag.com/en/iic-journal-of-innovation-12th-edition/0994713001573661267.

190 https://www.forbes.com/sites/bernardmarr/2017/03/06/what-is-digital-twin-technology-and-why-is-it-so-important.

191 Vilfredo, P. (1971). *Manual of Political Economy*. Translated by (ed. A.M. Kelley). MIT Press.

192 McLamore, E.S., Huffaker, R., Shupler, M., Ward, K., Austin Datta, S.P., Katherine Banks, M., Casaburi, G., Babilonia, J., Foster, J.S. (2019). Digital proxy of a bio-reactor (DIYBOT) combines sensor data and data analytics for wastewater treatment and wastewater management systems. (Nature Scientific Reports, in press) Draft copy of "DIYBOT" available from MIT Libraries. https://dspace.mit.edu/handle/1721.1/123983.

193 Hernlund, E. et al. (2013). Osteoporosis in the European Union: medical management, epidemiology and economic Burden: a report prepared in collaboration with the international osteoporosis foundation (IOF) and the European federation of pharmaceutical industry associations (EFPIA). *Archives of Osteoporosis* 8 (1–2): 136. https://doi.org/10.1007/s11657-013-0136-1, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3880487/pdf/11657_2013_Article_136.pdf.

194 Mithal, A., Dhingra, V. and Lau, E. (2009). The Asian Audit Epidemiology, costs and burden of osteoporosis in Asia 2009. http://www.iofbonehealth.org/.

195 www.iofbonehealth.org/sites/default/files/PDFs/Audit%20Eastern%20Europe_Central%20Asia/Russian_Audit-Bulgaria.pdf.

196 www.iofbonehealth.org/sites/default/files/PDFs/Audit%20Asia/Asian_regional_audit_Indonesia.pdf.

197 www.iofbonehealth.org/sites/default/files/PDFs/Audit%20Asia/Asian_regional_audit_India.pdf.

198 https://www.iofbonehealth.org/facts-statistics.

199 Global bone density test market grows substantially by 2023, asserts MRFR unleashing the forecast for 2017–2023. *Reuters*. www.reuters.com/brandfeatures/venture-capital/article?id=57229.

200 El Maghraoui, A. et al. (2009). Bone mineral density of the spine and femur in a group of healthy moroccan men. *Bone* 44 (5): 965–969. https://doi.org/10.1016/j.bone.2008.12.025.

201 Angell, M. (2005). The truth about the drug companies: how they deceive us and what to do about it. *Random House Trade Paperbacks* (2005). https://cyber.harvard.edu/cyberlaw2005/sites/cyberlaw2005/images/NYReviewBooksAngell.pdf.

202 https://www.keele.ac.uk/pharmacy-bioengineering/ourpeople/jamesnolan/.

203 https://people.eecs.berkeley.edu/~pabbeel/.

204 https://github.com/alirezadir/Production-Level-Deep-Learning.

205 Ritchie, H. (2018). What do people die from?. https://ourworldindata.org/what-does-the-world-die-from.

206 https://www.who.int/nmh/publications/ncd_report_chapter1.pdf.

207 Mendez, K.M., Broadhurst, D.I., and Reinke, S.N. (2019). The application of artificial neural networks in metabolomics: a historical perspective. *Metabolomics* 15 (11): 142. https://doi.org/10.1007/s11306-019-1608-0.

208 http://bit.ly/Farm-IoT-Ranveer.

209 https://www.theaaih.org/.

210 Gutterman, D.J. (2009). Silent myocardial ischemia. *Circulation Journal* 73: 785–797. https://www.jstage.jst.go.jp/article/circj/73/5/73_CJ-08-1209/_pdf.

211 Institute for Health Metrics and Evaluation, Human Development Network, The World Bank (2013). *The Global Burden of Disease: Generating Evidence, Guiding Policy — Sub-Saharan Africa Regional Edition*. Seattle, WA: IHME http://documents.worldbank.org/curated/en/831161468191672519/pdf/808520PUB0ENGL0Box0379820B00PUBLIC0.pdf.

212 World Health Organization (2009). *Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks*. ISBN: 978 92 4 156387 1 https://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf.

213 https://www.who.int/pmnch/media/press_materials/fs/fs_mdg4_childmortality/en/.

214 https://economictimes.indiatimes.com/magazines/panache/jack-dorsey-has-fallen-in-love-with-africa-plans-to-shift-there-for-3-6-months-next-year/articleshow/72275949.cms.

215 The concept of "bit dribbling" may not be new but may be attributed to Neil Gershenfeld (MIT) circa 1999 (*personal communication*). The notion is that very small amounts of data and information (bits) transmitted (dribbled) at the right time, are often enough to serve 80% of cases, in several circumstances.

216 Parker, M. (2018). Shut down the business school. https://www.theguardian.com/news/2018/apr/27/bulldoze-the-business-school.

217 Hedengren, J.D. Drilling automation and downhole monitoring with physics based models. https://apm.byu.edu/prism/index.php/Members/JohnHedengren.

218 https://apmonitor.com/do/index.php/Main/ShortCourse.

219 Osgood, N. (2019). Systems data science. https://www.youtube.com/watch?v=CPUOyqs9G3Q&feature=youtu.be.

220 Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* 53 (0): 370–418. http://www.rssb.be/bsn57/bsn57-6.pdf.

221 Eid, M.A. et al. (2016). A novel eye-gaze-controlled wheelchair system for navigating unknown environments: case study with a person with ALS. *IEEE Access* 4: 558–573. https://doi.org/10.1109/ACCESS.2016.2520093.

222 Robots for lonely hearts – Asian robotics review. https://asianroboticsreview.com/home29-html.

223 Oppenheimer, A. and Fitz, E.E. (2019). *The Robots Are Coming! The Future of Jobs in the Age of Automation*. Vintage Books, a division of Penguin Random House. ISBN: ISBN-13 978-0525565000.

224 Deming, D. (2020). The robots are coming. prepare for trouble. *The New York Times* (30 January 2020). https://www.nytimes.com/2020/01/30/business/artificial-intelligence-robots-retail.html.

225 http://www.philipkdickfans.com/mirror/websites/pkdweb/short_stories/Autofac.htm.

226 https://www.bloomberg.com/news/features/2017-10-18/this-company-s-robots-are-making-everything-and-reshaping-the-world.

227 Ford, M. (2016). *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books.

228 Pugliano, J. (2017). *The Robots Are Coming: A Human's Survival Guide to Profiting in the Age of Automation*. Ulysses Press.

229 The Dutch Safety Board (2009). Crashed during approach, Boeing 737-800, near Amsterdam Schiphol Airport. 25 February 2009. https://catsr.vse.gmu.edu/SYST460/TA1951_AccidentReport.pdf.

230 Pacaux-Lemoine, M.-P. and Flemisch, F. (2016). Layers of shared and cooperative control, assistance and automation. *IFAC-Papers OnLine* 49 (19): 159–164. https://doi.org/10.1016/j.ifacol.2016.10.479.

231 Wickens, C.D. et al. (2010). Stages and levels of automation: an integrated meta-analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54 (4): 389–393. https://doi.org/10.1177/154193121005400425.

232 Hamby, C. (2020). How boeing's responsibility in a deadly crash 'got buried. *The New York Times* (20 January 2020). https://www.nytimes.com/2020/01/20/business/boeing-737-accidents.html.

233 Hamby, C. and Moses, C. (2020). Boeing refuses to cooperate with new inquiry into deadly crash. *The New York Times* (6 February 2020). www.nytimes.com/2020/02/06/business/boeing-737-inquiry.html.

234 Solow, R.M. (1957). Technical change and the aggregate production function. *The Review of Economics and Statistics* 39 (3): 312. https://doi.org/10.2307/1926047, http://www.piketty.pse.ens.fr/files/Solow1957.pdf.

235 http://ghdx.healthdata.org/.

236 Howes, A. (2020). *Arts and Minds: How the Royal Society of Arts Changed a Nation*. Princeton University Press. ISBN: 9780691182643.

237 Endo, N., Ghaeli, N., Duvallet, C. et al. (2020). Rapid assessment of opioid exposure and treatment in cities through robotic collection and chemical analysis of wastewater. *Journal of Medical Toxicology* 16: 195–203. https://doi.org/10.1007/s13181-019-00756-5.

238 https://www.oxfordeconomics.com/recent-releases/how-robots-change-the-world.

239 Kasprzyk-Hordern, B. and Baker, D.R. (2012). Enantiomeric profiling of chiral drugs in wastewater and receiving waters. *Environmental Science & Technology* 46 (3): 1681–1691. https://doi.org/10.1021/es203113y, https://core.ac.uk/reader/161908112.

240 Matus, M. et al. (2019). 24-hour multi-omics analysis of residential sewage reflects human activity and informs public health. *Genomics* https://doi.org/10.1101/728022.

241 Wu, F. et al. (2020). SARS-CoV-2 titers in wastewater are higher than expected from clinically confirmed cases. *Infectious Diseases (except HIV/AIDS)* https://doi.org/10.1101/2020.04.05.20051540.

242 Wu, F. et al. (2020). SARS-CoV-2 titers in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. *Infectious Diseases (except HIV/AIDS)* https://doi.org/10.1101/2020.06.15.20117747.

243 Exceptions prove the rule. Not all Indians espouse the "body shop" mantra. Global business and tech leaders (2020) of Indian origin include Mrs Jayashree Ullal (CEO, Arista Networks), Mr Arvind Krishna (CEO, IBM, incoming), Mr Sundar Pichai (CEO, Alphabet/Google), Mr Satya Nadella (CEO, Microsoft), Mr Shantanu Narayen (CEO, Adobe), Mr Rajeev Suri (CEO, Nokia, outgoing), Mr V K Narasimhan (CEO, Novartis) and Mr Ajaypal Singh Banga (CEO, MasterCard). The market cap of these 8 companies (~$3T) may be 20% of the total market cap of the top 50 US companies (~$15T). www.iweblists.com/us.

244 Sabrie, G. (2019). Behind the rise of China's facial-recognition giants (03 September 2019). https://www.wired.com/story/behind-rise-chinas-facial-recognition-giants/.

245 https://www.purdue.edu/rosehub/.

246 Radjou, N., Prabhu, J. (2016). Frugal innovation: how to do more with less. http://naviradjou.com/book/frugal-innovation-how-to-do-more-with-less/.

247 Creighton, J. (2019). 5 Moon-landing innovations that changed life on Earth. https://theconversation.com/5-moon-landing-innovations-that-changed-life-on-earth-102700.

248 https://www.inc.com/bill-murphy-jr/27-innovations-we-use-constantly-but-that-you-probably-didnt-know-were-from-nasa-space-program.html.

249 https://spinoff.nasa.gov/.

250 Marcus, G. (2019). An epidemic of AI misinformation. https://thegradient.pub/an-epidemic-of-ai-misinformation/.

251 Vanegas, D.C., Patiño, L., Mendez, C. et al. (2018). Laser scribed graphene biosensor for detection of biogenic amines in food samples using locally sourced materials. *Biosensors* 8 (2): 42. https://doi.org/10.3390/bios8020042, https://www.mdpi.com/2079-6374/8/2/42.

252 Holmes, E.A. et al. (2016). Applications of time-series analysis to mood fluctuations in bipolar disorder to promote treatment innovation: a case series. *Translational Psychiatry* 6 (1): e720–e720. https://doi.org/10.1038/tp.2015.207, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5068881/pdf/tp2015207a.pdf.

253 Trautmann, S. et al. (2016). The economic costs of mental disorders: do our societies react appropriately to the burden of mental disorders? *EMBO Reports*

17 (9): 1245–1249. https://doi.org/10.15252/embr.201642951, www.ncbi.nlm.nih.gov/pmc/articles/PMC5007565/pdf/EMBR-17-1245.pdf.

254 Drevets, W.C. et al. (1997). Subgenual prefrontal cortex abnormalities in mood disorders. *Nature* 386 (6627): 824–827. https://doi.org/10.1038/386824a0.

255 Llamocca, P., Junestrand, A., Cukic, M., Urgelés, D., López, V. (2018). Data source analysis in mood disorder research. *XVIII Proceedings of the XVIII Conference of the Spanish Association for Artificial Intelligence.* Granada, Spain: CAEPIA. ISBN: 978-84-09-05643-9 F. Herrera et al. (Eds.), pp. 893–900.

256 Llamocca, P., Urgelés, D., Cukic, M., Lopez, V. (2019). Bip4Cast: Some advances in mood disorders data analysis. *Proceedings of the 1st International Alan Turing Conference on Decision Support and Recommender Systems*, London 2019.

257 Perrault, R., Shoham, Y., Brynjolfsson, E. et al. (2019). *The AI Index 2019 Annual Report*. Stanford, CA: AI Index Steering Committee, Human-Centered AI Institute, Stanford University https://hai.stanford.edu/sites/g/files/sbiybj10986/f/ai_index_2019_report.pdf.

258 Chen, Y.H. et al. (2019). Eyeriss v2: a flexible accelerator for emerging deep neural networks on mobile devices. (May 2019). http://arxiv.org/abs/1807.07928 ■ https://arxiv.org/pdf/1807.07928.pdf. https://www.rle.mit.edu/eems/wp-content/uploads/2019/04/2019_jetcas_eyerissv2.pdf.

259 Jo, E.S., and Timnit, G. (2019). Lessons from archives: strategies for collecting sociocultural data in machine learning. (December 2019) doi:https://doi.org/10.1145/3351095.3372829. https://arxiv.org/pdf/1912.10389.pdf.

260 Tembon, M.M., and Lucia, F. (2008). Girl's education in the 21st Century: gender equality, empowerment and growth. *The World Bank*. doi:https://doi.org/10.1596/978-0-8213-7474-0.

261 https://www.nytimes.com/2006/10/14/world/asia/14nobel.html.

262 Cohn, A. et al. (2019). Civic honesty around the Globe. *Science* 365 (6448): 70–73. https://doi.org/10.1126/science.aau8712.

263 Grosch, K., Rau, H. (2017). Gender differences in honesty: the role of social value orientation. Discussion Papers, No. 308, University of Göttingen, Center for European, Governance and Economic Development Research (CEGE), Göttingen, Germany. https://www.econstor.eu/bitstream/10419/156226/1/882555200.pdf.

264 Sen, A. (2009). *The Idea of Justice*. Harvard University Press https://dutraeconomicus.files.wordpress.com/2014/02/amartya-sen-the-idea-of-justice-2009.pdf.

265 Artificial intelligence makes bad medicine even worse. *Wired*. www.wired.com, https://www.wired.com/story/artificial-intelligence-makes-bad-medicine-even-worse/.

266 Datta, S. (2008). Arm chair essays in energy. https://dspace.mit.edu/handle/1721.1/45512.

267 Mackenzie, D. (1877). *The Flooding of the Sahara: An Account of the Proposed Plan for Opening Central Africa to Commerce and Civilization* (ed. S.L. Marston). Searle & Rivington https://ia902205.us.archive.org/23/items/floodingsaharaa01mackgoog/floodingsaharaa01mackgoog.pdf.

268 https://www.jstor.org/stable/1761255?seq=1#metadata_info_tab_contents.

269 https://www.nature.com/articles/019509a0.pdf.

270 Singh, L. et al. (2020). Bioelectrofuel synthesis by nanoenzymes: novel alternatives to conventional enzymes. *Trends in Biotechnology* 38 (5): S0167779919303129. https://doi.org/10.1016/j.tibtech.2019.12.017.

271 Rothman, J. (2020). The equality conundrum. (13 January 2020). www.newyorker.com/magazine/annals-of-inquiry. https://www.newyorker.com/magazine/2020/01/13/the-equality-conundrum.

272 https://www.mikespastry.com/.

273 https://www.modernpastry.com/.

274 Ilic, S. et al. (2018). Deep contextualized word representations for detecting sarcasm and irony. *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Association for Computational Linguistics, pp. 2–7: doi:https://doi.org/10.18653/v1/W18-6202. https://arxiv.org/pdf/1809.09795.pdf.

275 Datta, S. (2008). Convergence of bio, info, nano, eco: global public goods and economic growth. https://dspace.mit.edu/handle/1721.1/41909.

276 Steinmueller, W.E. (2010). Economics of technology policy. In: *Handbook of the Economics of Innovation*, 1ste, vol. 2 (eds. B.H. Hall and N. Rosenberg), 1181–1218. Elsevier https://doi.org/10.1016/S0169-7218(10)02012-5.

277 Wang, L., Chen, Y., Long, F. et al. (2020). Breaking the loop: tackling homoacetogenesis by chloroform to halt hydrogen production-consumption loop in single chamber microbial electrolysis cells. *Chemical Engineering Journal* https://doi.org/10.1016/j.cej.2020.124436.

278 https://patents.justia.com/inventor/andrew-e-fano.

279 https://patents.justia.com/patent/10095981.

280 https://claimparse.com/patent.php?patent_num=10275721.

281 https://www.ibm.com/downloads/cas/KMPVGB4W.

282 https://mcgovern.mit.edu/profile/tomaso-poggio/.

283 Fjelland, R. (2020). https://doi.org/10.1057/s41599-020-0494-4). Why general artificial intelligence will not be realized. *Humanit Soc Sci Commun* 7: 10. www.nature.com/articles/s41599-020-0494-4.pdf.

284 Faro, S.H. and Mohamed, F.B. (eds.) (2006). *Functional MRI: Basic Principles and Clinical Applications*. Springer. ISBN: 978-0-387-23046-7.

285 Suleiman Khan (2019). BERT, RoBERTa, DistilBERT, XLNet — which one to use. https://towardsdatascience.com/bert-roberta-distilbert-xlnet-which-one-to-use-3d5ab82ba5f8.

286 https://cloud.google.com/tpu/.

287 FloydHub Blog (2019). When not to choose the best NLP model. https://blog. floydhub.com/when-the-best-nlp-model-is-not-the-best-choice/.

288 Brown, T.B. et al. (2020). Language models are few-shot learners. *ArXiv* 2005: 14165 [Cs]. http://arxiv.org/abs/2005.14165. https://arxiv.org/pdf/2005. 14165.pdf.

289 Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L. (2018). Deep contextualized word representations. https://arxiv. org/pdf/1802.05365.pdf.

290 Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. https://arxiv.org/ pdf/1810.04805.pdf.

291 Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P. and Soricut, R. (2020). Albert: a lite bert for self-supervised learning of language representations. *8th International Conference on Learning Representations (2020)*. https://openreview. net/pdf?id=H1eA7AEtvS.

292 Sun, C. et al. (2019). A deep learning approach with deep contextualized word representations for chemical–protein interaction extraction from biomedical literature. *IEEE Access* 7: 151034–151046. https://doi.org/10.1109/ ACCESS.2019.2948155.

293 Tshitoyan, V., Dagdelen, J., Weston, L. et al. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* 571: 95–98. https://doi.org/10.1038/s41586-019-1335-8.

294 Pelillo, M. (2014). Alhazen and the nearest neighbor rule. *Pattern Recognition Letters* 38 (2014): 34–37. https://doi.org/10.1016/j.patrec.2013.10.022.

295 Babyak, M.A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine* 66 (3): 411–421. https://doi.org/10.1097/01. psy.0000127692.23278.a9, https://people.duke.edu/~mababyak/papers/ babyakregression.pdf.

296 Hamed, T. (2017). Recursive feature addition: a novel feature selection technique, including a proof of concept in network security. PhD thesis submitted to University of Guelph, Ontario, Canada. https://atrium.lib. uoguelph.ca/xmlui/bitstream/handle/10214/10315/Hamed_Tarfa_201704_PhD. pdf?sequence=1&isAllowed=y.

297 Nandy, A., Duan, C., Janet, J.P. et al. (2018). Strategies and software for machine learning accelerated discovery in transition metal chemistry. *Industrial and Engineering Chemistry Research* 57 (42): 13973–13986. https://doi.org/10.1021/ acs.iecr.8b04015.

298 Virshup, A.M. et al. Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds. *Journal of the*

American Chemical Society 135 (19): 7296–7303. https://doi.org/10.1021/ja401184g.

299 Lemonick, S. (2019). As DFT matures, will it become a push-button technology? *Chemical & Engineering News* 97 (35) https://cen.acs.org/physical-chemistry/computational-chemistry/DFT-matures-become-push-button/97/i35.

300 Bruna, J., Zaremba, W., Szlam, A. and LeCun, Y. (2014). Spectral networks and locally connected networks on graphs. http://arxiv.org/abs/1312.6203. https://arxiv.org/pdf/1312.6203.pdf.

301 Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O. and Dahl, G.E. (2017). Neural message passing for quantum chemistry. https://arxiv.org/pdf/1704.01212.pdf.

302 Swanson, K. (2019). Message passing neural networks for molecular property prediction. Master's Thesis. EECS, MIT. https://dspace.mit.edu/bitstream/handle/1721.1/123133/1128814048-MIT.pdf?sequence=1&isAllowed=y.

303 http://chemprop.csail.mit.edu/.

304 Ruddigkeit, L., van Deursen, R., Blum, L.C., and Reymond, J.-L. (2012). Enumeration of 166 billion organic small molecules in the chemical Universe database GDB-17. *Journal of Chemical Information and Modeling* 52 (11): 2864–2875. https://doi.org/10.1021/ci300415d.

305 Stokes, J.M., Yang, K., Swanson, K. et al. (2020). A deep learning approach to antibiotic discovery. *Cell* 180 (4): 688–702.e13. https://doi.org/10.1016/j.cell.2020.01.021, https://www.cell.com/cell/pdf/S0092-8674(20)30102-1.pdf.

306 http://news.mit.edu/2020/artificial-intelligence-identifies-new-antibiotic-0220.

307 Zhang, X. et al. (2019). Deep neural network hyperparameter optimization with orthogonal array tuning. (July 2019). http://arxiv.org/abs/1907.13359. https://arxiv.org/pdf/1907.13359.pdf.

308 Hansen, L.K. and Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (10) https://pdfs.semanticscholar.org/257d/c8ae2a8353bb2e86c1b7186e7d989fb433d3.pdf.

309 https://thenextweb.com/neural/2020/02/19/study-ai-expert-gary-marcus-explains-how-to-take-ai-to-the-next-level/.

310 Russell, S. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*, 3rde. Prentice Hall http://aima.cs.berkeley.edu/.

311 Winston, P.H. and Brown, R.H. (eds.) (1979). *Artificial Intelligence, an MIT Perspective*. MIT Press https://courses.csail.mit.edu/6.034f/ai3/rest.pdf.

312 Buchanan, B.G. and Shortliffe, E.H. (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley http://digilib.stmik-banjarbaru.ac.id/data.bc/2.%20AI/2.%20AI/1984%20Rule-Based%20Expert%20Systems.pdf.

313 Quinlan, J.R. (1987). *Applications of Expert Systems: Based on the Proceedings of the Second Australian Conference*, 1987. Turing Institute Press in association with Addison-Wesley Pub. Co.

314 Giarratano, J.C. and Riley, G. (1989). *Expert Systems: Principles and Programming*, 1ste. Boston: PWS-Kent Publishing Company. ISBN: 0-87835-335-6.

315 Buchanan, B.G. (1989). Can machine learning offer anything to expert systems? *Machine Learning* 4 (3–4): 251–254. https://doi.org/10.1007/BF00130712, https://link.springer.com/content/pdf/10.1007/BF00130712.pdf.

316 Valverde, S. (2016). Major transitions in information technology. *Philosophical Transactions of the Royal Society B* 371: 20150450. http://dx.doi.org/10.1098/rstb.2015.0450, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4958943/pdf/rstb20150450.pdf.

317 Mearian, L. (2017). CW@50: data storage goes from \$1M to 2 cents per gigabyte (+video). *Computerworld* (23 March 2017). https://www.computerworld.com/article/3182207/cw50-data-storage-goes-from-1m-to-2-cents-per-gigabyte.html.

318 Supernor, B. (2018). Why the cost of cloud computing is dropping dramatically. *App Developer Magazine*. https://appdevelopermagazine.com/why-the-cost-of-cloud-computing-is-dropping-dramatically/.

319 Wolpert, D.H. (2018). Why do computers use so much energy?. *Scientific American Blog Network*. https://blogs.scientificamerican.com/observations/why-do-computers-use-so-much-energy/.

320 Wolpert, D.H. (2019). The stochastic thermodynamics of computation. *Journal of Physics A: Mathematical and Theoretical* 2019 https://iopscience.iop.org/article/10.1088/1751-8121/ab0850/pdf.

321 Panesar, S.S., Kliot, M., Parrish, R. et al. (2019). Promises and perils of artificial intelligence in neurosurgery. *Neurosurgery* https://doi.org/10.1093/neuros/nyz471.

322 Hanhauser, E., Bono, M.S. Jr., Vaishnav, C. et al. (2020). Solid-phase extraction, preservation, storage, transport, and analysis of trace contaminants for water quality monitoring of heavy metals. *Environmental Science & Technology* https://doi.org/10.1021/acs.est.9b04695.

323 Sharma, R., Kamble, S.S., Gunasekaran, A. et al. (2020). A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Computers and Operations Research* https://doi.org/10.1016/j.cor.2020.104926.

324 Gyrard, A., Gaur, M., Shekarpour, S., Thirunarayan, K. and Sheth, A. (2018). Personalized health knowledge graph. *Contextualized Knowledge Graph Workshop, International Semantic Web Conference*, 2018. https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=1005&context=aii_fac_pub.

325 Apple agrees to settlement of up to \$500 million from lawsuit alleging it throttled older phones. *TechCrunch*. http://social.techcrunch.com/2020/03/02/apple-agrees-to-settlement-of-up-to-500-million-from-lawsuit-alleging-it-throttled-older-phones/.

326 C.,.G. (1909). Concentration and dependency ratios (in Italian). English translation in. *Rivista di Politica Economica (1997)* 87: 769–789.

327 www.reddit.com/r/MapPorn/comments/88mw3q/open_defecation_around_ the_world_2015_960_684/.

328 Nearly a billion people still defecate outdoors. Here's why. *Magazine* (25 July 2017). https://www.nationalgeographic.com/magazine/2017/08/toilet-defecate-outdoors-stunting-sanitation/.

329 Sadhu, B. et al. (2019). The more (antennas), the merrier: a survey of silicon-based mm-wave phased arrays using multi-IC scaling. *IEEE Microwave Magazine* 20 (12): 32–50.

330 Wang, T. and Kang, J.W. (2020). An integrated approach for assessing national e-commerce performance. Trade, Investment and Innovation Working Paper Series, No. 01/20, ESCAP Trade, Investment and Innovation Division, United Nations (UN). January 2020. Bangkok, Thailand. https://www.unescap.org/ sites/default/files/publications/Working%20Paper%20No.1_2020.pdf.

331 Andrew, G. (2015). *Open Defecation Around the World*. The World Bank.

332 http://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf.

333 Propaganda by erudite, credible and respectable scientists, are disturbing, devastating, desacralizing. Time, events and publications suggest that "giants" who are *good* are occasionally hypnotized by the slippery slope of metamorphosis from *good* to self-anointed "God" particles. Knowledge, which was once regarded as an oak tree, and supposed to usher in self-deprecation, modesty and humility, now, frequently suffers from bloating, sufficient to spill over the *black hole* of hubris. For example, a trio of brilliant male scientists, in the upper latitudes of North America, are acting as *Nostradamus*, stoked by greed and the quest for immortality, buoyed by corporate largesse, exclusively driven by the desire for wealth creation. A group of complicit organizations and ill-informed media are ever ready to quench the drab voices of reason and restraint, in favour of sensationalizing and amplifying this inane *Nostradamus Effect*. This harms society by generating derelict reports and reduces the credibility of august institutions and organizations which appear as pawns for corporate business development (https://knowledgegraphsocialgood.pubpub. org) often under a camouflage of so-called knowledge for social good. https:// knowledgegraphsocialgood.pubpub.org/programcommittee [This statement is the personal opinion of the author].

334 Boroush, M. (2020). *Research and Development: U.S. Trends and International Comparisons. Science and Engineering Indicators 2020. NSB-2020-3*. Alexandria, VA: National Science Board, National Science Foundation https://ncses.nsf.gov/ pubs/nsb20203/assets/nsb20203.pdf.

335 Cyranoski, D. (2018). China awaits controversial blacklist of 'poor quality' journals. *Nature* 562: 471–472. www.nature.com. https://doi.org/10.1038/ d41586-018-07025-5.

336 https://www.thehindu.com/profile/photographers/Prashant-Waydande/.

337 Indian research quality lags quantity. *Economic Times Blog* (24 December 2019). https://economictimes.indiatimes.com/blogs/et-editorials/indian-research-quality-lags-quantity/.

338 Jaishankar, D. (2013). The huge cost of India's discrimination against women. *The Atlantic* (18 March 2013). https://www.theatlantic.com/international/archive/2013/03/the-huge-cost-of-indias-discrimination-against-women/274115/.

339 Surbatovich, M., et al. (2017). Some recipes can do more than spoil your appetite: analyzing the security and privacy risks of IFTTT recipes. *Proceedings of the 26th International Conference on World Wide Web*, Intl WWW Conferences Steering Com, pp. 1501–10. doi:https://doi.org/10.1145/3038912.3052709. https://www.ftc.gov/system/files/documents/public_comments/2017/11/00026-141804.pdf.

340 https://coronavirus.jhu.edu/map.html.

341 Datta, S. (2020). CITCOM (unpublished) MIT Library. https://dspace.mit.edu/handle/1721.1/111021.

342 Cai, Y. et al. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science* https://doi.org/10.1126/science.abd4251.

343 Wiedmann, T. et al. (2020). Scientists' warning on affluence. *Nature Communications* 11 (1): 3107. https://doi.org/10.1038/s41467-020-16941-y, www.nature.com/articles/s41467-020-16941-y.pdf.

344 Giacobassi, C.A., Oliveira, D.A., Pola, C.C., Xiang, D., Tang, Y., Austin Datta, S.P., McLamore, E.S. and Gomes, C. (2020). Sense-Analyze-Respond-Actuate (SARA) Paradigm: Proof of concept systems spanning nanoscale and macroscale actuation for detection of Escherichia coli in water. (in press).

345 PDFs numbered 00, 01, 02, 03 and 04 in https://dspace.mit.edu/handle/1721.1/123984.

346 Yu, L.-R. et al. (2018). Aptamer-based proteomics identifies mortality-associated serum biomarkers in dialysis-dependent AKI patients. *Kidney International Reports* 3 (5): 1202–1213. https://doi.org/10.1016/j.ekir.2018.04.012.

347 Datta, S. (2020) CORONA: a social coronary. https://bit.ly/CORONARY.

348 Poire, N.P. (2011). The great transformation of 2021. ISBN-13 978-0557948901. http://www.lulu.com/shop/norman-poire/the-great-transformation-of-2021/paperback/product-14729057.html.

349 Victor Weisskopf. *MIT*. http://web.mit.edu/dikaiser/www/Kaiser.Weisskopf.pdf.

350 Gary, T. (2018). What if sugar is worse than just empty calories? An essay by gary taubes. *British Medical Journal*: j5808. www.bmj.com/content/bmj/360/bmj.j5808.full.pdf.

351 Marie, C. https://www.iupac.org/publications/ci/2011/3301/jan11.pdf.

352 Frankl, V.E. et al. (2020). *Yes to Life: In Spite of Everything.* Beacon Press. ISBN: ISBN-13 978-0807005552.

353 le Rond d'Alembert, J. (1886). Oeuvres de D'Alembert. ISBN-13 9781103953189.

354 Gordin, M.D. (2015). *Scientific Babel: How Science Was Done before and after Global English*. The University of Chicago Press.

355 https://citations.ouest-france.fr/citation-voltaire/medecins-administrent-medicaments-dont-savent-22351.html.

Market proponents of AI are individuals who blather about neural networks of which they know little, to solve problems using learning tools which they know less, for the society of human beings of whom they know nothing.

(**Adapted from** "Les médecins administrent des médicaments dont ils savent très peu, à des malades dont ils savent moins, pour guérir des maladies dont ils ne savent rien" – **Voltaire** [355])

(Doctors are men who prescribe medicines of which they know little, to cure diseases of which they know less, in human beings of whom they know nothing.)

← **Tweet**

Geoffrey Hinton
@geoffreyhinton

Suppose you have cancer and you have to choose between a black box AI surgeon that cannot explain how it works but has a 90% cure rate and a human surgeon with an 80% cure rate. Do you want the AI surgeon to be illegal?

3:37 PM · Feb 20, 2020 · Twitter Web App