Jared Bielby, Rachel Fischer

Introduction to AI, Ethics & Society – Part 2

The International Review of Information Ethics (IRIE) has pioneered ethics in technology and information studies for almost two decades. Leading on topics ranging from gaming to artificial intelligence, IRIE has ventured some of the most thought-provoking conversations of the digital age. In launching the second part of AI, Ethics & Society, we are called to reflect upon entering the second year of the Covid-19 pandemic and the various implications it has on society.

The first part of AI, Ethics & Society brought together a set of papers that were presented at a conference, hosted by the Kule Institute for Advanced Study that has made AI, Ethics and Society an area of research focus. The conference was held in May of 2019. Details can be found on the KIAS web site. The first part considered themes from ethical best practices for industry and government developing responsible AI services to aligning cultural and societal values in AI design, the role of researchers from social sciences and humanities disciplines in ethical innovation in the AI sector, and methods for interdisciplinary and intersectoral collaboration between interested in responsible AI. It looked at how nations can contribute to building a healthy AI sector through policy, research and innovation.

Part 2, presented now in Volume 29, is a continuation of the previous volume and presents articles inspired by the KIAS Conference of May 2019. In addition to this, Volume 29 contains papers presented during the AI4IA Conference hosted virtually on 28 September 2020, hosted in collaboration with the Kule Institute for Advanced Studies (KIAS) and AI for Society at the University of Alberta, Canada, as well as the International Centre for Information Ethics (ICIE). The UNESCO Information for All Program (IFAP) Working Group on Information Accessibility (WGIA) presented the Artificial Intelligence for Information Accessibility (AI4IA) and was held under the auspices of UNESCO IFAP, in observance of the International Day for Universal Access to Information (IDUAI).

As stated in the AI4IA report, the wide-ranging impact of the COVID-19 pandemic at every level of society has exposed a number of vulnerabilities in many countries. Nations must, in the short term, re-orientate their policies and legislation within various UNESCO areas of expertise. UNESCO and other intergovernmental organisations must also continue to address inequalities, particularly in terms of information and knowledge management, information accessibility and the challenges of illiteracy in the use of Information and Communication Technologies (ICTs) and Artificial Intelligence (AI). The following recommendations apply:

- Making AI accessible is a collaborative effort between the public sector, private sector and communities. Dialogues with civil society in national, regional and international levels are encouraged to ensure the inclusion of all in issues related to ethics of AI;
- Communities have an important role to play, we should not underestimate them in a fast-changing and ever evolving world;
- Cultural diversity must be central in design, roll-out and training of AI and tools towards ensuring information accessibility;
- In addition to the existing Information Ethics (IE) guidelines for schools and training institutions, specific skills for training of learners in coding should be formulated and creators of algorithms should receive intensive training in Information Ethics. Early childhood education must extend to formal, informal, and non-formal education as well as life-long learning;
- Ethics, transparency, human dignity and the rights of children must be promoted and implemented from the start of the development of any AI systems to their effective use; and
- Creation of special grants for small and developing countries to reduce the technological divide between the South and North, and the inequalities within (such as between rural and urban regions).

ISSN 1614-1687

Industry must be kept as accountable to its intentions and actions with AI as intergovernmental organizations like UNESCO are, since it is now industry, through digital innovation, that exerts more influence over, and thus more control of society than all government agencies combined. The transfer of power structures from government to corporate platforms is no secret and has been explored in depth alongside other digital disruptions and phenomena such as the changing dichotomies between nationalism, localism, and globalism. Indeed, the vices and virtues of the "Internet Giants" have been called out repeatedly, and all too often for their failures, as for instance with the embarrassment of the Facebook / Cambridge Analytica data abuse scandal (Isaak 58).

Since 2018, an emerging theme of an AI Ethics 'culture' has taken shape within the ICIE community, as reflected in the recent editions of IRIE as well as in the editorial of the current edition. The current articles of IRIE demonstrate a need for more than mere AI Ethics principles. Beyond even the potential to enforce such AI Ethics principles, an AI Ethics culture is the foundation upon which equitable AI societies can exist and is something that cannot be enforced by the many AI Ethics principles that have come to prominence in the last few years. As has been demonstrated time and time again throughout the inaugural age of AI, the ethics and inter-relations of humans and AI cannot be left to governments and corporations alone. In addition to the inability of organizations and institutes to enforce AI ethics guidelines – and if they are 'enforced', are they ethics, or law? – the monopoly of AI ethics by said organizations has done very little to date in terms of education and literacy for the general population.

While initial efforts have been made by a number of grassroots and intergovernmental organizations to bring an honest awareness of AI Ethics principles to the general population (Borenstein 62), the appropriation and commodification of the AI Ethics 'trend' by institutes and corporations has by far outshadowed efforts towards education and social responsibility. Just as was the case with the field of Information Ethics a decade ago, which found itself commodified for gain (Enright 105), so too does the emerging field of AI Ethics risk commodification. Indeed, any corporation that has not over the last year or two adopted or purchased some form of AI Ethics principles, either through external stakeholder agreements or through consultation with experts, may risk reputation, relevancy and profit loss. AI Ethics cannot be ignored as COVID-19 accelerates the demand for and use of AI backed technologies. Industry understands this. The need for COVID-19 contact-tracing apps and enforced quarantine surveillance policies have quickly been met by well-intentioned capitalist solutions. Several corporations have produced leading AI Ethics guidelines seemingly overnight, despite previously never having dealt in the field of ethics at all, let alone in information ethics (Hao). Following such popular data abuse debacles as Cambridge Analytica, business leaders now understand that a colorful demonstration of a company's AI Ethics principles could make or break the success of that company. Society now insists upon and expects that governments and corporations demonstrate an ethics prowess in response to emerging technologies. Every corporation that utilizes AI and algorithms -- are there any that don't? -- is now expected to prove their leadership in ethics. Ironically, the general population insisting on AI ethics from government and industry remain dangerously uneducated themselves about the importance of AI Ethics.

Thus, the authors represented in this Volume address the available recommendations insofar as they seek to expand on the complexities of AI and its impact on society, international law, children's rights, education and our understanding of information ethics in relation to the ethics of AI. While acknowledging the value of AI Ethics guidelines, and while pulling from those guidelines, the authors herein advocate for education, accountability, digital literacy and social responsibility.

In *Constructing AI: Examining how AI is shaped by data, models and people*, Katrina Ingram looks at how AI is becoming part of our digital infrastructure, and specifically at how it seems already intricately integrated. As she points out, there are examples of how these AI and their applications can be seen and experienced as harmful, unjust and discriminatory. She argues that these systems always exist within a

socio-cultural context that reflects the data used in their training, the design of their mathematical models and the values of their creators. Recommendations are made that the construction of AI needs to change if we want to build AI systems that benefit society; we need to change how we construct AI. The ethical and technical challenges of AI, specifical as it is used to moderate online content, is the theme of the article by Diogo Cortiz and Arkaitz Zubiaga. They conduct a case study, wherein they utilize an AI model to detect hate speech on social networks. Following a discussion on this case study, they argue that while AI can play a central role in dealing with information overload on social media, it could risk violating freedom of expression. The article concludes by recommending that AI can be used to monitor and assist in detecting online hate speech.

Contributing to the discussion on AI Innovation, Cordel Green and Anthony Clayton's contribution is situated within the context of the 4th Industrial Revolution (4th IR). In their article they deliberate on finding a way to mitigate the negatives of the 4th IR without impairing the extraordinary potential of AI to accelerate all areas of human development. They argue that implementing ethical AI will require a multi-modal and co-regulatory approach. Jandhyala Prabhakar Rao and Rambhatla Siva Prasad focus on the tangible and intangible impact of the use of AI. They also situate this discussion within the framework of the IDUAI in which information access is prioritized. They argue for the need to evolve nation-specific policies and regulations addressing the issues of inequalities and positions it within effective multi-stakeholder collaborative processes. Transitioning from nation-specific policies to international modalities, Fatima Roumate expressly discusses the role of International Human Rights law. She does so by addressing international mechanisms and ethics as new rules which can ensure the protection of human rights in the age of AI.

Isabella Henriques and Pedro Hartung's article considers the direct and indirect impact of AI on children in their article *Children's Rights by Design in AI Development for Education*. The article highlights the legal duty to respect and protect children's rights by all stakeholders involved in the design, implementation and usage of any AI-related technology or service. These recommendations are aligned with UNICEF's Policy Guidance on AI. In consideration of moral rights as assigned to humans and other natural entities, Howard Nye and Tugba Yolbas' article questions whether we can rationalise that an AI System can be seen to be a moral patient, assigned with both autonomy and owed duties of non-maleficence (i.e. to cause no harm). In *Artificial Moral Patients: Mentality, Intentionality, and Systematicity*, they argue that we already fail to respect and uphold the well-being of existing biological moral patients, and therefore there are compelling moral reasons to avoid treating AI Systems as moral entities.

Moving on from moral agency, international human rights and childrens' rights, the next three articles include considerations of value pluralism and a variety of ethics interpretations. Catharina Rudschies, Ingrid Schneider and Judith Simon state that in the current debate on the ethics of AI, much attention has been paid to find some "common ground" in the numerous AI ethics guidelines. They analyse the AI ethics landscape with a focus on divergences across actor types (public, expert, and private actors) and discuss the findings in the light of value pluralism. Conversely, in their article on the relationship (if any), between ethics of AI and Information Ethics, Coetzee Bester and Rachel Fischer unpack the various definitions and applications of these two examples of applied ethics. They acknowledge that although there certainly is convergence between the two, they do also stand separate to one another, predominantly due to Information Ethics being a longer established field than that of the ethics of AI or even data ethics. However, they recommend that popular dialogue and current research should not overcomplicate the differences and instead focus on the similarities, insofar as it contributes to much required awareness on the various implications of AI and the responsibilities of those who design and implement these technologies. Complementing the horizontal analysis of value pluralism, Jared Bielby explores what might be deemed a vertical analysis of value pluralism in his article, *Artful Intelligence: AI Ethics and the Automaton*, by bringing to light the historical origins of AI across time and culture. Bielby frames the

historical human-automaton relationship as a basis for value pluralism in mitigating issues such as function, bias, discrimination and unfairness in AI. He compares and contrasts perspectives on AI as influenced by Shinto influences on Japanese robot culture, the myth of the golem from Judaism, and accounts of automata from ancient Greek mythology.

This edition is complemented by a Book Review done by Rafael Capurro on *Informationsethik und Bibliotheksethik. Grundlagen und Praxism*. This is a book by Hermann Rösch, professor emeritus of Information Ethics at the Institute of Information Science, TH Köln, University of Applied Sciences. According to Capurro it is the first up-to-date presentation of Information Ethics and Library Ethics, that are intimately related but have their own foundational and practical issues. The book consists of four chapters, which look at ethics as a discipline within philosophy, the comparisons between normative and applied ethics as well the differences between personal and professional ethics. The book then proceeds to discuss information ethics and library ethics and the main issues found within these disciplines.

Thus, in the recognition that an equitable AI Ethics requires a deeper and more nuanced understanding of factors such as culture, social responsibility, law, inclusion, access, progress, and general human well-being, the authors of the current edition engage both unanswered questions and potential solutions to the development of an equitable AI-based culture. Moving beyond paying simple lip service to the many AI Ethics principles and guidelines available, they seek to bring the battle to the front lines – to parents wishing to make informed decisions about their children's digital exposure, to service workers and civil servants tasked with complex information and knowledge management practices, to the illiterate and literate alike whose experience and understanding of AI, algorithms and networks define their choices and options, where concerns of the 'digital divide' run so much deeper than mere access. Information accessibility and the challenges of illiteracy take on a new meaning in a COVID-19 world, where in many ways the playing field is levelled, and the lasting implications of an accelerated use of AI are ambiguous to designer and user alike.

References:

Borenstein, Jason, and Ayanna Howard. "Emerging challenges in AI and the need for AI ethics education." AI and Ethics 1.1 (2021): 61-65.

Enright, Nathaniel F. "You Can't Polish a Pumpkin: Scattered Speculations on the Development of Information Ethics." Journal of Information Ethics 20 (2011): 103-126.

Hao, Karen. "In 2020, let's stop AI ethics-washing and actually do something." (2019).

Isaak, Jim, and Mina J. Hanna. "User data privacy: Facebook, Cambridge Analytica, and privacy protection." Computer 51.8 (2018): 56-59.