IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96113

Five Advanced Computer Survey Foundations in AI

Ramesh Hanumanthappa Jaggal[1], Vidya S[2]

[1]PG Student, MCA, Bangalore Institute of Technology College, Bengaluru, India

[2]Asst. Professor, MCA, Bangalore Institute of Technology College, Bengaluru, India

Abstract – This article offers a summary of some of the most important research issues that the disciplines of artificial and computational intelligence are now grappling with (AI and CI). While AI comprises a number of techniques that enable robots to gather knowledge from data and perform autonomous tasks, CI acts as a means to that purpose by specializing in algorithms inspired by intricate natural occurrences (including the working of the brain). In this essay, we outline the major concerns pertaining to these sectors using five unique R's: (a)Reasonability, (b)Resilience, (c)Reproducibility, (d)Realism, and (e)Responsibility. Notably, the name AIR5, which refers to the five aforementioned R's collectively, refers to the five aforementioned R's, much as AIR is a basic component of biological life.

Keywords: Artificial intelligence, resilience, rationalizability, responsibility, realism, reproducibility

INTRODUCTION

AI was first developed with the goal of developing self-aware computers that could equal human intellect in particular fields. In a similar vein, the highly associated area in computational intelligence (CI) surfaced in an attempt to replicate the ideal learning and problem-solving ability observed in nature. Examples of CI span from approaches inspired by effective foraging behavioural patterns seen in ostensibly basic creatures like ants to cognitive computing case studies which imitate complicated human brain functions. Despite their (relatively) humble beginnings, the existing impacts of I quick access to humongous and growing quantities of data, (ii) rapid increase in computational power, and (iii) consistent advancements in data-driven machine learning (ML) algorithms [1]-[3] have contributed significantly to modern AI systems vastly surpassing humanly order to achieve project success across a wide range of applications. In this sense, some of the more noteworthy success tales that have garnered worldwide attention include IBM's Watson winning Jeopardy! [4]. Google Deep Mind's Alpha Go program defeated the greatest Go player in the world [5] their Alpha Zero chess algorithm defeated a world champion program [6] and Carnegie Mellon University's AI defeated four of the finest professional poker players in the world [7]. On January 2, 2019, the manuscript was received. On May 27, 2019, it was amended. On June 30, 2019, the document was approved. The School of Computer Science's Data Science and Artificial Intelligence Research Centre contributed to the funding of this study. (Yew-Soon Ong is the corresponding author.). A. Gupta is an associate professor at the Singapore Institute of Manufacturing Technology (SimTech), Agency for Science, Technology, and Research (ASTAR), Singapore. Digital Object Identifier Growing consensus holds that the area of artificial intelligence (AI) is set to have a big influence on society as a whole due to the industry's fast progress over the past ten years. Given that human intellect has contributed significantly to much of what humanity has accomplished, It is clear that the potential for augmenting cognitive capacities with AI (a combination often known as augmented intelligence) [8] has immense promise for better decision-making in high-impact industries including healthcare, environmental science, economics, and governance and so on. However, in order for the idea of AI to be more universally trusted, accepted, and smoothly woven into the fabric of society, there are still huge scientific problems that need to be resolved. We outline some of these issues in this article using the five distinct R's: (a)R1- rationalizability, (b) R2- resilience, (c) R3- reproducibility, (d) R4- realism, and (e) R5- responsibility, which we believe represent five key aspects of AI research that will support the discipline's sustained growth throughout the twenty-first century and beyond. To sum up, just as air is the fundamental component of biological life, the phrase AIR5, which stands for the five previously stated R's, is used to refer to some of the fundamental components of artificial life. The remaining part of the article is set up to give a concise overview of each of the five R's, highlighting their fundamental significance for the development of AI.

1. AI SYSTEMS REASONABILITY

The usage of deep neural networks (DNNs) is the foundation of many recent machine learning (ML)-based AI developments [2]- [3]. The human brain which is made up of the vast neural networks serves as a rough model for the creation of DNNs. It is not unexpected that this model has drawn a lot of attention because it is a key source of knowledge

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96113

about the natural world. “DNNs” are sometimes criticized, nevertheless, for being extremely opaque. The difficulty in interpreting these models-loosely defined as the science of understanding what a model might have done [9] and deriving explanations for why particular inputs result in the identified outputs/predictions/decisions stems from their layered non-linear structure, which often allows them to achieve remarkable prediction accuracy. DNN models have mostly been employed as "black boxes" because of their lack of transparency and causation [10]-[11]. In light of the aforementioned, it is asserted that in order for people to develop a better level of acceptability for contemporary AI systems, their operations and consequent outputs must be made more rationalizable that is, they must be rationalizable (interpreted and explained). Before an AI system is used in the real world, it is crucial to fully comprehend and confirm what it has learned in safety-critical applications, the requirement for rationalizability cannot be compromised. Examples include making medical diagnoses, using driverless vehicles, and other circumstances in which people's lives are in danger right away. One well-known example illustrating the risks of transparency in neural networks (NNs) is the predictions of patients death in the context of community-acquired pneumonia [12]. A unique (less effective but more understandable) rule-based approach was developed to identify the ensure that all aspects from one of pneumonia datasets, despite the fact that NNs seemed to become the most adequate model for this job (when tested on the given test data). Having asthma (x) reduces your risk of dying [13]. The data which used train the system had a distinct (albeit obviously erroneous) trend that was shown by the inferred rule; this pattern may have also hampered the NN. Unfortunately, in these delicate situations, it is usually impossible to verify and assess the correctness of trained NNs. Thus, it is possible to significantly reduce potential risks brought on by the unintended learning of erroneous patterns from raw data by developing rationalizable models based on accepted theories [14]-[15]. The heart of rationalizability, according to this argument, is interpretable and explicable AI, but this is not the whole story. Even when a model is able to draw reasons for its own predictions from hitherto unnoticed input data, the level of confidence in those predictions will not be accurately recorded and displayed. Such uncertainties are logically expected, particularly when an input point is located beyond the coverage of the dataset which was used to train the model. speculative possibility While DNNs are (justly) considered as cutting-edge ML approaches in this regard, it is important to note that they do not (yet) properly represent uncertainty [16]. With some basic work in constructing a rigorous Bayesian interpretation of popular deep learning algorithms previously reported in [17]-[18] this paves the way for future research in probabilistic AI and machine learning.

2. SUITABILITY OF AI SYSTEMS

Even the most sophisticated models, like DNNs, have an exceptional propensity to be readily deceived, according to new study, despite the tremendous advances made in AI [19]. Well-known instances of how a trained DNN classifier's output may be significantly changed by making a tiny additive change to an input images have appeared in the computer vision field [20]. The additional disruption, often referred to as an adversarial assault, is typically so slight as to be undetectable to the human eye yet results in incorrect classification by the DNN. In extreme circumstances, it has been shown that assaulting a single picture pixel is sufficient to trick several types of DNNs [21]. A very useful illustration of the general phenomena is given. An image recognition AI was tricked into categorizing a "Stop" sign as a "Speed Limit 45" sign by placing a few black and white stickers on it in [22] which provides a particularly illuminating example of the general phenomena. Notably, voice recognition apps have reported comparable results [23]. The aforementioned ("Stop" sign) case-study is particularly troubling for businesses like self-driving automobiles, even if the repercussions of such flagrant misrepresentation are evident. Therefore, there have been focused attempts in recent years to strengthen DNNs' resilience, or their capacity to maintain high anticipated accuracy even when under assault from enemies (input perturbations). To that end, part of the suggested data poisoning has emerged as a different type of attack that can explicitly harm the training process, in additional to adversarial attacks that are meant to occur after that a highly trained model is placed into operation. In this case, the attacker's goal is to slightly change the training dataset by adding new data points [24] or changing existing ones [25] to drive the learner to choose a flawed model. As the training data the essential component of all machine learning systems is obtained from the outside world, where it is susceptible to purposeful or accidental manipulation, maintaining performance resilience against such assaults is unquestionably of utmost significance [26]. For contemporary ML paradigms like federated learning, the challenges are even more challenging.

3. AI SYSTEMS CAPABILITY OF REPRODUCTION

A topic that is commonly mentioned while training DNNs and ML models in general is the replication dilemma [27]. In essence, it has been challenging for others to duplicate some of the major findings described in the literature. Reproducibility is a prerequisite for every claim to be believable and illuminating, as stated in [28]. Therefore, assuring reproducible AI system performance through the creation and adherence to precise software standards, as well as through thorough system testing and validation on common datasets and benchmarks, is essential for preserving their credibility. Then, we'll briefly go through two other approaches that are supplementary to the final goal. A fundamental barrier to effectively recreating published findings is the vast number of hyperparameters for example, neural architecture decisions,

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96113

learning algorithm settings, etc. that must be carefully specified before training a model on any particular dataset [29]. The setup of these configurations can have a substantial impact on the effectiveness of the learning process, even though they are often viewed as a secondary consideration among the main components of a model or learning algorithm. As a result, the trained model may not function well if the best hyperparameter selection techniques are not used. Or to put it another way, as may have been published in a scientific article, the model falls short of its real potential. Given the aforementioned considerations, automating the entire process by formulating it all as a global optimization problem is an effective substitute for manually setting the hyperparameters.

Numerous strategies have been proposed to do this, including stochastic evolutionary algorithms [30]-[31] and Bayesian optimization techniques [32] which allow for the automatic selection of almost optimal hyperparameters (thus preventing human inaccuracies). The entire strategy is covered under the term AutoML (automatic machine learning) [33] which has lately gained popularity among ML professionals. Continuous work is being done to develop algorithms that can automatically exchange and reuse gained knowledge across datasets, problems, and domains [34]. The goal is to make AI more generalizable so that its performance may be replicated in other activities that are comparable to its own by sharing common knowledge building blocks, as opposed to being limited to a specific (narrow) activity. Transfer and multitask learning [35]-[37] and its extensions to the field of global optimization are promising research projects in this area (through transfer and multitask optimization) [35]. Memetic computation is a related research area that is currently being developed in the area of nature-inspired CI.

A fundamental unit of information that exists in the brain and is duplicated from one brain to another through imitation is how memories were initially characterized sociologically in [36]. Since then, memories have evolved to stand in for a variety of computationally encoded sorts of information that may be transferred from one job to another in order to enhance performance. A more immediate step toward boosting AI reproducibility is to encourage the sharing of datasets and executable code from fully documented scientific publications, in addition to the long-term development of algorithms that can automate the selection of hyperparameters. Despite the fact that open collaborations and the creation of open-source software are growing in popularity in the field of artificial intelligence, a recent study found that leading AI conferences continue to use documentation practices that make stated findings mostly unreplaceable [35]. To put it another way, in order to make the assessment of AI technologies easier, globally recognized software standards pertaining to code documentation, data formats, testing environment setup, etc. remain essential.

4. THE REALITY OF AI SYSTEMS

The performance effectiveness and accuracy of AI systems have been the three R's main areas of study up to this point. The development of emotional intelligence in machines is the main topic of this section since it is thought to be equally crucial for the successful eventual integration of AI into society. The daily usage of smart speakers (such Google Home gadgets and Amazon's Alexa), the advancement of education through virtual instructors [35] and even offering psychological assistance to Syrian refugees via chat-bots [36] are all areas where AI has showed promise. Such human-aware AI systems [37] must not only be reliable but also display traits that humans have, such relatability, goodness, and honesty. The ongoing pursuit of high accuracy and automation must be balanced with the creation of machine behaviours that result in more satisfying human-computer interaction if we are to attain realism in intelligent systems. Numerous research lines have come into focus in this area.

One-way affective computing attempts to advance human comprehension is by researching how to make AI more adept at recognizing, comprehending, and expressing genuine emotions and feelings [37]. The creation of systems capable of recognizing and analysing multimodal data streams is one of the key challenges facing the field. The driving argument is based on the fact that individuals express themselves differently and to diverse degrees through a variety of communication techniques (such as speech, body language, facial emotions, etc.).

As a consequence, as compared to the finest immoral analysis approaches, which analyse individual emotional signals in isolation, integrating visual and auditory information cues typically results in a more comprehensive comprehension of a person's mood [38]-[39]. Collective intelligence is a meta-concept that supports openly leveraging the collective knowledge of the people, as opposed to affective computing, which focuses on a specific group of concerns connected to human-centered learning. The development of techniques that can detect and manage multimodal data streams is one of the major issues facing the area. The driving argument is based on the fact that individuals express themselves differently and use various forms of communication to differing degrees (such as speech, body language, facial expressions, and so on). Therefore, as compared to the finest immoral analysis tools, which analyse individual emotional signals independently, integrating visual and auditory information cues generally yields a more comprehensive knowledge of a person's mood [38]-[39]. Collective Intelligence is a meta-concept which encourages explicitly tapping into the collective wisdom of the people, in contrast to emotional computing, which focuses on a specific group of difficulties associated with human-centered learning.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96113

5. RESULT OF AI SYSTEMS RESPONSIBILITY

According to the IEEE guidelines for designing systems that are ethically compliant, " We need to establish societal and regulatory guidelines to make sure that automated and intelligent systems remain centred on people and uphold humanity's values as their use and influence spread.". As a result, we include the objective of incorporating ethics into AI under the final R. Here, we presume that the term "ethics" refers to a normative practical philosophical study that examines how one should behave toward others. While the goal of realism emphasizes tight cooperation between humans and machines, responsibility is a general idea that needs to be incorporated at all levels of an AI system.

The capacity to effectively understand complicated patterns from massive volumes of data, as was previously said, has been one amazing result of current AI technology, frequently leading to performance levels that are above the capabilities of humans. Naturally, given that the concept of the future when robots rule the planet is now quite popular, their incredible power has also given rise to significant fear. In light of this, the current Taking cues from the imagined governing principles of Isaac Asimov's robotic-based world, the AI research community has begun to recognise that computer ethics play a key role throughout the design of intelligent autonomous systems are designed to be a part of a bigger ecosystem consisting of human stakeholders. To properly define what ethical machine behaviour is in order to develop accurate legal frameworks for it, however, is a difficult problem. Although current frameworks have generally put the responsibility for ethics codification on It has been suggested that effective issues with intelligent machines may be outside the purview of system designers, so take note, AI developers. In fact, a careful assessment is required of a number of nuanced problems, including those relating to privacy, public policy, and national security. The following list serves as an instance of the kind of problems that are certain to come up but that are impossible or extremely difficult to find an objective solution to.

• From phone-lines, camera surveillance or emails how much information should AI systems be able to access in the sake of performance optimization?

• How can self-driving car insurance plans be drafted to strike a balance between the chance of small human injury and the very likely occurrence of significant material damage to private or public property?

• How should autonomous weapons in national security and defence applications adhere to humanitarian law while keeping their original design goals?

Reaching an agreement on such matters will be challenging, especially as ethical truth is sometimes arbitrary and differs between civilizations and people. Because of this, there is an undeniable need for urgent worldwide research investment in the vision of infusing ethics into AI.

CONCLUSION

It is critical to keep in mind that the many concepts from R1 (rationalizability) through R4 (realism), which allow autonomous systems to reliably function and defend their actions in the context of human ethics and emotions, serve as stepping stones toward greater accountability in AI. The General Data Protection Regulation of the European Union, which implies a "right to explanation" really stipulates that the capability to do so be available.

REFERENCES

[1] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Netw., vol. 61, pp. 85–117, 2015.
[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[3] K. O. Stanley, J. Clune, J. Lehman, and R. Miikkulainen, "Designing neural networks through neuroevolution," Nature Mach. Intell., vol. 1, no. 1, pp. 24–35, 2019.
[4] D. A. Ferrucci, "Introduction to 'This is Watson'," IBM J. Res. Develop., vol. 56, no. 3.4, pp. 1.1–1.15, 2012.
[5] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, no. 7587, pp. 484–489, 2016.
[6] D. Silver et al., "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," Science, vol. 362, no. 6419, pp. 1140–1144, 2018.
[7] T. Sandholm, "Super-human AI for strategic reasoning: beating top pros in heads-up no-limit texas hold'em," in Proc. 26th Int. Joint Conf. Artif. Intell., Aug. 2017, pp. 24–25.
[8] E. Szathmáry, M. J. Rees, T. J. Sejnowski, T. Nørretranders, and W. B. Arthur, "Artificial or augmented intelligence? The ethical and societal implications," in Grand Challenges for Science in the 21st Century, vol. 7. Singapore: World Scientific, 2018, pp. 51–68.
[9] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: An overview of interpretability of machine learning," in Proc. IEEE 5th Int. Conf. Data Sci. Adv. Anal., Oct. 2018, pp. 80–89.
[10] W. Samek, T. Wiegand, and K. R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," arXiv:1708.08296, 2017.

DOI: 10.17148/IARJSET.2022.96113

[11] Z. Zeng, C. Miao, C. Leung, and C. J. Jih, "Building more explainable artificial intelligence with argumentation," in Proc. 23rd AAAI/SIGAI Doctoral Consortium, 2018, pp. 8044–8045.
[12] G. F. Cooper et al., "An evaluation of machine-learning methods for predicting pneumonia mortality," Artif. Intell. Med., vol. 9, no. 2, pp. 107–138, 1997.
[13] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," in Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2015, pp. 1721–1730.
[14] A. Karpatne et al., "Theory-guided data science: A new paradigm for scientific discovery from data," IEEE Trans. Knowl. Data Eng., vol. 29, no. 10, pp. 2318–2331, Oct. 2017.
[15] X. Jia et al., "Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles," in Proc. SIAM Int. Conf. Data Mining, May 2019, pp. 558–566.
[16] Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," Nature, vol. 521, no. 7553, pp. 452–459, 2015.
[17] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in Proc. 33rd Int. Conf. Mach. Learn., Jun. 2016, pp. 1050–1059.
[18] Y. Gal and Z. Ghahramani, "A theoretically grounded application of dropout in recurrent neural networks," in Proc. 30th Int. Neural Inf. Process. Syst., 2016, pp. 1019–1027.
[19] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2015, pp. 427–436.
[20] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," IEEE Access, vol. 6, pp. 14410–14430, 2018.
[21] J. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," IEEE Trans. Evol. Comput., to be published.
[22] K. Eykholt et al., "Robust physical-world attacks on deep learning visual classification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 1625–1634.
[23] N. Carlini and D. Wagner, "Audio adversarial examples: Targeted attacks on speech-to-text," in Proc. IEEE Secur. Privacy Workshops, May 2018, pp. 1–7.
[24] B. Biggio, B. Nelson, and P. Laskov, "Poisoning attacks against support vector machines," in Proc. 29th Int. Conf. Int. Conf. Mach. Learn., Jun. 2012, pp. 1467–1474.
[25] M. Zhao, B. An, W. Gao, and T. Zhang, "Efficient label contamination attacks against black-box learning models," in Proc. 26th Int. Joint Conf. Artif. Intell., Aug. 2017, pp. 3945–3951.
[26] J. Steinhardt, P. W. W. Koh, and P. S. Liang, "Certified defenses for data poisoning attacks," in Proc. 31st Int. Conf. Adv. Neural Inf. Process. Syst., 2017, pp. 3517–3529.
[27] M. Hutson, "Artificial intelligence faces reproducibility crisis," Science, vol. 359, no. 6377, pp. 725–726, 2018.
[28] K. Bollen, J. T. Cacioppo, R. M. Kaplan, J. A. Krosnick, and J. L. Olds, "Social, behavioral, and economic sciences perspectives on robust and reliable science: Report of the subcommittee on replicability in science, advisory committee to the national science foundation directorate for social, behavioral, and economic sciences," 2015.
[29] A. Klein, E. Christiansen, K.Murphy, and F. Hutter, "Towards reproducible neural architecture and hyperparameter search," 2018. [Online]. Available: https://openreview.net/pdf?id=rJeMCSnml7
[30] I. Loshchilov and F. Hutter, "CMA-ES for hyperparameter optimization of deep neural networks," in Proc. ICLR Workshop, 2016.
[31] R. Miikkulainen et al., "Evolving deep neural networks," In Artificial Intelligence in the Age of Neural Networks and Brain Computing. New York, NY, USA: Academic, 2019, pp. 293–312.
[32] B. Shahriari, K. Swersky, Z.Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of Bayesian optimization,"Proc. IEEE, vol. 104, no. 1, pp. 148–175, Jan. 2016.
[33] M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, and F. Hutter, "Efficient and robust automated machine learning," in Proc. 28th Int. Conf. Adv. Neural Inf. Process. Syst., 2015, pp. 2962–2970.
[34] J. N. van Rijn and F. Hutter, "Hyperparameter importance across datasets," in Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2018, pp. 2367–2376.
[35] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," J. Big Data, vol. 3, no. 9, pp. 1–40, 2016.
[36] B. Da, Y. S. Ong, A. Gupta, L. Feng, and H. Liu, "Fast transfer Gaussian process regression with large-scale sources," Knowl.-Based Syst., vol. 165, pp. 208–218, 2019.
[37] R. Caruana, "Multitask learning," Mach. Learn., vol. 28, no. 1, pp. 41–75, 1997.
[38] A. Gupta, Y. S. Ong, and L. Feng, "Insights on transfer optimization: Because experience is the best teacher," IEEE Trans. Emerg. Topics Comput. Intell., vol. 2, no. 1, pp. 51–64, Feb. 2018.
[39] D. Yogatama and G. Mann, "Efficient transfer learning method for automatic hyperparameter tuning,"Artif. Intell. Statist., vol. 33, pp. 1077–1085, Apr. 2014.