Centring dignity in algorithm development: testing a Dignity Lens

Centring dignity in algorithm development: testing a Dignity Lens

Lorenn P. Ruster

School of Cybernetics, Australian National University, lorenn.ruster@anu.edu.au

Paola Oliva-Altamirano

SmartyGrants Innovation Lab, Our Community, paolao@ourcommunity.com.au

Katherine A. Daniell

3A Institute, School of Cybernetics & Fenner School of Environment and Society, Australian National University, katherine.daniell@anu.edu.au

Against a backdrop of algorithms that disempower, dehumanise, disenfranchise and discriminate, there are increasing calls to centre the human in AI development processes and to humanise AI development in practice; centring dignity in AI development could provide a way forward. Despite the inclusion of dignity in many Artificial Intelligence (AI) ethics frameworks, like many other AI ethics principles, there is little operational understanding of what dignity can look like in practice when it comes to the development of algorithms. Drawing on cybernetics and a model of dignity developed in the field of international conflict resolution, this paper presents our work-in-progress tool - the Dignity Lens - for considering dignity throughout the AI development lifecycle, and practitioner reflections from using the tool. This work is an initial step towards articulating what dignity-centred AI development could look like in practice, assisting practitioners designing and developing algorithms to actively consider dignity.

CCS CONCEPTS • Human-centered computing ~Human computer interaction (HCI) ~HCI theory, concepts and models

Additional Keywords and Phrases: Dignity, AI Ethics, Interaction Design, Cybernetics

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 INTRODUCTION

Artificial Intelligence (AI) ethics instruments – frameworks, principles, guidelines, policies, tools - have enabled new conversations around what principles could and should be at the heart of AI systems and how they could manifest in practice [1,30,44]. A plethora of these instruments have emerged in recent years; [29] identified at least 106 instruments alone, and a range of others have been compiled by [4] and many others. Although AI ethics instruments serve as a needed step towards addressing concerns around the social and ethical issues surrounding AI systems, turning these principles into practice is a known and deeply felt challenge, reflected in increasing calls for further work to find ways to operationalise AI ethics principles [see for example 43].

To date, much of the work to operationalise AI ethics principles focuses on principles such as fairness [26], accountability [39], and transparency [11,38,55]. There are, however, other principles also worthy of exploration. The principle of interest in this paper, is that of dignity. Dignity has received relatively little attention in the field of AI ethics, particularly when it comes to how to operationalise it in practice. Like many of the AI ethics principles, the meaning of dignity can seem nebulous and even appear paradoxical [52] and, as a result, is certainly challenging to operationalise. However, the importance of dignity cannot be understated: dignity has been described as central to being human, critical to human rights, and even a core value underpinning democracy [33].

A review of eighty-four AI ethics instruments [25] classified eleven commonly referenced principles, one of which was dignity. For [25]:

> "While dignity remains undefined in existing guidelines, save one specification that it is a prerogative of humans but not robots, there is frequent reference to what it entails: dignity is intertwined with human rights or otherwise means avoiding harm, forced acceptance, automated classification and unknown human–AI interaction. It is argued that AI should not diminish or destroy, but respect, preserve or even increase human dignity. Dignity is believed to be preserved if it is respected by AI developers in the first place and promoted through new legislation, through governance initiatives, or through government-issued technical and methodological guidelines." [25]

This summary of what dignity means in the context of existing AI ethics instruments raises many questions: How might AI developers practically respect dignity? What tools may help them to do so? And what role might developers play beyond compliance with legislation, governance initiatives and/or government-issued guidelines?

This paper shares our work-in-progress Dignity Lens which aims to assist practitioners to practically respect dignity, by operationalising dignity in the context of AI system design, development and implementation. It draws upon Hicks' 10 essential elements of dignity, as published in the field of international conflict resolution [20] and organisational leadership [21] (see Table 1 below). For Hicks [20], at the heart of dignity is a desire to be seen, heard, listened to, and treated fairly; to be recognized, understood, and to feel safe in the world. In addition to incorporating this view of dignity, the Dignity Lens also employs approaches known in cybernetics to study systems such as focusing on feedback loops, relationships and dynamics. In doing so, it extends Hicks' [20] model of dignity from 10 elements to an ecosystem of protective and proactive mechanisms that exist in a dynamic balance.

2 RELATED WORK

The creation of the Dignity Lens is grounded in Hicks' [20] dignity model and cybernetic approaches which, together, yield a conceptual view of dignity as an ecosystem.

2.1 Adapting Hicks' 10 essential elements of dignity [20]

Hicks' dignity model [20] stems from her experiences in international conflict resolution and mediation and has since been applied in a variety of organisational leadership contexts [21]. The model describes 10 essential elements of dignity and 10 temptations to violate dignity (see Table 1).

Table 1: Hicks' 10 essential elements of dignity and 10 temptations to violate dignity [20]

10 essential elements of dignity [20]	10 temptations to violate dignity [20]
1. Acceptance of identity	1. Taking the bait
2. Inclusion	2. Saving face
3. Safety	3. Shirking responsibility
4. Acknowledgement	4. Seeking false dignity
5. Recognition	5. Seeking false security
6. Fairness	6. Avoiding conflict
7. Benefit of the doubt	7. Being the victim
8. Understanding	8. Resisting feedback
9. Independence	9. Blaming and shaming others to deflect your own guilt
10. Accountability	10. Engaging in false intimacy and demeaning gossip

Discourse analysis was applied to how dignity was used in Hicks' book [20] to understand the types of unwritten mechanisms and actions associated with dignity in practice. From this analysis, three types of mechanisms and actions were identified. Firstly, protective mechanisms and actions associated with preventing dignity violations and/or remedying dignity violations; secondly, mechanisms and actions associated with promoting dignity; finally, there is a recognition that both protective and proactive mechanisms are underpinned by mechanisms and actions acknowledging dignity.

2.2 Employing cybernetics approaches

The establishment of the Dignity Lens prototype employs cybernetic approaches. Cybernetics is interested in the science of communication and control between machines and people, specifically with how information flows within and between purposeful systems [19]. [37] reflects on the connection between HCI and Cybernetics:

> "…cybernetics is a "science of interactions" for which human-computer interaction is a subset…With its practical tools for modelling purpose, feedback, and autonomy, cybernetics has something to offer [HCI].". [37]

Further, Cybernetics and artificial intelligence are intimately connected; several members of the Macy conferences which enabled the flourishing of cybernetics, went on to be involved in establishing artificial intelligence as a research agenda [7]. Although cybernetics is not explicitly focused on ethics, one of the founders of cybernetics, Norbert Wiener, is seen as critical to the founding of information ethics [22]. It is thus unsurprising that cybernetic approaches are relevant to this work.

Cybernetics spans a wide range of disciplines, engaging with philosophers, computer scientists, anthropologists, mathematicians, physicists and others and accordingly has many of its own approaches and ways of thinking. Influential to the development of the Dignity Lens are cybernetic ideas surrounding feedback loops, emphasising relationships, thinking in systems and transdisciplinarity [see for example 2,6,18,41,56]. These ideas are represented in the visualisation of dignity as an ecosystem which takes Hicks' [20] use of the term dignity and transforms it into the roles that individuals, organisations, governments and others can play *in relation to* dignity (protective, proactive, acknowledging roles). In doing so, it takes a systems view and privileges a dynamic sense of dignity to be used across disciplines to interrogate how dignity could be, or has been, put into practice.

3 BRIDGING THE PRINCIPLES-TO-PRACTICE GAP: THE PROTOTYPE DIGNITY LENS

The AI ethics principles-to-practice gap is well documented [see for example 45] and is the focus of much research effort, particularly when it comes to principles like transparency, fairness and accountability. Dignity, despite being one of the common principles highlighted by [25] and present in thirteen (13) of the eighty-four (84) AI ethics instruments reviewed, has received limited attention in terms of how to move from principles to practice. Even for [44], who proclaim to have "compiled the most comprehensive document collecting existing guidance which can guide practical action [on AI ethics]", the principle of dignity appears elusive. Guidance on dignity speaks to respecting human beings' intrinsic value, ensuring dignity is not violated and being clear about when a user is interacting with an AI and not another human [44]. Although a helpful starting point, this "most comprehensive document" is far from achieving its purpose of guiding practical action.

To bridge the principles-to-practice gap when it comes to dignity, a different approach was taken in the work of this paper. Instead of compiling what has already been presented in existing AI ethics instruments regarding dignity, the Dignity Lens draws on the principle of interdisciplinarity, also found in cybernetics [see for example 18], by reviewing what dignity as a concept could look like in practice from a variety of perspectives, including philosophy and human rights [5,9,52], nursing practice [14], constitutional law [40], international development [31,32] and non-Western perspectives [8,23,24,28]. Hicks' model of dignity [20], comprising 10 essential elements of dignity, was chosen as an initial starting point to understand its applicability to AI development. This model was infused with approaches common to the field of cybernetics, such as thinking in systems, relationships and dynamics [see for example 2,6,18,41,56]. As a result, an ecosystem view of dignity underpinning the Dignity Lens was generated (see Figure 1).

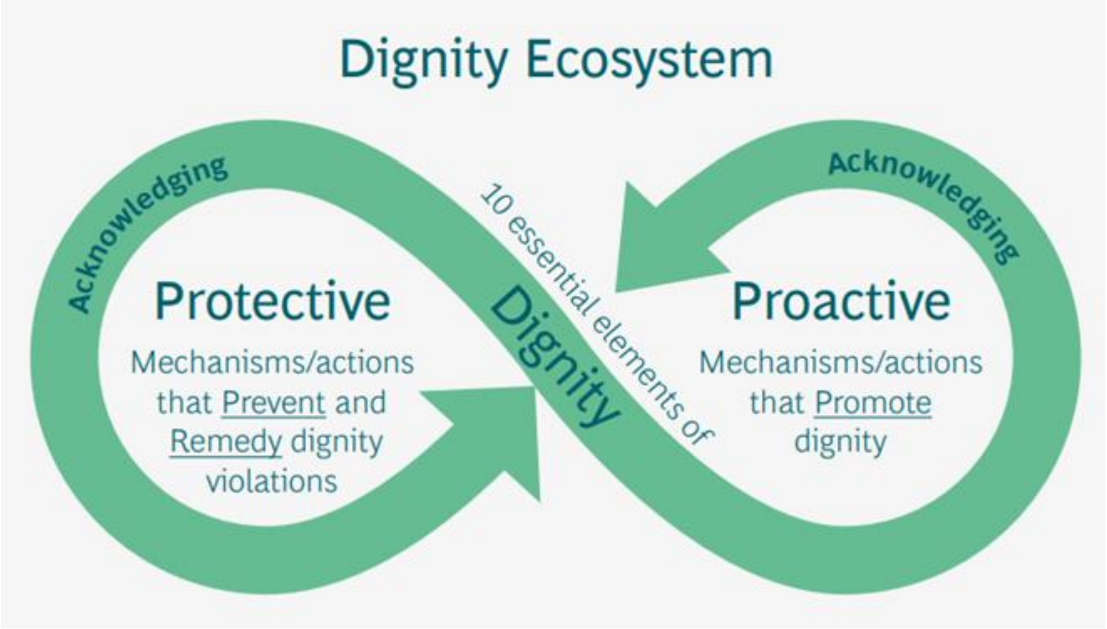
Dignity Ecosystem
Acknowledging
Acknowledging
10 essential elements of
Dignity
Protective
Mechanisms/actions that Prevent and Remedy dignity violations
Proactive
Mechanisms/actions that Promote dignity

Figure 1: Dignity as an ecosystem, taken from [42]

The ecosystem view of dignity shows a dynamic interplay between protective, proactive and acknowledging mechanisms and actions. Mechanisms and actions that take a protective stance prevent dignity violations from occurring and/or remedy for those that do occur. Examples could include ensuring that algorithms comply with anti-discrimination laws or taking steps to identify and mitigate against bias in datasets. Those that take a proactive stance, actively promote dignity. Examples of proactive mechanisms and actions include actively creating feedback processes for end-users to be heard and their needs reflected in algorithmic design and monitoring, and the hiring of diverse teams with lived experience related to where the algorithm will be used. Both protective and proactive mechanisms and actions are underpinned by an acknowledgement that dignity exists and is important. The dignity ecosystem assumes that there is a required balance between protective and proactive mechanisms to enable dignity.

3.1 Prototype Dignity Lens

To assist practitioners with what dignity-centred development looks like in practice, the first version of the prototype Dignity Lens sought to identify mechanisms and actions that enabled a dignity ecosystem. These mechanisms and actions are organised by two dimensions: firstly, Hicks'10 essential elements of dignity [20] and secondly, whether the mechanism or action was protective (that is, preventing or remedying for a dignity violation) and/or proactive (that is, promoting dignity) in nature. It is envisioned that the Dignity Lens will be used in contexts where designers, developers and other practitioners are keen to embed dignity prospectively in their design and/or check for the extent to which their algorithm upholds dignity as a retrospective, reflective process.

As described by [42], a first test of the Dignity Lens was conducted in 2021 to understand the extent to which dignity was present in the AI ethics instruments of the governments of Australia, Canada and the United Kingdom. This work was conducted by using the Dignity Lens to analyse publicly available documents. Although it provided an initial helpful starting point for understanding the usefulness of the application of the Dignity Lens, we were keen to trial it in collaboration with practitioners.

Table 2: Example of applying Dignity Lens in practice, prototype 1

Essential element of dignity [20]	Mechanisms/ actions		
	Protective		Proactive
	Prevention	Remedy	
1. Acceptance of identity	• Compliance with antidiscrimination laws • Impact assessments (e.g., unintended consequences assessment, privacy impact assessments, Equality Impact Assessment etc.) • Risk assessments • Testing for unintended (data) biases	• Bias mitigation • Other mitigation	• Consultation with affected populations • Involvement of diverse expertise (e.g., external stakeholders) • User-centricity

4 EXPLORATORY STUDY: APPLYING THE DIGNITY LENS TO CLASSIEFIER

The main purpose of our exploratory study was to iterate upon the Dignity Lens by co-reflecting on its usefulness with practitioners keen to apply it in their own context. The approach of co-reflection and iteration leans into the connection between cybernetics and design: namely that design can be seen as the practical arm of cybernetics, embodying notions of circularity, iteration and unknowing [16]. We actively leaned into these elements through the exploratory study, iterating on the Dignity Lens in real time, returning to how to apply it in a circular fashion and embracing the unknown of how useful it will be in this context. In this case, the practitioner involved was a data scientist designing and developing an algorithm, however it is thought it may have wider applicability to engineers, policymakers and others. Similarly, this study focuses on the retrospective application of the Dignity Lens to the design and development of CLASSIEfier – a text auto-classification system used to classify grantmaking records - however, it is thought to also have potential applicability in design.

4.1 CLASSIEfier

CLASSIEfier is a keyword-matching model used as a data science solution within a cloud-based grants administration system, developed by and housed within an Australian social enterprise, Our Community. The grants administration system is used by over 370 government, philanthropic and corporate grantmakers and manages the flow of more than 4 billion Australian dollars in grants every year across tens of thousands of applicants in 32 countries [48]. CLASSIEfier was developed to enable better tracking of Australian grants by systematically classifying social sector initiatives and entities. It classifies against CLASSIE - a taxonomy adapted from the Philanthropic Classification System - and against the Sustainable Development Goals [35]. Other taxonomies will be added in the future. CLASSIEfier enables benchmarking among grantmakers and a deeper understanding of their funding distributions and impact, informing prioritisation of grant assessments and longer-term program planning.

The risks of CLASSIEfier producing an incorrect classification are wide-ranging. Classification errors could mislead funding distribution decisions for individual grantmakers and affect the validity of impact evaluations and decisions based on CLASSIEfier's aggregated data outputs (see [49,50] for example outputs). Given the high concentration of Australian grantmakers and applicants (over 50,000 grant officers, over 10,000 grant programs on the platform and nearly half a million applications submitted using the platform in the past year [51]), the potential impact of errors is significant.

The data scientist who developed CLASSIEfier has taken several steps to improve the transparency and explainability of the algorithm and to ensure stakeholder engagement, testing and iteration based on feedback. However, there were still many questions regarding the extent to which dignity was upheld in the design and development of CLASSIEfier. Since the grants administration system sits in a social enterprise serving over 80,000 not-for-profit members and has a service pledge to "be human" in their interactions [36], interrogating the development of CLASSIEfier through the lens of dignity was seen as a valuable endeavour. To date, the data scientist who developed CLASSIEfier has not found a helpful tool to assist in such an exploration. Upon release of first prototype of the Dignity Lens [42], the data scientist reached out to use it to analyse how they fared when it came to dignity when developing CLASSIEfier.

5 FINDINGS AND PRACTITIONER REFLECTIONS

An account of using this tool in conjunction with a data scientist who created an automated classification system is reflected upon, as well as the changes to the tool that were made as a result of using it in practice. The tool was applied retrospectively to an algorithm already in deployment; however, it is believed that it may also have use in design and thereby be of potential interest to the HCI community.

5.1 Updates to the Dignity Lens

Upon trialling the Dignity Lens prototype with the data scientist, the importance of the AI development lifecycle became apparent. It was difficult for the data scientist to think from the perspective of the element of dignity in the first instance, whereas thinking about the different decisions made throughout the AI development process was a more intuitive way to have the conversation. Accordingly, the Dignity Lens was iterated upon to foreground the decisions made and actions taken at each stage of the AI development lifecycle, and then the actions taken were connected to Hicks' dignity elements [20] and protective/proactive stances (see Table 3). Stages of the AI development lifecycle were co-created with the data scientist to reflect their process and broadly align with other development lifecycle perspectives [see for example 47].

Table 3: Dignity Lens prototype 2, one example per AI development phase, excerpt from trial with CLASSIEfier [adapted from 34]

In this phase	A decision was made to:	The element(s) of dignity upheld through this decision include:	Protective/ Proactive
Planning and data exploration	Train the model without using personal data	*Safety:* We followed the Innovation Lab data science guidelines to prevent violation of stakeholder privacy and keep data secure.	Protective
Development	Prepare a training dataset which attempts to mitigate for identified data breaches	*Acceptance of identity:* We acknowledged the importance of fairly representing all subjects and populations when training the algorithm *Inclusion and fairness:* We chose to mitigate bias by using only SmartyGrants data in the training dataset. Although public data appeared useful on face value, on closer examination we found that the bias inherent in the public data could harm the dignity of the populations represented in the data and the data owners. For example, data coming from news can unfairly link specific populations to alcohol consumption, family violence etc. *Understanding:* We analysed the different outcomes generated by different training datasets, and their implications for stakeholders' dignity. We adjusted the model to account for our findings	Protective
Testing	Open the algorithm's code for review, and invite data scientist Kabir Manandhar Shrestha to join the Innovation Lab for three months to review CLASSIEfier. His review is summarised in the article "Ethical considerations in multilabel text classifications" [27].	*Acknowledgement and fairness:* We acknowledged that the algorithm developers had limited expertise and welcomed an external reviewer to identify and mitigate bias. *Understanding and accountability:* We collected feedback from the external reviewer and then were held to account to implement the feedback gathered.	Protective – identifying ways to mitigate for bias
Release	Publish the results in plain English.	*Inclusion and accountability:* We tried to cater to diverse audiences by publishing the results using plain and simple English and using clear and understandable visualisations. In doing this, we aimed to increase the likelihood that people would use the outputs to inform their own work. Thus we promoted not only inclusion but also greater accountability – if more people can understand the results, more can hold us to account.	Proactive – maximises accessibility and holds us to account
Review and monitoring	Make CLASSIEfier a live tool, open to feedback.	*Acknowledgement, understanding and accountability*: We are actively collecting and incorporating feedback from users on an ongoing basis. For example, data users and owners can suggest changes to the keywords, the system integration and the user interface(s). This feedback mechanism not only acknowledges their different experiences but also gives them a way of taking control of their experience. In this way we continue to seek understanding of different users' experiences of the tool. Improvements are released iteratively based on lessons learnt.	Proactive – users are seen, heard and listened to on a regular basis

Following this initial data capture, the information could then be rearranged according to element of dignity to better understand where mechanisms and actions that upheld dignity were situated across the stages of AI development and where there were gaps. See Table 4 for an example.

Table 4: Mechanisms and Actions upholding dignity organised by element of dignity, protective/proactive stance and stage of AI development lifecycle, excerpt taken from [34]

Essential element of dignity [20]	Protective / Proactive	Stages of AI development				
		Planning & data exploration	Development	Testing	Release	Review & monitoring
1. Acceptance of Identity "Approach people as being neither inferior nor superior to you; give others the freedom to express their authentic selves without fear of being negatively judged; interact without prejudice or bias, accepting that characteristics such as race, religion, gender, class, sexual orientation, age, and disability are at the core of their identities." [20]	Protective	Data distribution investigation	Preparing a training dataset to mitigate against identified biases Model assessment and selection aligned to needs	Testing with data owners		
	Proactive		Feedback on keywords used in model		Customised user interface	

It also became apparent that it was important to explicitly identify the stakeholders involved at each stage of the development process, as a way of thinking through the various decisions made and their impacts from different perspectives. More explicitly incorporating the identification of stakeholders as a preliminary step before using the Dignity Lens has been a useful procedural addition. Further iteration of the Dignity Lens to encompass stakeholder perspectives is an avenue of future exploration, drawing upon stakeholder theory [17].

5.2 Practitioner reflections

Co-reflection was used as a way to rapidly iterate on the design of the Dignity Lens [54]. [57] defines co-reflection as a way for individuals to explore their experiences and reach new understandings through collaborative meaning making, fostering collaboration and reflective practices [46]. Co-reflection practices also draws upon second-order cybernetics [12,15] to think about how our relative positions as a part of the AI development lifecycle system impacts the use and usefulness of the Dignity Lens. This combination of approaches led to three key learnings.

5.2.1 Mapping to dignity elements and protective/proactive stance prompts new thinking

Firstly, arranging the different mechanisms and actions taken by element of dignity and by proactive or protective stance allowed for new considerations. For example, this mapping of mechanisms to dignity elements revealed that some elements of dignity were less considered or in the case of 'benefit of the doubt', completely missing. The team reflected that consideration of benefit of the doubt could have encouraged useful debate around deciding on where the edges of the system of interest are, also known as 'boundary judging' [3]. For example, considering benefit of the doubt may have prompted consideration about the ‘edge case’ use of the system and the governance structures in place to account for such use.

As a result of the mapping according to protective/ proactive stance, the team also reflected on there being more protective mechanisms than ones that actively promoted dignity. Using the Dignity Lens in this way served as a feedback loop to the team regarding where they have been implicitly focusing. Seeing this imbalance prompted them to think about how they could have created more proactive mechanisms. For the element of safety, for example, they were acting in a protective way purely through mechanisms such as removing personal data from the training set and publishing results in aggregate. Ways to embed the promotion of physical and psychological safety into CLASSIEfier were noted as worthy of further consideration.

In addition, it was reflected that one mechanism could be used in both protective and proactive ways. In their case, a mechanism like scenario planning was employed protectively – to prevent harms and think about worst case scenarios – but equally could have been employed in a proactive stance to maximise dignity in design. Overall, the reflections emanating from the mapping were considered useful and prompted calls to use the tool prospectively in the design phase of future tools (not just retrospectively).

5.2.2 Dignity is considered to different extents across the AI development lifecycle

Secondly, considering the mechanisms according to AI development stage assisted in taking a holistic view that may have otherwise been neglected. For example, the team reflected that more mechanisms were operating in earlier stages of AI development, which was reflective of their experiences as designers and thereby comfort in the design phase. The Dignity Lens revealed to the team what [13] refer to as ‘individual preexisting bias’, and how this was entering the system implicitly and unconsciously. Accordingly, it highlighted opportunities for deeper consideration of mechanisms within the release, review and monitoring and de-commissioning stages. Consideration across the lifecycle also showed them how values that they were carrying into the development process, for example around the value of feedback, were applied inconsistently across the lifecycle phases. That is, despite a conscious effort to preference feedback by the team, only some of the development phases incorporated feedback; it could have had relevance in other phases as well. In doing so, the tool provided a check for the team on the extent to which they were ‘walking the talk’ of putting their values into practice consistently and assisted in identifying opportunities for improvement, not only for future tool development, but also for the continued review and monitoring of CLASSIEfier and its potential de-commissioning in the future. Although it felt somewhat foreign to give weight across all phases of the lifecycle, consideration of mechanisms and actions at different stages of a lifecycle is common within circular economy research [see for example 10]; future directions could look towards this body of research for inspiration regarding how to embed meaningful engagement across different lifecycle stages.

5.2.3 Dignity Lens as a boundary object for increased confidence and integrity in decision-making

Finally, the team reflected that the Dignity Lens gave a language regarding what dignity could mean in practice. The importance of having a way to put words to dignity and have an approach that could be used to have one conversation is reflective of the Dignity Lens presenting as a boundary object that was plastic enough to mould to the context of the team, yet robust enough to enable a united conversation [53]. Being able to use the Dignity Lens in this way increased the team’s confidence that their decisions are aligned with their values and provided a way of documenting decisions for continual improvement. They believed that the robustness of Dignity Lens would hold even if more parties were involved, for example in the design phase. They also believed that using it prospectively in design would enable even more confidence and better decision-making.

6 CONCLUSION AND FUTURE DIRECTIONS

This paper details our work-in-progress tool – the Dignity Lens – which assists in identifying the operationalisation of dignity in the design and development of algorithms, ultimately assisting designers and developers to guardrail against algorithms that dehumanise and disempower. We initially developed this prototype tool through applying Hicks'10 essential elements of dignity [20], integrating discourse analysis on how dignity is used (in protective, proactive and acknowledging ways) and cybernetic approaches. Preliminary investigation using the tool with a data scientist practitioner led to an iteration of the Dignity Lens that more overtly spans across the AI development lifecycle. The testing also pointed toward the importance of undertaking stakeholder mapping before using the tool to make it more contextually relevant. It is believed that more could be done to integrate stakeholder perspectives into the Dignity Lens, drawing upon stakeholder theory [17] and second-order cybernetics, which actively considers the observer of a system (in this case, the person/team applying the Dignity Lens) as a part of the system itself [see for example 12,15,19].

There are many future directions of this research, including testing the Dignity Lens in the design phase to understand whether the Dignity Lens is useful prospectively. In addition, consideration of how to use the Dignity Lens in the context of continual algorithmic monitoring is also of interest, particularly its role in forming a feedback loop between users and developers regarding dignity impacts of an algorithm over time. A core limitation of this work is a lack of detailed justification regarding the choice of dignity model. Accordingly, next steps will further consider dignity models from interdisciplinary contexts in order to test the choice of Hicks' model [20] as a starting point. In addition, it would be helpful to understand the extent to which the Dignity Lens can serve as a boundary spanning object in the face of more complex AI systems, by testing the Dignity Lens in contexts where more complex AI models are being developed and deployed. Nevertheless, the work-in-progress Dignity Lens provides the beginning of a way forward in terms of bridging the AI ethics principles-to-practice gap regarding dignity.

ACKNOWLEDGMENTS

The authors would like to thank members of Our Community, the Centre for Public Impact and the Australian National University's School of Cybernetics for their contributions to the development and iteration of the Dignity Lens. This work was supported by a Florence Violet McKenzie scholarship and an Australian Government Research Training Scholarship.

REFERENCES

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), Association for Computing Machinery, New York, NY, USA, 1–13. DOI:https://doi.org/10.1145/3290605.3300233

[2] Gregory Bateson. 1967. Cybernetic explanation. American behavioral scientist 10, 8 (1967), 29–29.

[3] Gregory Bateson. 2000. Steps to an ecology of mind (University of Chicago Press ed ed.). University of Chicago Press, Chicago.

[4] Kathy Baxter. 2021. Ethical AI frameworks, tool kits, principles, and certifications - Oh my! Salesforce Research. Retrieved January 28, 2022 from https://blog.salesforceairesearch.com/frameworks-tool-kits-principles-and-oaths-oh-my/

[5] Rachel Bayefsky. 2013. Dignity, Honour, and Human Rights: Kant's Perspective. Political Theory 41, 6 (December 2013), 809–837. DOI:https://doi.org/10.1177/0090591713499762

[6] Stafford Beer. 2002. What is cybernetics? Kybernetes 31, 2 (March 2002), 209–219. DOI:https://doi.org/10.1108/03684920210417283

[7] Genevieve Bell. 2021. Talking To Ai: An Anthropological Encounter with Artificial Intelligence. In The SAGE Handbook of Cultural Anthropology. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road London EC1Y 1SP, 442–458. DOI:https://doi.org/10.4135/9781529756449.n25

[8] Jacob Daniel. 2019. Dignity as Respect: A Contemporary Hindu Understanding of Human Dignity. PhD Thesis.

[9] Stephen Dilley, Nathan J. Palpant, Nathan J. Palpant, and Ana Smith Iltis. 2012. Human Dignity in Bioethics : From Worldviews to the Public Square. Taylor & Francis Group, London, UNITED KINGDOM. Retrieved from http://ebookcentral.proquest.com/lib/anu/detail.action?docID=1122866

[10] Ellen MacArthur Foundation. 2019. Artificial Intelligence and the Circular Economy: AI as a tool to accelerate the transition. EllenMacArthur Foundation. Retrieved from https://emf.thirdlight.com/link/dl06eujbcbet-wx40o7/@/preview/1?o

[11] H. Felzmann, E. Fosch-Villaronga, C. Lutz, and A. Tamò-Larrieux. 2020. Towards Transparency by Design for Artificial Intelligence. Science and Engineering Ethics 26, 6 (2020), 3333–3361. DOI:https://doi.org/10.1007/s11948-020-00276-4

[12] Heinz von Foerster. 2003. Cybernetics of Cybernetics. In Understanding Understanding. Springer New York, New York, NY, 283–286. DOI:https://doi.org/10.1007/0-387-21722-3_13

[13] Batya Friedman and Helen Nissenbaum. 1996. Bias in computer systems. ACM Trans. Inf. Syst. 14, 3 (July 1996), 330–347. DOI:https://doi.org/10.1145/230538.230561

[14] Ann Gallagher. 2004. Dignity and Respect for Dignity - Two Key Health Professional Values: implications for nursing Practice. Nurs Ethics 11, 6 (November 2004), 587–599. DOI:https://doi.org/10.1191/0969733004ne744oa

[15] Ranulph Glanville. 2004. The purpose of second-order cybernetics. Kybernetes 33, 9/10 (October 2004), 1379–1386. DOI:https://doi.org/10.1108/03684920410556016

[16] Ranulph Glanville. 2009. A (Cybernetic) Musing: Design and Cybernetics. Cybernetics and human knowing 16, 3–4 (2009), 175–186.

[17] Vincent de Gooyert, Etiënne Rouwette, Hans van Kranenburg, and Edward Freeman. 2017. Reviewing the role of stakeholders in Operational Research: A stakeholder theory perspective. European Journal of Operational Research 262, 2 (October 2017), 402–410. DOI:https://doi.org/10.1016/j.ejor.2017.03.079

[18] Steve Joshua Heims. 1991. The Cybernetics Group. The MIT Press. DOI:https://doi.org/10.7551/mitpress/2260.001.0001

[19] Francis Heylighen and Cliff Joslyn. 2001. Cybernetics and Second-Order Cybernetics. Encyclopedia Phys. Sci. Technol. 4, (March 2001). DOI:https://doi.org/10.1016/B0-12-227410-5/00161-7

[20] Donna Hicks. 2013. Dignity: the essential role it plays in resolving conflict in our lives and relationships. Yale University Press, New Haven, Conn.; London.

[21] Donna Hicks. 2018. Leading with dignity: how to create a culture that brings out the best in people. Yale University Press, New Haven.

[22] Kenneth Einar Himma and Herman T. Tavani (Eds.). 2008. The Handbook of Information and Computer Ethics. John Wiley & Sons, Inc., Hoboken, NJ, USA. DOI:https://doi.org/10.1002/9780470281819

[23] Polycarp A. Ikuenobe. 2016. The Communal Basis for Moral Dignity: An African Perspective. Philosophical Papers 45, 3 (September 2016), 437–469. DOI:https://doi.org/10.1080/05568641.2016.1245833

[24] Polycarp A. Ikuenobe. 2018. Human rights, personhood, dignity, and African communalism. Journal of Human Rights 17, 5 (October 2018), 589–604. DOI:https://doi.org/10.1080/14754835.2018.1533455

[25] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. Nat Mach Intell 1, 9 (September 2019), 389–399. DOI:https://doi.org/10.1038/s42256-019-0088-2

[26] Michael Madaio, Lisa Egede, Hariharan Subramonyam, Jennifer Wortman Vaughan, and Hanna Wallach. 2022. Assessing the Fairness of AI Systems: AI Practitioners' Processes, Challenges, and Needs for Support. Proc. ACM Hum.-Comput. Interact. 6, CSCW1 (March 2022), 1–26. DOI:https://doi.org/10.1145/3512899

[27] Kabir Manandhar Shrestha. Ethical considerations in multilabel text classifications. Our Community. Retrieved August 9, 2022 from https://smartygrants.com.au/research/ethical-considerations-in-multilabel-text-classifications

[28] Thaddeus Metz. 2012. African Conceptions of Human Dignity: Vitality and Community as the Ground of Human Rights. Hum Rights Rev 13, 1 (March 2012), 19–37. DOI:https://doi.org/10.1007/s12142-011-0200-4

[29] Jessica Morley, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2021. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. In Ethics, Governance, and Policies in Artificial Intelligence, Luciano Floridi (ed.). Springer International Publishing, Cham, 153–183. DOI:https://doi.org/10.1007/978-3-030-81907-1_10

[30] Michael Muller, Christine T. Wolf, Josh Andres, Michael Desmond, Narendra Nath Joshi, Zahra Ashktorab, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Evelyn Duesterwald, and Casey Dugan. 2021. Designing Ground Truth and the Social Life of Labels. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21), Association for Computing Machinery, New York, NY, USA, 1–16. DOI:https://doi.org/10.1145/3411764.3445402

[31] Martha Nussbaum. 2008. Human Dignity and Political Entitlements. In Human dignity and bioethics, Barbara T. Lanigan (ed.). Nova Science Publishers, Incorporated, New York, 245–264.

[32] Martha Nussbaum. 2011. Creating capabilities the human development approach. Belknap Press of Harvard University Press, Cambridge, Mass. Retrieved September 18, 2021 from http://site.ebrary.com/id/10488676

[33] Josiah Ober. 2012. Democracy's Dignity. Am Polit Sci Rev 106, 4 (November 2012), 827–846. DOI:https://doi.org/10.1017/S000305541200038X

[34] Paola Oliva-Altamirano and Lorenn P Ruster. 2022. The ethics of automated classification: a case study using a dignity lens. Our Community, Melbourne, Victoria. Retrieved May 11, 2022 from https://smartygrants.com.au/research/the-ethics-of-automated-classification-a-case-study-using-a-dignity-lens

[35] Our Community. 2021. CLASSIE - Classification of Social Sector Initiatives and Entities. Retrieved August 6, 2022 from https://www.ourcommunity.com.au/classie

[36] Our Community Pty Our Community. About Us - Our Community. Retrieved August 6, 2022 from https://www.ourcommunity.com.au/aboutus

[37] Paul Pangaro. 2001. The Cybernetics of HCI: A pragmatic approach. In Seminar on People, Computers and Design. Retrieved August 6, 2022 from https://hci.stanford.edu/courses/cs547/abstracts/01-02/020215-pangaro.html

[38] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Association for Computing Machinery, New York, NY, USA, 1–13.

DOI:https://doi.org/10.1145/3173574.3173677

[39] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20), Association for Computing Machinery, New York, NY, USA, 33–44. DOI:https://doi.org/10.1145/3351095.3372873

[40] N Rao. 2008. On the Use and Abuse of Dignity in Constitutional Law. Columbia Journal of European Law 14, 2 (2008), 69–73.

[41] Thomas Rid. 2017. Rise of the machines: a cybernetic history. Scribe Publications, Brunswick, VIC.

[42] Lorenn P Ruster and Thea Snow. 2021. Exploring the role of dignity in government AI Ethics instruments. Centre for Public impact. Retrieved March 19, 2021 from https://www.centreforpublicimpact.org/partnering-for-learning/cultivating-a-dignity-ecosystem-in-government-ai-ethics-instruments

[43] Emma Ruttkamp-Bloem. 2020. The Quest for Actionable AI Ethics. In Artificial Intelligence Research, Aurona Gerber (ed.). Springer International Publishing, Cham, 34–50. DOI:https://doi.org/10.1007/978-3-030-66151-9_3

[44] Mark Ryan and Bernd Carsten Stahl. 2020. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. Journal of Information, Communication and Ethics in Society 19, 1 (January 2020), 61–86. DOI:https://doi.org/10.1108/JICES-12-2019-0138

[45] Daniel Schiff, Bogdana Rakova, Aladdin Ayesh, Anat Fanti, and Michael Lennon. 2021. Explaining the Principles to Practices Gap in AI. IEEE Technology and Society Magazine 40, 2 (June 2021), 81–94. DOI:https://doi.org/10.1109/MTS.2021.3056286

[46] Donald A Schon. 2017. The reflective practitioner: how Professionals Think in Action. Retrieved August 9, 2022 from http://www.dawsonera.com/depp/reader/protected/external/AbstractView/S9781315237473

[47] Saad Shafiq, Atif Mashkoor, Christoph Mayr-Dorn, and Alexander Egyed. 2021. A Literature Review of Using Machine Learning in Software Development Life Cycle Stages. IEEE Access 9, (2021), 140896–140920. DOI:https://doi.org/10.1109/ACCESS.2021.3119746

[48] SmartyGrants. 2020. Grantmakers: Join the FunderStorm! North Melbourne, Australia. Retrieved June 8, 2022 from https://smartygrants.com.au/uploads/general/SG/SmartyGrantsBrochure.pdf

[49] SmartyGrants. 2021. The Future of Funding: What are the priorities and directions of…. Our Community, North Melbourne, Australia. Retrieved August 6, 2022 from https://smartygrants.com.au/research/the-future-of-funding-what-are-the-priorities-and-directions-of-australian-grantmakers

[50] SmartyGrants. 2022. The Future of Funding: How well are Australian grants addressing the…. Our Community. Retrieved August 6, 2022 from https://smartygrants.com.au/research/the-future-of-funding-how-well-are-australian-grants-addressing-the-un-sustainable-development-goals

[51] SmartyGrants- SmartyGrants. Home. SmartyGrants. Retrieved August 6, 2022 from https://smartygrants.com.au/

[52] Herbert Spiegelberg. 1976. Human Dignity. A Challenge to Contemporary Philosophy. In Human Dignity: this Century and the Next, R Gotesky and E Laszlo (eds.). Gordon & Breach, Amsterdam, 39–64.

[53] Susan Leigh Star (Ed.). 1995. Ecologies of knowledge: work and politics in science and technology. State University of New York Press, Albany.

[54] Oscar Tomico, Joep W. Frens, and C. J. Overbeeke. 2009. Co-reflection: user involvement for highly dynamic design processes. In CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09), Association for Computing Machinery, New York, NY, USA, 2695–2698. DOI:https://doi.org/10.1145/1520340.1520389

[55] Daniel Karl I. Weidele, Justin D. Weisz, Erick Oduor, Michael Muller, Josh Andres, Alexander Gray, and Dakuo Wang. 2020. AutoAIViz: opening the blackbox of automated artificial intelligence with conditional parallel coordinates. In Proceedings of the 25th International Conference on Intelligent User Interfaces (IUI '20), Association for Computing Machinery, New York, NY, USA, 308–312. DOI:https://doi.org/10.1145/3377325.3377538

[56] Norbert Wiener. 2019. Cybernetics or Control and Communication in the Animal and the Machine (Reissue Of The 1961 Second Edition ed.). MIT Press, Cambridge, MA, USA.

[57] Joyce Yukawa. 2006. Co-reflection in online learning: Collaborative critical thinking as narrative. Computer Supported Learning 1, 2 (June 2006), 203–228. DOI:https://doi.org/10.1007/s11412-006-8994-9