Global Faculty Initiative

The Faculty Initiative

seeks to promote the integration

of Christian faith and academic disciplines

by bringing theologians into conversation with scholars

across the spectrum of faculties

in research universities

worldwide.

www.globalfacultyinitiative.net

Disciplinary Brief

ARTIFICIAL INTELLIGENCE FOR SOCIAL GOOD

Jacqueline C.K. Lam

Associate Professor, Department of Electrical and Electronic Engineering, University of Hong Kong

Victor O.K. Li

Chair Professor of Information Engineering, Department of Electrical and Electronic Engineering, University of Hong Kong

Nigel Biggar (2022) argues that there is a created order in the world, but then humans, due to their sinful nature, have degraded and distorted that order.

AI and the Observation of Order

In our work on Artificial Intelligence (AI), we also believe that there exists a created natural order, which we can observe through analysis of different datasets, this may include image datasets, text datasets, audio datasets, etc. An AI model is trained using a collected dataset as input. For example, we may use a dataset containing 10,000 photos of animals as input, with each photo labelled with the identity of the animal, to train an AI model so it will be able to determine the animal shown in a photo. One popular AI model is the feed-forward neural network (NN). It consists of multiple layers, each consisting of nodes or neurons, with the input images (the input layer) feeding the lowest/first layer of the NN, which is connected to the second layer, then the third layer, etc., until the top layer, which is connected to the output layer (the label of the input image). The neurons of each layer are connected to the neurons of the next higher layer with edges, each assigned a weight/parameter. An input image, coded as the intensity of the pixels of the photo, is multiplied by the weights as it goes up the layers, eventually reaching the output layer. Obviously, we want an input image of a cat to eventually produce the output label of a cat, rather than another animal. The goal of training is to adjust the parameters so that they fit the majority of input images. Once the model is trained, i.e., the parameters have been optimized, one can give the model a new image, outside of the original images used in the training, and ask the model to determine which animal it is. We note that there is hierarchical structural order in this NN. The first layer determines the edges in the image. The second layer determines corners, or intersection of edges, and as we go up hierarchies, the layers identify increasingly complex structures in the image, such as the face of the

animal, the eyes, etc.

While we believe that there is an emergent, created order in the natural world, our observations via data collection and the hierarchical AI techniques may not be able to give us a true or comprehensive representation of this emergent order [1] [2]. Whatever AI models, hierarchical or non-hierarchical alike, the accuracy is dependent on the quality of the dataset used for model training. If the original 10,000 photos are those of dogs and cats, and one presents the AI model a photo of an elephant, it will not be able to determine that it is an elephant. Even if we increase the number of photos to one million in the dataset, and include many more different animals, we may never be able to include all the animals in the world, as there may be certain rare animals for which a photo is unavailable for training or, due to the modern view of the genetic 'tree of life' (which swept away Platonic and Aristotelian ideas of 'form'), there may be animals which very closely resemble each other. Furthermore, the photos may be blurry, and the model trained with such photos will not be accurate.

AI in Pursuit of God's Created Order

How can AI facilitate the type of "order" and "goods" that God destines human beings to uphold?

As implied by Biggar, "the order which … God impresses on the created world is not merely physical, but also value-laden", "the co-originality of matter and value" implies that humans care about a range of goods not just material ones, "immaterial goods such as moral integrity, the virtue of charity, and relations of justice can be very powerful motives". We fully share Biggar's view concerning the value-laden nature of God's created order. It is therefore our desire to encourage academics to appreciate such values or immaterial goods, as we embed such values into our discipline of AI for Social Good.

While engineers and natural scientists typically work on programmes that strive to help humanity flourish by means of material prosperity, such as, developing sustainable energy systems, creating smart cities, etc., social scientists are keen on investigating non-material goods, such as mental wellness, family relationships and social welfare. Many of these objectives are constrained by a humanistic perspective, failing to appreciate the goods that originate from the Creator. Biggar (2022) is inclined to associate "goods" with things that are "desirable", which also implies the objective of "human flourishing". So, our question becomes, what type of "goods" does the Creator of "order" and "goods" want to achieve via AI for Social Good?

Questions on AI for Social Good: Case Studies on Alzheimer's Disease and Air Quality Estimation

Along this line, below are some questions that a Christian-infused AI for Social Good should address:

1. What types of AI model design and operation align with "order" and "goods" in biblical terms?

While AI is designed by human beings, most of the time an NN is a black box which runs automatically without human intervention. Does this mean that AI can automatically operate without human input/guidance? If so, can an NN better reflect the "order" of the created system? Can the results generated by a self-governing NN be free from any human bias or subjectivity, and generate significance and meaning for humanity/society? Thinking about this differently: if there are divinely created orders that operate at an emergent level within the world, can self-governing AI make these visible? Put in Biggar's terms, can AI create value-driven prosperity, fostering human flourishing, drawing people to better understand the good and perfect order of our Creator, with a stronger reflection on truths underlying natural orders?

Take the specific example of using AI to screen existing drugs for their suitability in early treatment of Alzheimer's Disease (AD) to prevent its symptomatic development. This ideally requires a large number of datasets of known drugs and drug-drug interactions, complete human reactome data incorporating both protein-protein interactions, drug-target associations, and the relative involvement of each pathological pathway in AD. Recent development of tools such as AlphaFold [3], which opens the prospect of predicting protein structure from genetic data, and innovative drug-target prediction methods [4], provide potential directions to remove human bias in data selection.

Take another example of AD biomarker identification. Whilst big datasets may help mitigate human bias, due to the enormous number of interactions and the volume of data that would result from NN incorporating such big data, considerable human biases remain when only a small sample of selected data are inputted to the NN. Additional biases can be introduced through human perception of pathogenesis of AD: the molecular mechanism of Alzheimer's Disease remains ambiguous, and debates linger over the relative contribution of β-amyloid plaques, Neuro-Fibrillary Tangles, the immune system, oxidation pathways, and somatic mutagenesis, etc. Exploitation of NN for screening biomarkers of AD thus requires big datasets covering a wide range of subjects and potential biomarkers, such as genes and proteins, linguistic markers, behavioural markers of AD, coupled with the latest scientific/expert knowledge concerning AD pathogenesis, as demonstrated in our expert-guided AI-driven framework, exploiting different and big AD-associated datasets to identify early biomarkers governing early AD onset. This, along with datasets consisting of many subjects used to screen potential genes/biomarkers associated with AD, provide the probability of a particular gene/biomarker's contribution to AD.

To better reflect the created natural order, NN should be built with a capacity to remove possible human biases, while at the same time ready to incorporate human knowledge which reflects the more objective scientific/natural order. The wisdom of determining when human inputs are noises or valuable insights is an art rather than a science.

2. What are considered as biblically "desirable" AI outcomes?

Along the line of Biggar's elaboration on "goods", what outcomes generated by AI are considered as biblically good/desirable? How can AI generate more goods than harms to humanity/society? How can we avoid harmful consequences of AI? Take our example of AI-driven treatment for AD, how can we make good use of the results generated from AI causal models to extend justice and healing on vulnerable dementia population, instead of generating wrong prediction of treatment leading to poor medical advice or judgement? By justice in the Alzheimer's context, we mean that each patient, with the advancement in AI-driven drug discovery and biomarker identification technologies, is given equal right to access to advanced AI-driven AD diagnostics and treatments, irrespective of one's race, age, and socio-economic background, etc.

3. How do we help the public understand how AI works, its potentials and limitations, and acknowledge humbly what we do and do not know with AI?

If AI is meant for Social Good, along the lines of human flourishing, then it must be understood by the general public, and must address the needs of the society. Given that humans are created and non-omnipotent, humans are faulty instead of faultless, the AI system that humans create, even in cases when the system runs automatically, free of human control, would still be far from perfect. While the world is fascinated by what the AI and big data system can offer, it is important that we recognize both its potentials and limitations. For instance, while AI can help make estimation or prediction of the onset of a disease, such as AD, its results might still be subject to uncertainties, biases, and errors in input data and model design. Hence, an essential element of AI for Social Good should include a humble acknowledgment of (i) the limitations and flaws of AI, and (ii) the possible oversight of the orders/truths, virtues that Biggar (2022) affirmed in scholarship infused by God's moral order.

Criteria of AI for Social Good

Hence, in future, AI for Social Good should aim to answer the theologically-driven questions above. The system we champion is guided by values derived from what is biblically desirable, such as building capacities for reasoning and interpretability, reducing biased decision-making, allowing humans to provide expert inputs and guide AI operations, and to take control of the system whenever life-critical decisions have to be made; aiming to meet the needs of our society, such as improving the quality of life and enhancing the ability of the public to make sense of results generated by AI models [5]. It should also be able to build in itself a capacity to acknowledge limitations and partialities, or uncertainties that AI decisions will generate, leading us to acknowledge that we are simply human and we do not hold a complete picture of the ultimate truth and reality. In the following, we discuss the criteria in terms of AI methods, processes, and outcomes.

1. Methods

1.1 Reasoning and Interpretability

AI for Social Good should allow users to comprehend and trust the outputs created by AI algorithms by adopting interpretable or explainable AI methodologies. Injecting reasoning and interpretability into AI algorithms is crucial for organizations and individuals to build trust and confidence in AI algorithms, by understanding how it works, what works best, and their limitations. This is consistent with Biggar's idea that the rationality of the creative order, as it applies to the natural order, must leave rooms for proper reasoning, for reflections on what are truths and non-truths, through repeated experiments and verifications [A].

1.2 Bias Removal

AI for Social Good should develop AI algorithms and datasets to reduce biases. They can be designed with debiasing techniques that avoid creating unjust results, such as biases due to incomplete or small datasets, missing and noisy data, human-in-the-loop subject data, lack of expert guidance etc. [B] While biases are unavoidable due to human limitations and imperfections, attempts to remove biases are important as they imply that humans are constantly searching for the perfect natural order, but are yet attaining such perfection.

2. Processes

2.2 Transparency and Accountability

AI for Social Good must account for the incentives and limitations of any human involvements in AI design, and provide appropriate opportunities for feedback, relevant explanations, and appeal. AI models should build a capacity to allow appropriate human direction and guidance whenever expert knowledge is critical to guide model operations and improve the quality of model outputs.

2.2 Security and Privacy

AI for Social Good must process data in a secure fashion and fully respect the principles of human rights and privacy. There is a need for balancing the privacy of users while achieving greater social good via AI models. For instance, in order to improve the accuracy of AD biomarkers identification, AD subjects' data used for AI model training must be handled with care. While collecting and using such data for AI inputs can potentially benefit a large number of people suffering from AD, the rights and privacy of individual subjects providing clinical information must be safe-guarded.

3. Outcomes

3.1 Moral and Ethical Implications

AI for Social Good must ensure that "every person involved in AI development must exercise caution by anticipating … the adverse consequences of Artificial Intelligent Systems use and by taking the appropriate measures to avoid them." [9]. While it is still debatable whether a machine or an AI algorithm should be delegated with moral or ethical responsibility, namely, a "moral machine" [10], it would be important for human beings to be able to take back the control of making moral or ethical decisions, in case of emergency.

3.2 Public Understanding

AI for Social Good projects must enhance the ability of public to understand and make sense of the results for their own decision-makings. As implied by Biggar, along with his idea on value-driven prosperity and human flourishing, what is considered as social good can only happen when the public are free to make decisions for their own good, and for the good of their own societies [C].

Acknowledgment

References

[1] Gulick, W. (2020) Forms of Emergence, Tradition and Discovery, the Journal of Polanyi Society, 46, 1, 55-59
[http://polanyisociety.org/TAD%20WEB%20ARCHIVE/TAD46-1/Gulick-TAD46-1-pg55-59-pdf.pdf]

[2] Agler, D.W. (2020) Emergence from Within and Without: Juarrero on Polanyi's Account of the External Origin of Emergence, Tradition and Discovery: the Journal of Polanyi Society, 40, 3, 23-35
[http://polanyisociety.org/TAD%20WEB%20ARCHIVE/TAD40-3/TAD40-3-fnl-pg23-35-pdf.pdf

[3] Jumper, J. et al. (2021) Highly accurate protein structure prediction with AlphaFold, Nature, 596, 2021, 583-589 [https://www.nature.com/articles/s41586-021-03819-2]

[4] Lim, S.G. et al. (2021) A review on compound-protein interaction prediction methods: Data, format, representation and model, Computational and Structural Biotechnology Journal, 19, 2021, 1541-1596. [https://www.sciencedirect.com/science/article/pii/S2001037021000763]

[5] Li, V.O.K., Lam, J.C.K., Cui, J., "AI for Social Good: AI and Big Data Approaches for Environmental Decision-making," Elsevier Environmental Science and Policy, Vol. 125, Nov 2021, pp. 241 - 246.

[6] Zhang, Q., Han, Y., Li, V.O.K, Lam, J.C.K., "Deep-AIR: A hybrid CNN-LSTM framework for fine-grained air pollution estimation and forecast in metropolitan cities," accepted for publication, IEEE Access.

[7] Han, Y., Li, V.O.K, Lam, J.C.K., and Pollitt, M., "How blue is the sky? Estimating the air quality data in Beijing during the Blue Sky Day Period (2008-2012) by the Bayesian Multi-Task LSTM approach," Elsevier Environmental Science and Policy, Vol. 116, Feb 2021, pp. 69 -77.

[8] Li, V.O.K, "AI for social good: A case study of near real-time street-level air pollution estimation and public health management," *IEEE HK Section 50th Anniversary Magazine*, IEEE, Hong Kong, Nov 2022, p. 12 – 13.

[9] Canada-ASEAN Business Council (2021). *The Montreal Declaration for the responsible development of artificial intelligence launched*.
https://www.canasean.com/the-montreal-declaration-for-the-responsible-development-of-artificial-intelligence-launched/ [Retrieved on 3rd August 2021]

[10] MIT Media Lab (2021). *Moral machine*. https://www.moralmachine.net [Retrieved on 2nd August 2021]

End Notes

[A] For example, in [6], we have developed Deep-AIR, an AI model to estimate and forecast air quality based on input data which influence air quality, such as traffic congestion, wind direction, wind speed, and urban morphology. Based on saliency scores generated by Deep-AIR, we can determine the relative contributions of different input features on the output air quality, expected impacts and potential biases.

[B] Recent statistical research examining China's air quality data has posed questions on data accuracy, especially data reported during the Blue Sky Day (BSD) period (2000–2012). In [7], we propose a multi-task machine-learning model to re-estimate the official air quality data during the recent BSD period, from 2008 to 2012. Results have shown that average re-estimated daily air qualities are respectively 56% and 55% higher than the official ones, for air quality index (AQI) and AQI equivalent $PM_{2.5}$, during the BSD period, from 2008 to 2012.

[C] In [8], we described an AI/big data project which resulted in the deployment of UMeAir, a smartphone app based on Deep-AIR, which not only displays estimated and forecasted air quality throughout the city, but also gives travel route advice so the citizens can avoid bad air and enjoy improved quality of life. For example, with UMeAir app, asthmatics can use the timely information provided to avoid bad air; this is especially important during the high episode days. For severe cases, such simple alert and advice can be life-saving.

For more information