Article

Approximate and Situated Causality in Deep Learning

Jordi Vallverdú [1,*]

[1] Philosophy Department - UAB; jordi.vallverdu@uab.cat
* Correspondence: jordi.vallverdu@uab.cat; Tel.: +345811618

Abstract: Causality is the most important topic in the history of Western Science, and since the beginning of the statistical paradigm, it meaning has been reconceptualized many times. Causality entered into the realm of multi-causal and statistical scenarios some centuries ago. Despite of widespread critics, today Deep Learning and Machine Learning advances are not weakening causality but are creating a new way of finding indirect factors correlations. This process makes possible us to talk about approximate causality, as well as about a situated causality.

Keywords: causality; deep learning; machine learning; counterfactual; explainable AI; blended cognition; mechanisms; system

1. Causalities in the 21st Century.

In classic Western Philosophies, causality was observed as an obvious observation of the divine regularities which were ruling Nature. From a dyadic truth perspective, some events were true while others were false, and those which were true followed strictly the Heaven's will. That ontological perspective allowed early Greek philosophers (inspired by Mesopotamian, Egyptian and Indian scientists) to define causal models of reality with causal relations deciphered from a single origin, the arche (ἀρχή). Anaximander, Anaximenes, Thales, Plato or Aristotle, among others, created different models about causality, all of them connected by the same idea: hazard or nothingness was not possible. Despite those ideas were defended by atomists (who thought on a Nature with both hazard and void), any trace of them was deleted from researches. On the other hand, Eastern philosophers departed from the opposite ontological point of view: at the beginning was the nothingness, and the only true reality is the continuous change of things [1]. For Buddhist (using a four-valued logic), Hindu, Confucian or Taoist philosophers, causality was a reconstruction of the human mind, which is also a non-permanent entity. Therefore, the notion of causality is ontologically determined by situated perspectives about information values [2], which allowed and fed different and fruitful heuristic approaches to reality [3], [4]. Such situated contexts of thinking shape the ways by which people perform epistemic and cognitive tasks[5]–[7].

These ontological variations can be justified and fully understood once we assume the Duhem-Quine Thesis, that is, that it is impossible to test a scientific hypothesis in isolation, because an empirical test of the hypothesis requires one or more background assumptions (also called auxiliary assumptions or auxiliary hypotheses). Therefore, the history of the idea of causality changes coherently across the geographies and historical periods, entering during late 19th Century into the realm of statistics and, later in 20th Century, in multi-causal perspectives [8]. The statistical nature of contemporary causality has been involved into debates between schools, mainly Bayesians and a broad range of frequentist variations. At the same time, the epistemic thresholds has been changing, as the recent debate about statistical significance has shown, desacralizing the *p*-value. The most recent and detailed academic debate on statistical significance was extremely detailed into the #1 Supplement of the Volume 73, 209 of the journal *The American Satistician*, released in March, 20th 2019. But during the last decades of 20th Century and the beginning of the 21st Century, computational tools have become the backbone of cutting scientific researches [9], [10]. After the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2019 doi:10.20944/preprints201907.0110.v1

great advances produced by machine learning techniques (henceforth, ML), several authors have asked themselves whether ML can contribute to the creation of causal knowledge. We will answer to this question into next section.

2. Deep Learning, Counterfactuals, and Causality

Is in this context, where the statistical analysis rules the study of causal relationships, that we find the attack to Machine Learning and Deep Learning as no suitable tools for the advance of causal and scientific knowledge. The most known and debated arguments come from the eminent statistician Judea Pearl [11], [12], and have been widely accepted. The main idea is that machine learning do not can create causal knowledge because the skill of managing counterfactuals, and following his exact words, [11] page 7: "Our general conclusion is that human-level AI cannot emerge solely from model-blind learning machines; it requires the symbiotic collaboration of data and models. Data science is only as much of a science as it facilitates the interpretation of data – a two-body problem, connecting data to reality. Data alone are hardly a science, regardless how big they get and how skillfully they are manipulated". What he is describing is the well-known problem of the black box model: we use machines that process very complex amounts of data and provide some extractions at the end. As has been called, it is a GIGO (Garbage In, Garbage Out) process [13], [14]. It could be affirmed that GIGO problems are computational versions of Chinese room mental experiment [15]: the machine can find patterns but without real and detailed causal meaning. This is what Pearl means: the blind use of data for establishing statistical correlations instead that of describing causal mechanisms. But is it true? In a nutshell: not at all. I'll explain the reasons.

2.1.Deep Learning is not a data driven but a context driven techology: made by humans for humans.

Most of epistemic criticisms against AI are always repeating the same idea: machines are still not able to operate like humans do [16], [17]. The idea is always the same: computers are operating with data using a blind semantic perspective that makes not possible that they understand the causal connections between data. It is the definition of a black box model [18], [19]. But here happens the first problem: deep learning (henceforth, DL) is not the result of automated machines creating by themselves search algorithms and after it, evaluating them as well as their results. DL is designed by humans, who select the data, evaluate the results and decide the next step into the chain of possible actions. At epistemic level, is under human evaluation the decision about how to interpret the validity of DL results, a complex, but still only, technique [20]. But even last trends in AGI design include causal thinking, as DeepMind team has recently detailed [21], and with explainable properties. The exponential growth of data and their correlations has been affecting several fields, especially epidemiology [22], [23]. Initially it can be expressed by the agents of some scientific community as a great challenge, in the same way that astronomical statistics modified the Aristotelian-Newtonian idea of physical cause, but with time, the research field accepts new ways of thinking. Consider also the revolution of computer proofs in mathematics and the debates that these techniques generated among experts [24], [25].

In that sense, DL is just providing a situated approximation to reality using correlational coherence parameters designed by the communities that use them. It is beyond the nature of any kind of machine learning to solve problems only related to human epistemic envisioning: let's take the long, unfinished, and even disgusting debates among the experts of different of statistical schools [8]. And this is true because data do not provide or determine epistemology, in the same sense that groups of data do not provide the syntax and semantics of the possible organization systems to which they can be assigned. Any connection between the complex dimensions of any event expresses a possible epistemic approach, which is a (necessary) working simplification. We cannot understand the world using the world itself, in the same way that the best map is not a 1:1 scale map, as Borges wrote (1946, *On Exactitude in Science):* "…In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer

 doi:10.20944/preprints201907.0110.v1

satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography"

Then, DL cannot follow a different information processing process, a specific one completely different from those run by humans. As any other epistemic activity, DL must include different levels of uncertainties if we want to use it [26]. Uncertainty is a reality for any cognitive system, and consequently, DL must be prepared to deal with it. Computer vision is a clear example of that set of problems [27]. Kendall and Gal have even coined new concepts to allow introduce uncertainty into DL: homocedastic, and heterocedastic uncertainties (both aleatoric) [28]. The way used to integrate such uncertainties can determine the epistemic model (which is a real cognitive algorithmic extension of ourselves). For example, Bayesian approach provides an efficient way to avoid overfitting, allow to work with multi-modal data, and make possible use them in real-time scenarios (as compared to Monte Carlo approaches) [29]; or even better, some authors are envisioning Bayesian Deep Learning [30]. Dimensionality is a related question that has also a computational solution, as Yosuhua Bengio has been exploring during last decades [31]–[33].

In any case, we cannot escape from the informational formal paradoxes, which were well-known at logical and mathematical level once Gödel explained them; they just emerge in this computational scenario, showing that artificial learnability can also be undecidable [34]. Machine learning is dealing with a rich set of statistical problems, those that even at biological level are calculated at approximate levels [35]–[37], a heuristics that are being implemented also into machines. This open range of possibilities, and the existence of mechanisms like informational selection procedures (induction, deduction, abduction), makes possible to use DL in a controlled but creative operational level [38].

2.2. Deep learning is already running counterfactual approaches.

The second big Pearl critics is against DL because of its incapacity of integrating counterfactual heuristics. First we must affirm that counterfactuals do not warrant with precision any epistemic model, just add some value (or not). From a classic epistemic point of view, counterfactuals do not provide a more robust scientific knowledge: a quick look at the last two thousands of both Western and Estern sciences can give support to this view [4]. Even going beyond, I affirm that counterfactual can block thinking once it is structurally retaled to a close domain or paradigm of well-established rules; otherwise is just fiction or an empty mental experiment. Counterfactals are a fundamental aspect of human reasoning [39]–[42], and their algorithmic integration is a good idea [43]. But at the same time, due to the underdetermination [44]–[46], counterfactual thinking can express completely wrong ideas about reality. DL can no have an objective ontology that allows it to design a perfect epistemological tool: because of the huge complexity of the involved data as well as for the necessary situatedness of any cognitive system. Uncertainty would not form part of such counterfactual operationability [47], once it should ascribed to any not-well known but domesticable aspect of reality; nonetheless, some new ideas do not fit with the whole set of known facts, the current paradigm , nor the set of new ones. This would position us into a sterile no man's land, or even block any sound epistemic movement. But humans are able to deal with it, going even beyond [48]. Opportunistic blending, and creative innovating are part of our set of most valuable cognitive skills [49] .

2.3. DL is not Magic Algorithmic Thinking (MAT).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2019 doi:10.20944/preprints201907.0110.v1

Our third and last analysis of DL characteristics is related to its explainability. Despite of the evidence that the causal debate is beyond any possible resolution provided by DL, because it belongs to ontological perspectives that require a different holistic analysis, it is clear that the results provided my DL must be nor only coherent but also explainable, otherwise we would be in front of a new algorithmic form of magic thinking. By the same reasons by which DL cannot be just mere a complex way of curve fitting, it cannot become a fuzzy domain beyond human understanding. Some attempts are being held to prevent us from this, most of them rules by DARPA: Big Mechanisms [50] or XAI (eXplainable Artificial Intelligence) [51], [52]. An image from DARPA website:

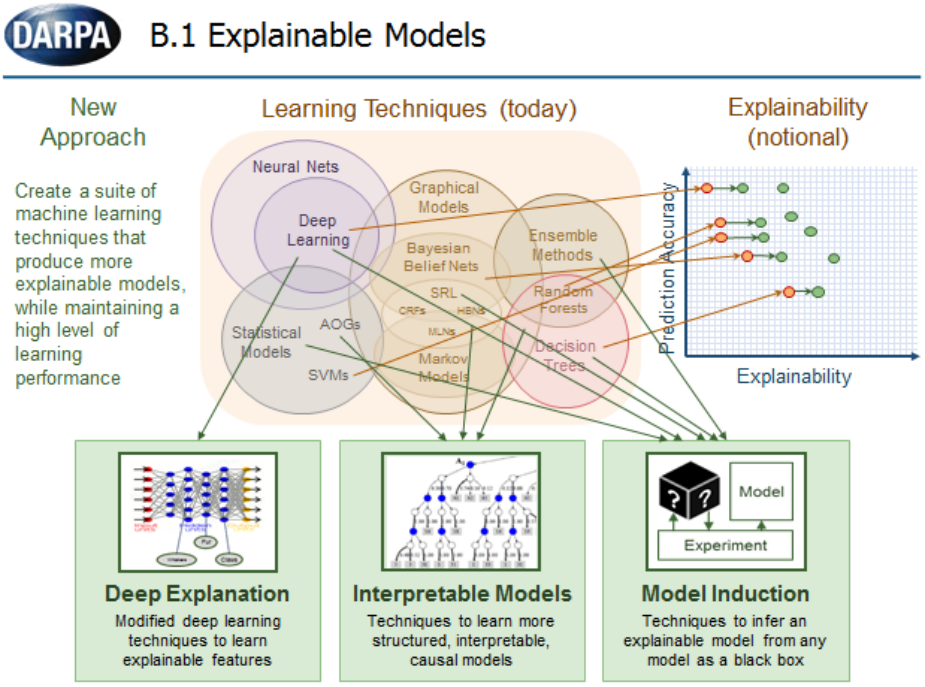

Figure 1. Explanability in DL

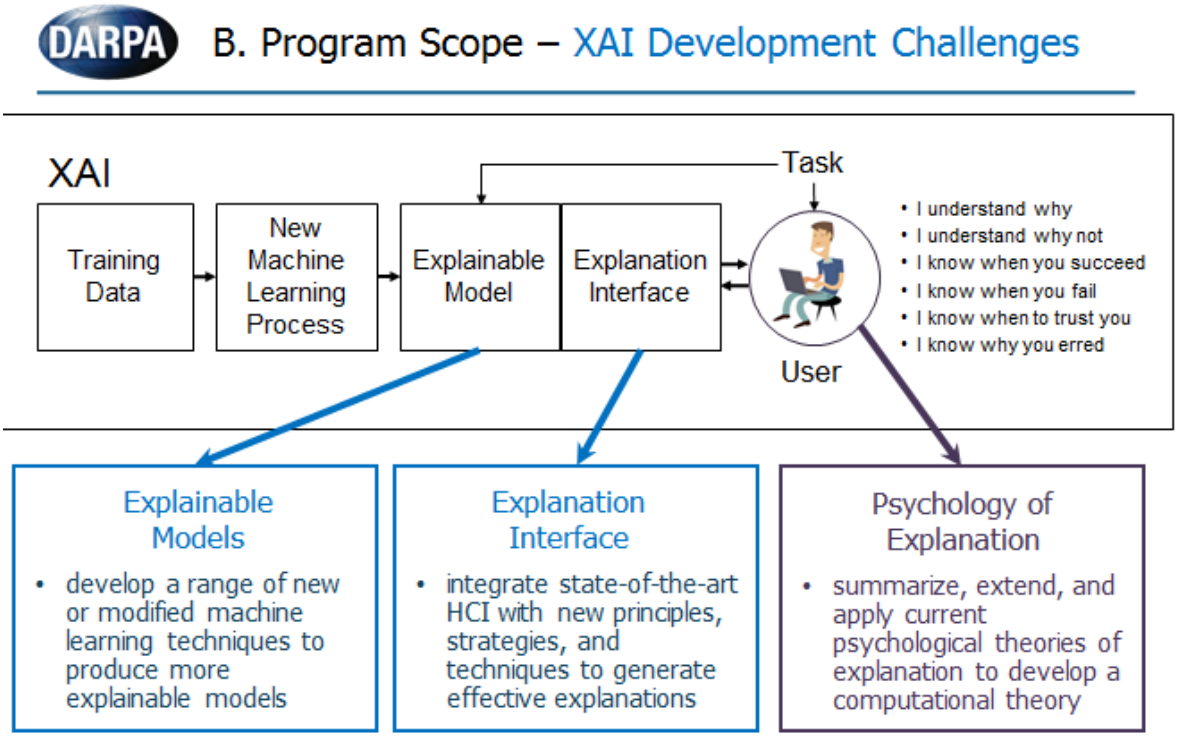

Figure 2. XAI from DARPA

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2019 doi:10.20944/preprints201907.0110.v1

Again, these approaches answer to a request: how to adapt new epistemic tools to our cognitive performative thresholds and characteristics. There is not a bigger revolution from a conceptual perspective than the ones that happened during Renaissance with the use of telescopes or microscopes. DL systems are not running by themselves interacting with the world, automatically selecting the informational events to be studied, or evaluating them in relation to a whole universal paradigm of semantic values. Humans neither do.

Singularity debates are useful and exploring possible conceptual frameworks and must be held [53]–[55], but at the same time cannot become fallacious fatalist arguments against current knowledge. Today, DL is a tool used by experts in order to map new connections between sets of data. Epistemology is not automated process, despite of minor and naïve attempts to achieve it [56], [57]. Knowledge is a complex set of explanations related to different systems that is integrated dynamically by networks of epistemic (human, still) agents who are working with AI tools. Machines could postulate their own models, true, but the mechanisms to verify or refine them would not be beyond any mechanism different from the used previously by humans: data do not express by itself some pure nature, but offers different system properties that need to be classified in order to obtain knowledge. And this obtaining is somehow a creation based on the epistemic and body situatedness of the system.

.

3. Extending bad and/or good human cognitive skills through DL.

It is beyond any doubt that DL is contributing to improve the knowledge in several areas some of them very difficult to interpret because of the nature of obtained data, like neuroscience [58]. These advances are expanding the frontiers of verifiable knowledge beyond classic human standards. But even in that sense, they are still explainable. Anyhow, are humans who fed up DL systems with scientific goals, provide data (from which to learn patterns) and define quantitative metrics (in order to know how close are you getting to success). At the same time, are we sure that is not our biased way to deal with cognitive processes that mechanism that allows us to be creative? For such reason, some attempts to reintroduce human biased reasoning into machine learning are being explored [59]. This re-biasing [60], even replying emotional like reasoning mechanisms [61], [62].

My suggestion is that after great achievements following classic formal algorithmic approaches, it now time for DL practitioners to expand horizons looking into the great power of cognitive biases.

For example, machine learning models with human cognitive biases are already capable of learning from small and biased datasets [63]. This process reminds the role of Student test in relation to frequentist ideas, always requesting large sets of data until the creation of the *t*-test, but now in the context of machine learning.

In [63] the authors developed a method to reduce the inferential gap between human beings and machines by utilizing cognitive biases. They implemented a human cognitive model into machine learning algorithms and compared their performance with the currently most popular methods, naïve Bayes, support vector machine, neural networks, logistic regression, and random forests. This even could make possible one-shot learning systems [64]. Approximate computing can

 doi:10.20944/preprints201907.0110.v1

boost the potentiality of DL, diminishing the computational power of the systems as well as adding new heuristic approaches to information analysis [35], [65]–[67].

Finally, a completely different type of problems, but also important, are how to reduce the biased datasets or heuristics we provide to our DL systems [68] as well as how to control the biases that make us not to interpret DL results properly [69]. Obviously, if there is any malicious value related to such bias, it must be also controlled [70].

4. Causality in DL: the epidemiological case study

Several attempts has been implemented in order to allow causal models in DL, like [71] and the Structural Causal Model (SCM) (as an abstraction over a specific aspect of the CNN. We also formulate a method to quantitatively rank the filters of a convolution layer according to their counterfactual importance), or Temporal Causal Discovery Framework (TCDF, a deep learning framework that learns a causal graph structure by discovering causal relationships in observational time series data) by [72]. But my attempt here will be twofold: (1) first, to consider about the value of "causal data" for epistemic decisions in epidemiology; and (2) second, to look how DL could fit or not with those causal claims into the epidemiological field.

4.1.Do causality affects at all epidemiological debates?

According to the field reference [73], MacMahon and Pugh [74] created the one of the most frequently used definitions of epidemiology: "Epidemiology is the study of the distribution and determinants of disease frequency in man". Note the absence of the term 'causality' and, instead, the use of the one of 'determinant'. This is the result of the classic prejudices of Hill in his paper of 1965: "*I have no wish, nor the skill, to embark upon philosophical discussion of the meaning of 'causation'. The 'cause' of illness may be immediate and direct; it may be remote and indirect underlying the observed association. But with the aims of occupational, and almost synonymous preventive, medicine in mind the decisive question is where the frequency of the undesirable event B will be influenced by a change in the environmental feature A. How such a change exerts that influence may call for a great deal of research, However, before deducing 'causation' and taking action we shall not invariably have to sit around awaiting the results of the research. The whole chain may have to be unraveled or a few links may suffice. It will depend upon circumstanc*es." After this philosophical epistemic positioning, Hill numbered his 9 general qualitative association factors, also commonly called "Hill's criteria" or even, which is frankly sardonic, "Hill's Criteria of Causation". For such epistemic reluctances epidemiologists abandoned the term "causation" and embraced other terms like "determinant" [75], "determining conditions" [76], or "active agents of change" [77]. For that reason, recent researches have claimed for a pluralistic approach to such complex analysis [78]. As a consequence we can see that even in a very narrow specialized field like epidemiology the meaning of cause is somehow fuzzy. Once medical evidences showed that causality was not always a mono-causality [22], [79] but, instead, the result of the sum of several causes/factors/determinants, the necessity of clarifying multi-causality emerged as a first-line epistemic problem. It was explained as a "web of causation" [80]. Some debates about the logics of causation and some popperian interpretations were held during two decades [81], [82]. Pearl himself provided a graphic way to adapt human cognitive visual skills to such new epidemiological multi-causal reasoning [83], as well do-calculus [84], [85], and directed acyclic graphs (DAGs) are becoming a fundamental tool [86], [87]. DAGs are commonly related to randomized controlled trials (RCT) for assessing causality. But RCT are not a Gold Standard beyond any critic [88], [89], because as [90] affirmed, RCT are often flawed, mostly useless, although clearly indispensable (it is not so uncommon that the same author claim against classic p-value suggesting a new 0,005, [91]). Krauss has even defended the impossibility of using RCT without biases [92], although some authors defend that DAGs can reduce RCT biases [93].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2019 doi:10.20944/preprints201907.0110.v1

But there a real case that can shows us a good example about the weight of causality in real scientific debates. We will see the debates about the relation between smoke and lung cancer. As soon as in 1950 were explained the causal connections between smoking and lung cancer [94]. But far from being accepted, these results were replied by tobacco industry using scientific experimental regression. Perhaps the most famous generator of silly counterarguments was R.A. Fisher, the most important frequentist researcher of 20th Century. In 1958 he published a paper in *Nature* journal [95] in which he affirmed that all connections between tobacco smoking and lung cancer were due to a false correlation. Even more: with the same data could be inferred that "smoking cigarettes was a cause of considerable prophylactic value in preventing the disease, for the practice of inhaling is rare among patients with cancer of the lung that with others" (p. 596). Two years later he was saying similar silly things in a high rated academic journal [96]. He even affirmed that Hill tried to plant fear into good citizens using propaganda, and entering misleadingly into the thread of overconfidence. The point is: did have Fisher real epistemic reasons for not to accepting the huge amount of existing causal evidences against tobacco smoking? No. And we are not affirming the consequent after collecting more data not available during Fisher life. He has strong causal evidences but he did not wanted to accept them. Still today, there are evidences that show how causal connections are field biased, again with tobacco or the new e-cigarettes [97]–[99].

As a section conclusion can be affirmed that causality has strong specialized meanings and can be studied under a broad range of conceptual tools. The real example of tobacco controversies offers such long temporal examples.

4.2.Can DL be of sume utility for the epidemiological debates on causality?

The second part of my argumentation will try to elucidate whether DL can be useful for the resolution of debates about causality in epidemiological controversies. The answer is easy and clear: yes. But it is directly related to a specific idea of causality as well as of a demonstration. For example, can be found a machine learning approach to enable evidence based oncology practice [100]. Thus, digital epidemiology is a robust update of previous epidemiological studies [101][102]. The new possibilities of finding new causal patterns using bigger sets of data is surely the best advantages of using DL for epidemiological purposes [103]. Besides, such data are the result of integrating multimodal sources, like visual combined with classic informational [104], but the future with mode and more data capture devices could integrate smell, taste, movements of agents,...deep convolutional neural networks can help us, for example, to estimate environmental exposures using images and other complementary data sources such as cell phone mobility and social media information. Combining fields such as computer vision and natural language processing, DL can provide the way to explore new interactions still opaque to us [105], [106].

Despite of the possible benefits, it is also true that the use of DL in epidemiological analysis has a dangerous potential of unethicality, as well as formal problems [107], [108]. But again, the evaluation of involved expert agents will evaluate such difficulties as things to be solved or huge obstacles for the advancement of the field.

5. Conclusion: causal evidence is not a result, but a *process*.

Author has made an overall reply to main critics to Deep Learning (and machine learning) as a reliable epistemic tool. Basic arguments of Judea Pearl have been criticized using real examples of DL, but also making a more general epistemic and philosophical analysis. The systemic nature of knowledge, also situated and even biased, has been pointed as the fundamental aspect of a new

algorithmic era for the advance of knowledge using DL tools. If formal systems have structural dead-ends like incompleteness, the bioinspired path to machine learning and DL becomes a reliable way [109], [110] to improve, one more time, our algorithmic approach to nature. Finally, thanks to the short case study of epidemiological debates on causality and their use of DL tools, we've seen a real implementation case of such epistemic mechanism. The advantages of DL for multi-causal analysis using multi-modal data have been explored as well as some possible critics.

Funding: This work has been funded by the Ministry of Science, Innovation and Universities within the State Subprogram of Knowledge Generation through the research project FFI2017-85711-P Epistemic innovation: the case of cognitive sciences. This work is also part of the consolidated research network "Grup d'Estudis Humanístics de Ciència i Tecnologia" (GEHUCT) ("Humanistic Studies of Science and Technology Research Group"), recognised and funded by the Generalitat de Catalunya, reference 2017 SGR 568.

Acknowledgments: I thank Mr. Isard Boix for his support all throughout this research. Best moments are those without words, and sometimes this lack of meaningfulness entails unique meanings.

Conflicts of Interest: The author declares no conflict of interest.

References

[1] J. W. Heisig, *Philosophers of Nothingness : An Essay on the Kyoto School*. University of Hawai'i Press, 2001.

[2] J. Vallverdú, "The Situated Nature of Informational Ontologies," in *Philosophy and Methodology of Information*, WORLD SCIENTIFIC, 2019, pp. 353–365.

[3] M. J. Schroeder and J. Vallverdú, "Situated phenomenology and biological systems: Eastern and Western synthesis.," *Prog. Biophys. Mol. Biol.*, vol. 119, no. 3, pp. 530–7, Dec. 2015.

[4] J. Vallverdú and M. J. Schroeder, "Lessons from culturally contrasted alternative methods of inquiry and styles of comprehension for the new foundations in the study of life," *Prog. Biophys. Mol. Biol.*, 2017.

[5] A. Carstensen, J. Zhang, G. D. Heyman, G. Fu, K. Lee, and C. M. Walker, "Context shapes early diversity in abstract thought," *Proc. Natl. Acad. Sci.*, pp. 1–6, Jun. 2019.

[6] A. Norenzayan and R. E. Nisbett, "Culture and causal cognition," *Curr. Dir. Psychol. Sci.*, vol. 9, no. 4, pp. 132–135, 2000.

[7] R. E. Nisbet, *The Geography of Thought: How Asians and Westerners Think Differently...and Why: Richard E. Nisbett: 9780743255356: Amazon.com: Books*. New York: Free Press (Simon & Schuster, Inc.), 2003.

[8] J. Vallverdú, *Bayesians versus frequentists : a philosophical debate on statistical reasoning*. Springer, 2016.

[9] D. Casacuberta and J. Vallverdú, "E-science and the data deluge.," *Philos. Psychol.*, vol. 27, no. 1, pp. 126–140, 2014.

[10] J. Vallverdú Segura, "Computational epistemology and e-science: A new way of thinking," *Minds Mach.*, vol. 19, no. 4, pp. 557–567, 2009.

[11] J. Pearl, "Theoretical Impediments to Machine Learning," *arXiv Prepr.*, 2018.

[12] J. Pearl and D. Mackenzie, *The book of why : the new science of cause and effect*. Basic Books, 2018.

[13] S. Hillary and S. Joshua, "Garbage in, garbage out (How purportedly great ML models can be screwed up by bad data)," in *Proceedings of Blackhat 2017*, 2017.

[14] I. Askira Gelman, "GIGO or not GIGO," *J. Data Inf. Qual.*, 2011.

[15] J. Moural, "The Chinese room argument," in *John searle*, 2003.

[16] H. L. Dreyfus, *What Computers Can't Do: A Critique of Artificial Reason*. 1972.

[17] H. L. Dreyfus, S. E. Drey-fus, and L. a. Zadeh, "Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer," *IEEE Expert*, vol. 2, no. 2, pp. 237–264, 1987.

[18] W. N. Price, "Big data and black-box medical algorithms," *Sci. Transl. Med.*, 2018.

[19] J. Vallverdú, "Patenting logic, mathematics or logarithms? The case of computer-assisted proofs," *Recent Patents Comput. Sci.*, vol. 4, no. 1, pp. 66–70, 2011.

[20] F. Gagliardi, "The necessity of machine learning and epistemology in the development of categorization theories: A case study in prototype-exemplar debate," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009.

[21] T. Everitt, R. Kumar, V. Krakovna, and S. Legg, "Modeling AGI Safety Frameworks with Causal Influence Diagrams," Jun. 2019.

[22] M. Susser and E. Susser, "Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology," *Am. J. Public Health*, vol. 86, no. 5, pp. 674–677, 1996.

[23] A. Morabia, "Hume, Mill, Hill, and the sui generis epidemiologic approach to causal inference.," *Am. J. Epidemiol.*, vol. 178, no. 10, pp. 1526–32, Nov. 2013.

[24] T. C. Hales, "Historical overview of the Kepler conjecture," *Discret. Comput. Geom.*, vol. 36, no. 1, pp. 5–20, 2006.

[25] N. Robertson, D. P. Sanders, P. D. Seymour, and R. Thomas, "The Four–Colour Theorem," *J. Comb. Theory, Ser. B*, vol. 70, pp. 2–44, 1997.

[26] Yarin Gal, Y. Gal, and Yarin Gal, "Uncertainty in Deep Learning," *Phd Thesis*, 2017.

[27] A. G. Kendall, "Geometry and Uncertainty in Deep Learning for Computer Vision," 2017.

[28] A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?," Mar. 2017.

[29] J. Piironen and A. Vehtari, "Comparison of Bayesian predictive methods for model selection," *Stat. Comput.*, 2017.

[30] N. G. Polson and V. Sokolov, "Deep learning: A Bayesian perspective," *Bayesian Anal.*, 2017.

[31] Y. Bengio and Y. Lecun, "Scaling Learning Algorithms towards AI To appear in ' Large-Scale Kernel Machines ',," *New York*, 2007.

[32] J. P. Cunningham and B. M. Yu, "Dimensionality reduction for large-scale neural recordings," *Nat. Neurosci.*, vol. 17, no. 11, pp. 1500–1509, 2014.

[33] S. Bengio and Y. Bengio, "Taking on the curse of dimensionality in joint distributions using neural networks," *IEEE Trans. Neural Networks*, 2000.

[34] S. Ben-David, P. Hrubeš, S. Moran, A. Shpilka, and A. Yehudayoff, "Learnability can be undecidable," *Nat. Mach. Intell.*, 2019.

[35] C. B. Anagnostopoulos, Y. Ntarladimas, and S. Hadjiefthymiades, "Situational computing: An innovative architecture with imprecise reasoning," *J. Syst. Softw.*, vol. 80, no. 12 SPEC. ISS., pp. 1993–2014, 2007.

[36] K. Friston, "Functional integration and inference in the brain," *Progress in Neurobiology*, vol. 68, no. 2. pp. 113–143, 2002.

[37] S. Schirra, "Approximate decision algorithms for approximate congruence," *Inf. Process. Lett.*, vol. 43, no. 1, pp. 29–34, 1992.

[38] L. Magnani, "AlphaGo, Locked Strategies, and Eco-Cognitive Openness," *Philosophies*, 2019.

[39] A. A. Baird and J. A. Fugelsang, "The emergence of consequential thought: Evidence from neuroscience," *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2004.

[40] T. Gomila, *Verbal Minds : Language and the Architecture of Cognition.* Elsevier Science, 2011.

[41] J. Y. Halpern, *Reasoning About Uncertainty*, vol. 2003. 2003.

[42] N. Van Hoeck, "Cognitive neuroscience of human counterfactual reasoning," *Front. Hum. Neurosci.*, 2015.

[43] J. Pearl, "The algorithmization of counterfactuals," *Ann. Math. Artif. Intell.*, 2011.

[44] D. Tulodziecki, "Underdetermination," in *The Routledge Handbook of Scientific Realism*, 2017.

[45] L. Laudan, "Demystifying Underdetermination," in *Philosophy of Science. The Central Issues*, 1990.

[46] D. Turner, "Local Underdetermination in Historical Science*," *Philos. Sci.*, 2005.

[47] D. Lewis, "Counterfactual Dependence and Time's Arrow," *Noûs*, 2006.

[48] M. Ramachandran, "A counterfactual analysis of causation," *Mind*, 2004.

[49] J. Vallverdú and V. C. Müller, *Blended cognition : the robotic challenge*. Springer, 2019.

[50] A. Rzhetsky, "The Big Mechanism program: Changing how science is done," in *CEUR Workshop Proceedings*, 2016.

[51] A. Wodecki *et al.*, "Explainable Artificial Intelligence (XAI) The Need for Explainable AI," 2017.

[52] T. Ha, S. Lee, and S. Kim, "Designing Explainability of an Artificial Intelligence System," 2018.

[53] R. Kurzweil, "The Singularity Is Near: When Humans Transcend Biology," *Book*, vol. 2011. p. 652, 2005.

[54] R. V Yampolskiy, "Leakproofing the Singularity - Artificial Intelligence Confinement Problem," *J. Conscious. Stud.*, vol. 19, no. 1–2, pp. 194–214, 2012.

[55] J. Vallverdú, "The Emotional Nature of Post-Cognitive Singularities," S. A. Victor Callaghan, James Miller, Roman Yampolskiy, Ed. Springer Berlin Heidelberg, 2017, pp. 193–208.

[56] R. D. King *et al.*, "The automation of science," *Science (80-.).*, 2009.

[57] A. Sparkes *et al.*, "Towards Robot Scientists for autonomous scientific discovery," *Automated Experimentation*. 2010.

[58] A. Iqbal, R. Khan, and T. Karayannis, "Developing a brain atlas through deep learning," *Nat. Mach. Intell.*, vol. 1, no. 6, pp. 277–287, Jun. 2019.

[59] D. D. Bourgin, J. C. Peterson, D. Reichman, T. L. Griffiths, and S. J. Russell, "Cognitive Model Priors for Predicting Human Decisions."

[60] J. Vallverdu, "Re-embodying cognition with the same 'biases'?," *Int. J. Eng. Futur. Technol.*, vol. 15, no. 1, pp. 23–31, 2018.

[61] A. Leukhin, M. Talanov, J. Vallverdú, and F. Gafarov, "Bio-plausible simulation of three monoamine systems to replicate emotional phenomena in a machine," *Biologically Inspired Cognitive Architectures*, 2018.

[62] J. Vallverdú, M. Talanov, S. Distefano, M. Mazzara, A. Tchitchigin, and I. Nurgaliev, "A cognitive architecture for the implementation of emotions in computing systems," *Biol. Inspired Cogn. Archit.*, vol. 15, pp. 34–40, 2016.

[63] H. Taniguchi, H. Sato, and T. Shirakawa, "A machine learning model with human cognitive biases capable of learning from small and biased datasets," *Sci. Rep.*, 2018.

[64] B. M. Lake, R. R. Salakhutdinov, and J. B. Tenenbaum, "One-shot learning by inverting a compositional causal process," in *Advances in Neural Information Processing Systems 27 (NIPS 2013)*, 2013.

[65] Q. Xu, T. Mytkowicz, and N. S. Kim, "Approximate Computing: A Survey," *IEEE Des. Test*, vol. 33, no. 1, pp. 8–22, 2016.

[66] C.-Y. Chen, J. Choi, K. Gopalakrishnan, V. Srinivasan, and S. Venkataramani, "Exploiting approximate computing for deep learning acceleration," in *2018 Design, Automation & Test in Europe Conference &*

Exhibition (DATE), 2018, pp. 821–826.

[67] J. Choi and S. Venkataramani, "Approximate Computing Techniques for Deep Neural Networks," in *Approximate Circuits*, Cham: Springer International Publishing, 2019, pp. 307–329.

[68] M. A. Gianfrancesco, S. Tamang, J. Yazdany, and G. Schmajuk, "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data," *JAMA Internal Medicine*. 2018.

[69] T. Kliegr, Š. Bahník, and J. Fürnkranz, "A review of possible effects of cognitive biases on interpretation of rule-based machine learning models," Apr. 2018.

[70] B. Shneiderman, "Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight," *Proc. Natl. Acad. Sci.*, 2016.

[71] T. Narendra, A. Sankaran, D. Vijaykeerthy, and S. Mani, "Explaining Deep Learning Models using Causal Inference," Nov. 2018.

[72] M. Nauta, D. Bucur, C. Seifert, M. Nauta, D. Bucur, and C. Seifert, "Causal Discovery with Attention-Based Convolutional Neural Networks," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 1, pp. 312–340, Jan. 2019.

[73] W. Ahrens and I. Pigeot, *Handbook of epidemiology: Second edition*. 2014.

[74] B. MacMahon and T. F. Pugh, *Epidemiology; principles and methods*. Little, Brown, 1970.

[75] R. Lipton and T. Ødegaard, "Causal thinking and causal language in epidemiology: it's in the details.," *Epidemiol. Perspect. Innov.*, vol. 2, p. 8, 2005.

[76] P. Vineis and D. Kriebel, "Causal models in epidemiology: Past inheritance and genetic future," *Environmental Health: A Global Access Science Source*. 2006.

[77] J. S. Kaufman and C. Poole, "Looking Back on 'Causal Thinking in the Health Sciences,'" *Annu. Rev. Public Health*, 2002.

[78] J. P. Vandenbroucke, A. Broadbent, and N. Pearce, "Causality and causal inference in epidemiology: The need for a pluralistic approach," *Int. J. Epidemiol.*, vol. 45, no. 6, pp. 1776–1786, 2016.

[79] M. Susser, "Causal thinking in the health sciences concepts and strategies of epidemiology," *Causal Think. Heal. Sci. concepts*, 1973.

[80] N. Krieger, "Epidemiology and the web of causation: Has anyone seen the spider?," *Soc. Sci. Med.*, vol. 39, no. 7, pp. 887–903, 1994.

[81] M. Susser, "What is a cause and how do we know one? a grammar for pragmatic epidemiology," *American Journal of Epidemiology*. 1991.

[82] C. Buck, "Popper's philosophy for epidemiologists," *Int. J. Epidemiol.*, 1975.

[83] S. Greenland, J. Pearl, and J. M. Robins, "Causal diagrams for epidemiologic research," *Epidemiology*, 1999.

[84] D. Gillies, "Judea Pearl Causality: Models, Reasoning, and Inference, Cambridge: Cambridge University Press, 2000," *Br. J. Philos. Sci.*, 2001.

[85] R. R. Tucci, "Introduction to Judea Pearl's Do-Calculus," Apr. 2013.

[86] S. Greenland, J. Pearl, and J. M. Robins, "Causal diagrams for epidemiologic research.," *Epidemiology*, vol. 10, no. 1, pp. 37–48, 1999.

[87] T. J. Vanderweele and J. M. Robins, "Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect," *Am. J. Epidemiol.*, 2007.

[88] C. Boudreau, E. C. Guinan, K. R. Lakhani, and C. Riedl, "The Novelty Paradox & Bias for Normal Science: Evidence from Randomized Medical Grant Proposal Evaluations," 2016.

[89] R. M. Daniel, B. L. De Stavola, and S. Vansteelandt, "Commentary: The formal approach to quantitative

causal inference in epidemiology: Misguided or misrepresented?," *International Journal of Epidemiology*, vol. 45, no. 6. pp. 1817–1829, 2016.

[90] J. P. A. Ioannidis, "Randomized controlled trials: Often flawed, mostly useless, clearly indispensable: A commentary on Deaton and Cartwright," *Soc. Sci. Med.*, vol. 210, pp. 53–56, Aug. 2018.

[91] J. P. A. Ioannidis, "The proposal to lower P value thresholds to .005," *JAMA - Journal of the American Medical Association*. 2018.

[92] A. Krauss, "Why all randomised controlled trials produce biased results," *Ann. Med.*, vol. 50, no. 4, pp. 312–322, May 2018.

[93] I. Shrier and R. W. Platt, "Reducing bias through directed acyclic graphs," *BMC Medical Research Methodology*. 2008.

[94] R. Doll and A. B. Hill, "Smoking and Carcinoma of the Lung," *Br. Med. J.*, vol. 2, no. 4682, pp. 739–748, Sep. 1950.

[95] R. A. FISHER, "Cancer and Smoking," *Nature*, vol. 182, no. 4635, pp. 596–596, Aug. 1958.

[96] R. A. FISHER, "Lung Cancer and Cigarettes?," *Nature*, vol. 182, no. 4628, pp. 108–108, Jul. 1958.

[97] B. Wynne, "When doubt becomes a weapon," *Nature*, vol. 466, no. 7305, pp. 441–442, 2010.

[98] C. Pisinger, N. Godtfredsen, and A. M. Bender, "A conflict of interest is strongly associated with tobacco industry–favourable results, indicating no harm of e-cigarettes," *Prev. Med. (Baltim).*, vol. 119, pp. 124–131, Feb. 2019.

[99] T. Grüning, A. B. Gilmore, and M. McKee, "Tobacco industry influence on science and scientists in Germany.," *Am. J. Public Health*, vol. 96, no. 1, pp. 20–32, Jan. 2006.

[100] N. Ramarajan, R. A. Badwe, P. Perry, G. Srivastava, N. S. Nair, and S. Gupta, "A machine learning approach to enable evidence based oncology practice: Ranking grade and applicability of RCTs to individual patients.," *J. Clin. Oncol.*, vol. 34, no. 15_suppl, pp. e18165–e18165, May 2016.

[101] M. Salathé, "Digital epidemiology: what is it, and where is it going?," *Life Sci. Soc. Policy*, vol. 14, no. 1, p. 1, Dec. 2018.

[102] E. Velasco, "Disease detection, epidemiology and outbreak response: the digital future of public health practice," *Life Sci. Soc. Policy*, vol. 14, no. 1, p. 7, Dec. 2018.

[103] C. Bellinger, M. S. Mohomed Jabbar, O. Zaïane, and A. Osornio-Vargas, "A systematic review of data mining and machine learning for air pollution epidemiology," *BMC Public Health*. 2017.

[104] S. Weichenthal, M. Hatzopoulou, and M. Brauer, "A picture tells a thousand…exposures: Opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology," *Environment International*. 2019.

[105] G. Eraslan, Ž. Avsec, J. Gagneur, and F. J. Theis, "Deep learning: new computational modelling techniques for genomics," *Nat. Rev. Genet.*, vol. 20, no. 7, pp. 389–403, Jul. 2019.

[106] M. J. Cardoso *et al.*, Eds., *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, vol. 10553. Cham: Springer International Publishing, 2017.

[107] C. Kreatsoulas and S. V Subramanian, "Machine learning in social epidemiology: Learning from experience.," *SSM - Popul. Heal.*, vol. 4, pp. 347–349, Apr. 2018.

[108] "In the age of machine learning randomized controlled trials are unethical." [Online]. Available: https://towardsdatascience.com/in-the-age-of-machine-learning-randomized-controlled-trials-are-unethical-74acc05724af. [Accessed: 03-Jul-2019].

[109] T. F. Drumond, T. Viéville, and F. Alexandre, "Bio-inspired Analysis of Deep Learning on Not-So-Big Data Using Data-Prototypes," *Front. Comput. Neurosci.*, vol. 12, p. 100, Jan. 2019.

 doi:10.20944/preprints201907.0110.v1

[110] K. Charalampous and A. Gasteratos, "Bio-inspired deep learning model for object recognition," in *2013 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2013, pp. 51–55.