AI Design and Policy for Education

Jonathan Adams, Toi Ohomai / Te Pūkenga
Christine Cheyne, Toi Ohomai / Te Pūkenga
Josh Burrell, Toi Ohomai / Te Pūkenga

Corresponding Author: Jonathan Adams, Jonathan.Adams@toiohomai.ac.nz, 07 5578969

Abstract

The increasing availability and testing of Artificial Intelligence in Education (AIED) is highlighting the concurrent gap and demand for ethical design and use. This paper proposes the design thinking framework for use in AI design. Design thinking inverts the current AI development process which builds the AI application first, then looks to apply this to human problems. In contrast, the human-centred focus of design thinking in AI development places empathy and agency with users and marginalised or affected parties at the heart of the design process.

Design thinking shifts the dominant discourse from the technological merits of AI development to the merits of the AI design for the needs and interests of ākonga (students) and kaiako (teachers), as defined by them. It ensures that AI tools are not just those that are feasible but desirable from end-users' perspectives.

By applying design thinking principles, the AI applications are intrinsically aligned to ākonga needs. We consider design thinking to be grounded in consideration of human-centric ethical and cultural influences that shape educational technology uptake in Aotearoa New Zealand.

Introduction

The rapid increase in the availability of AI applications has heightened the urgency of addressing ethical considerations associated with existing technology. Most recent and contemporary commercial AI development – as with other technologies - is technologically-driven. By this, we mean that commercial applications are developed based on the availability of technology. Advancing technology is the primary focus. Ethical questions are then identified when the technology is developed and ready for uptake or when it is being utilised. In contrast, design thinking, which emphasises deeply understanding user requirements, provides a framework for ethical development and use of AIED.

Design thinking inverts the current AI development process which builds the AI application first, then looks to apply this to human problems. In contrast, design thinking in AI development places empathy for and agency with users and marginalised or affected parties at the heart of the design process. Without this empathy and agency, there are potential risks associated with AI development. In the next section, we identify some ethical problems in commercial AI using human categorisation. This highlights the need for effective governance of AI. We briefly note some attempts to govern AI in Aotearoa New Zealand which to date remain underdeveloped. We then turn to design thinking which emphasises deeply understanding user requirements throughout the design and development process. aligns with developments in public policy that highlight the importance We discuss how design thinking principles of engagement, collaboration and co-design with ākonga/learners and kaiako/teachers ensure human rights including indigenous rights are at the centre of the design of AIED tools and applications. We conclude by noting that design thinking has enhanced outcomes in many spheres including health care, urban and infrastructure planning as well as conventional retail product development. While there may be challenges in implementing design thinking lessons can be drawn from those other spheres.

AI commercialisation

Commercial software products are typically licensed with an ‘as-is’ disclaimer leaving it to the customer to fit and use the software for their own use and context. For business and

government customers, including the education sector, these licenses are accompanied by variable fees for consulting, integration and services to use or customise the software. With rapid advances in AI (especially large language models), this approach in the acquisition and use of AI technology is inherently flawed and dangerous. It is based on a procurement model that assumes transparency of the application, that we know what the software can do with the data we input. It is a software licensing model that is suited for relatively simple business productivity applications with low risk. AI software products can be considered in two broad categories:

1. Standalone applications that customers connect to with their own customer data. This results in outputs that are already codified by a stand-alone application, a black box for the user;

and

2. AI configuration and training tools that utilise existing proprietary and open-source foundational AI models, hosted in cloud platforms such as Amazon Web Services (AWS), Azure (Microsoft), Google and Alibaba Cloud. This option allows customers to use machine learning to develop their own applications or use their own training data without any customer or end-user engagement. Within this second category are 2.1 the open-source models, available on GitHub and cloud providers, and 2.2 branded models, often built off the same training data sets as open-source (GitHub, n.d.).

In the context of AIED, and focusing on commercial branded providers, we see even within this well-resourced and documented offering a design process that preferences technological not human centric problems. Critically, it leaves the ethics assessment to developers, with no challenge to the implicit bias this brings in the absence of a user-centric engagement.

Challenges for AIED

The significant benefits using AI in human health, disease identification and environmental science show us why we should embrace technology such as machine learning or vision AI when it is directed with and for us. The accessibility opportunities for vision impaired

users provides a compelling ethical use for features such as user or built-environment identification.

Human identity and identification are among the most sensitive capabilities already recalibrating ethical boundaries, reflected in changes to previously available AI features. For example, face and identity attributes such as gender and age that were previously able to be used in Microsoft AI products were retired on 30th June 2023 (Bird, 2022) but gender labelling is still a feature for Microsoft direct customers and access determined by the technology vendor in their internal process. In New Zealand there is legal protection from discrimination within Section 21(1) of the HRA, including sex, race, age and employment status, which would include AI-supported decision-making by government agencies. These are all factors that have previously led to controversy in relation to algorithmic decisions, notably in the use of human attributes and profiling in corrections and policing (Gavaghan et al., 2019). A vendor-only decision on deploying AI features risks a subjective world view being used to grant access to these capabilities and subsequent applications. As highlighted by West et al. (2019, p. 8) “the functional logics of a given technology echo the gender and racial dynamics of the industry that produced it”.

By focussing on education value and society’s needs we can address inequities in existing business and public service operations, before it is codified and perpetuated in the digital world. An example of this conscious implementation is the Equity Adjustor Score, a change to including ethnicity as a criterion for elective surgery to offset health outcomes disparity in New Zealand Māori and Pacific populations (Arora & O'Callaghan, 2023; Royal Australasian College of Surgeons, 2023).

To understand these human dimensions Renz and Vladova (2021, p. 8) proposed that AIED use a Human-centred AI (HCAI) design, describing it as a “design thinking approach that puts humans at the centre of AI development, rather than considering AI automation as a replacement for human agency and control”. Our argument is to extend this further, where the AI developer functions are subsumed to that of order-takers. In an AIED context the wishes of ākonga and their whānau/family, and kaiako are included in a human-centred design framework that meets ethical guidelines codified as standards. This is how risk for other technologies that can cause harm, such as transport certification and marine surveys for shipping, is managed.

Governance of AI

Having a strong governance framework for AI in Aotearoa might provide scope to ensure that users are involved in development of AI technologies. To date, however, there have been only limited and patchy efforts in this regard although, as we are argue, there are helpful principles in some existing public policy and also some community and industry-led initiatives.

The Algorithm Charter for Aotearoa New Zealand was published in July 2020 as an initiative to foster public confidence in how government agencies use algorithm data (New Zealand Government, 2020). Numerous government agencies are signatories to the Charter. The Charter contains six core commitments:

1. Transparency - explaining how decisions are informed by algorithms.
2. Partnership – respecting government's Treaty partner through embedding a Te Ao Māori perspective in the development and use of algorithms consistent with the principles of the Treaty of Waitangi.
3. People - identifying and actively engaging with people, communities and groups who have an interest in algorithms, and consulting with those impacted by their use.
4. Data – in particular, understanding its limitations and identifying and managing bias.
5. Privacy, ethics, and human rights – safeguarding these by regularly peer reviewing algorithms to assess for unintended consequences and act on this information.
6. Human oversight – ensuring that there is a point of contact for public inquiries about algorithms, providing a channel for challenging or appealing of decisions informed by algorithms and clearly explaining the role of humans in decisions informed by algorithms (New Zealand Government, 2020).

The Algorithm Charter aligns closely with the private sector's AI Forum's AI principles also published in a 2020 document called "Trustworthy AI in Aotearoa New Zealand". The principles were developed to build public trust with AI use and development, providing guidance for "anyone involved in designing, developing and using artificial intelligence in

New Zealand (AI stakeholders), with the goal of ensuring New Zealanders have access to trustworthy AI" (AI Forum New Zealand, 2020, p. 1) with five principles all of which have a strong focus on people.

While the Algorithm Charter appears to offer a promising framework for guiding AI, its delivery and impact has not matched the aspirational language. According to Courtney (2021, p. 30),

> Despite the triumphant claims from actors involved that the Algorithm Charter is a 'worldfirst' instrument in governing the public sector, the Charter falls significantly short of any legitimate attempt at governance of automated decision-making. Recognition of the Charter and its principles is voluntary for public sector agencies, leading to absences from those who have authorising provisions in their deeming legislation, or from those whose absences appear to be unexplained. Compliance is unenforceable and largely untraceable, as the Charter has no force on its signatories as a legally binding instrument.

The 2023 Digital Technologies Industry Transformation Plan describes a vision where "The world looks to Aotearoa New Zealand as a leader in ethical, innovative, inclusive and sustainable digital technologies" (New Zealand Government, 2023, p.7). As with this Algorithm Charter this appears to be highly aspirational and far from being attained.

Internationally, there are very different approaches to govern AI, which makes it very unlikely that an international co-ordination will emerge (Hutson, 2023). The EU has taken a 'regulation by risk' approach but this is very different from, say, the United States and China.

UK-based NESTA (the National Endowment for Science, Technology and the Arts) has toolkits for EU scenarios and explains why it is important for citizens and users at the centre of the design. This is supported by the EU legislation for AI that starts with the Amendment 15 statement that "As a pre-requisite, artificial intelligence should be a human-centric technology" European Parliament (2023, n.p.).

Within the CANZUS (Canada, Australia, New Zealand, and the United States) countries colonised by Anglo-centric cultures (Oliver et al., 2024) data sovereignty advocacy by

indigenous cultures now facing digital colonialism (Taiuru, 2020) is strengthening. This is more than avoiding negative cultural impacts of data use but asserting data ownership and control. A recent leading example of this is the Māori Data Lens created collaboratively with AWS and Māori data and technology advisers (Amazon Web Services, 2024).

In western societies, public participation including user involvement in design of policy and services has long been regarded as a feature of democratic society and associated with the rights of citizens to influence the polity to which they belong (Cheyne et al., 2008). The focus on users reflected a view that the public sector needed to be more consumer or customer focused (Boston et al., 1991). Public sector reforms in New Zealand in the late 1980s included a new focus on making central government organisations more responsive to citizens. Similarly, local government reforms in the late 1980s brought in significant new statutory requirements for consultation intended to ensure local government was more responsive and accountable to communities (Cheyne, 2015). Involving users is therefore a widely accepted feature of democratic societies and one that is particularly important in the era of AI.

Existing legislation provides for co-design in line with Item 3 of the NZ Algorithm Charter:

> Focus on people by identifying and actively engaging with people, communities and groups who have an interest in algorithms, and consulting with those impacted by their use (New Zealand Government, 2020).

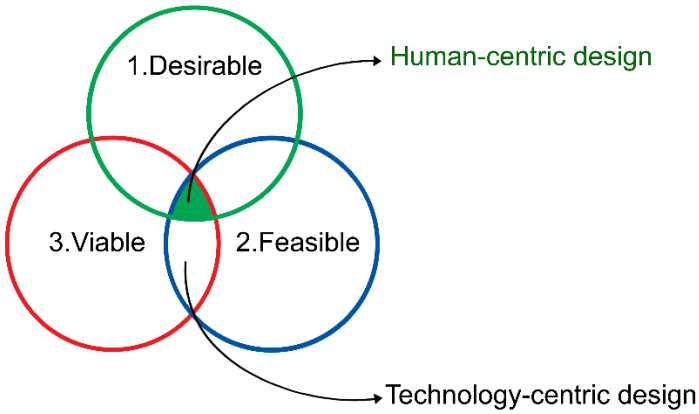
Design thinking principles are arguably implicit in the requirements in the Privacy Act 2020, Human Rights Act 1993 (HRA) and Health and Safety at Work Act 2015, where organisations must design their software or services in a way that protects or does not discriminate against any group of users and minimises the risk of harm. An organisation can only do this by understanding users' and other stakeholders' needs and expectations.

The New Zealand AI forum noted there are many projects with Māori tikanga and mātauranga Māori that are clear indicators of user-centric governance (AI Forum New Zealand, 2020). The Crown's responsibility to give effect to Te Tiriti o Waitangi in the Education and Training Act 2020 would require tertiary education institutions to implement ethical use of AIED through consultation (*Education and Training Act 2020*, 2020).

Insights from design thinking

As we noted in the Introduction, design thinking inverts the current AI development process which builds the AI application first, then looks to apply this to human problems Conventional AI development is typically focussed on what is possible, referred to in design thinking as the ‘Feasible,’ in conjunction with the commercial imperative of ‘Viability,’ to discover what can become a sustainable business model. In contrast, design thinking starts with ‘Desirability’ by directly asking the question; will people find this of value? (Figure 1) This is the starting point of a design thinking process.

Figure 1
How does Design Thinking Work?



Source: Adapted from (IDEO U, n.d.)

The design thinking process occurring at the nexus of Desirability, Feasibility and Viability contrasts with the technology race developing and releasing AI applications and models illustrated in the technology-centric overlap in Figure 1, where speed to market is preferred over arguably quality or ethics. The commercialisation of research by technologists is the reverse of human-centric design and may even exclude or ignore what is desirable in order to achieve commercial leadership with their technology. The absence of a genuine human-centric approach to foundational AI models was highlighted by Ehad Mostaque, founder of AI application provider Stability.ai in an interview in January 2023: “There aren't many people that train these models. We don't invite the whole community and you have 100 people training a model. It's usually five to ten, plus a supercomputer and a data

team, and things like that" (Estaque, cited in Guo, 2023, p. n.p.). This small cohort of model trainers contrasts with the high-volume dataset training by millions of casual minimum-wage labour workers, echoing the exploitation of resources of earlier industrial revolutions (Adams & Riddle, 2022).

A key aspect of human-centred design is authentic empathy and understanding of users, their cultural values and needs informed by ākonga and kaiako themselves, not assumed by the developers of AIED. It is unlikely to be possible to remove historical bias entirely from data for applications that use human attributes as historical datasets are a product of their time and place. There is still a gap in how to "better operationalize the various values that arise during the development of AI systems, rather than only applying rules and guidelines after AI deployment" (Renz & Vladova, 2021, p. 6).

Focusing on the right problem is more likely to lead to the right solution(s) during ideation and prototyping phases. Those are the solutions that prioritise stakeholder needs over commercial needs. Further, in the software development process itself, design thinking has shown it can improve the productivity and quality of AI design (Chang & Tsai, 2021). Successful design teams experience a knowledge convergence (Dong, 2005) in a process of creating a shared mental model or design requirements within a team that then feeds into the evolving design (Du et al., 2012). The benefits are explained by developers who created an algorithm for fracture detection that "[by] prioritizing the needs and preferences of end-users, design thinking can lead to the creation of more effective and user-friendly tools" (Ouyang et al., 2023, p. 735). The most logical way to approach an AI design, both its development and in the way it works is to start with the output, the human-centred need that is to be created (Dorst, 2011).

In a survey of AIED researchers about ethical considerations, Holmes et al. (2022) found that there was a heavy preoccupation with data issues (including ownership, privacy, security, control of interpretation) and also with representativeness and equity (including access to technology). In contrast, Holmes et al. (2022) emphasised the need for a broader consideration of ethics, noting that ethics of AIED should not be focused exclusively on being 'preventative' (that is, harm prevention) but 'protective' and 'facilitative' ensuring the benefits are experienced by all stakeholders. Interestingly, they

called for AIED to be approached with a ‘design science orientation’ (Holmes et al., 2022, p. 519) as this

> points to another practical outworking of ethical principles – namely, the use of human-centred design methods that give stakeholders genuine agency in shaping digital tools, thus increasing the chances of producing tools that are usable, effective, organisationally acceptable, and ethically sound.

Conclusion

Design thinking is well suited to complex problems such as AIED design, as it promotes a human-centric view, leading to a better design, more accurate identification of unmet needs and better outcomes, including application accuracy (Ouyang et al., 2023). Design thinking offers a framework to guide inform and guide the development of AIED that places ākonga, kaiako and other stakeholders at the centre of technology development. This shift of focus is likely to prompt AIED designers to confront their own assumptions and uncover inherent or unconscious bias.

Insights can be found other sectors and spheres (in particular, health care, housing, urban planning and architecture) where design thinking and human centric design has been incorporated in how to engage users and other stakeholders through use of effective communication tools and resourcing of user engagement. In their study of the use of design thinking to improve health care delivery in the intensive care unit, (Krolikowski et al., 2022) concluded that

> The tools of DT can support meaningful iteration and in-depth stakeholder engagement. Nevertheless, DT is not a standalone approach to solving complex healthcare problems. Interventions designed through DT need to be rigorously tested to provide evidence of both safety and effectiveness. The future of DT in healthcare is promising and, with thoughtful application and acknowledgement of limitations, DT has the potential to improve healthcare delivery.

In highlighting the potential for design thinking to enhance AIED development we also note that there are practical considerations, and possible pitfalls but these do not necessarily outweigh the potential gains.

References

Adams, J., & Riddle, K. (2022). AI Design Issues in Education. ITP National Research Symposium Proceedings. ITP National Research Symposium Proceedings., Invercargill. https://doi.org/10.34074/proc.2302012

AI Forum New Zealand. (2020). *Trustworthy AI in Aotearoa New Zealand. AI Principles*. AI Forum New Zealand. https://aiforum.org.nz/wp-content/uploads/2020/03/Trustworthy-AI-in-Aotearoa-March-2020.pdf

Amazon Web Services. (2024). *Māori Data Lens : AWS Well-Architected Framework*. https://docs.aws.amazon.com/pdfs/wellarchitected/latest/maori-data-lens/maori-data-lens.pdf

Arora, S., & O'Callaghan, J. (2023). *New ranking system prioritising patients on ethnicity to improve equity, experts say*. Stuff. Retrieved 19 June from https://www.stuff.co.nz/national/health/300908591/new-ranking-system-prioritising-patients-on-ethnicity-to-improve-equity-experts-say

Bird, S. (2022, 21 June). Responsible AI investments and safeguards for facial recognition. https://azure.microsoft.com/en-us/blog/responsible-ai-investments-and-safeguards-for-facial-recognition/

Boston, J., Martin, J., Pallot, J., & Walsh, P. (1991). *Public Management. The New Zealand Model*. Oxford University Press.

Chang, Y.-S., & Tsai, M.-C. (2021). Effects of design thinking on artificial intelligence learning and creativity. *Educational Studies*, 1-18. https://doi.org/10.1080/03055698.2021.1999213

Cheyne, C. (2015). Changing Urban Governance in New Zealand: Public Participation and Democratic Legitimacy in Local Authority Planning and Decision-Making 1989–2014. *Urban Policy and Research*, *33*(4), 416-432. https://doi.org/10.1080/08111146.2014.994740

Cheyne, C., O'Brien, M., & Belgrave, M. (2008). *Social Policy in Aotearoa New Zealand* (4 ed.). Oxford University Press.

Courtney, B. (2021). *Unchartered Waters: The insufficiencies of New Zealand's 'world-first' Algorithm Charter in governance of automated decision-making* [LLB (Honours), Victoria University of Wellington]. https://ir.wgtn.ac.nz/bitstream/handle/123456789/18163/paper_access.pdf?sequence=1

Dong, A. (2005). The latent semantic approach to studying design team communication. *Design Studies*, *26*(5), 445-461. https://doi.org/https://doi.org/10.1016/j.destud.2004.10.003

Dorst, K. (2011). The core of 'design thinking' and its application. *Design Studies*, *32*(6), 521-532. https://doi.org/https://doi.org/10.1016/j.destud.2011.07.006

Du, J., Jing, S., & Liu, J. (2012). Creating shared design thinking process for collaborative design. *Journal of Network and Computer Applications*, *35*(1), 111-120. https://doi.org/https://doi.org/10.1016/j.jnca.2011.02.014

Education and Training Act 2020. (2020). https://www.legislation.govt.nz/act/public/2020/0038/latest/LMS280244.html

European Parliament. (2023). *Artificial Intelligence Act*. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

Gavaghan, C., Knott, A., Maclaurin, J., Zerilli, J., & Liddicoat, J. (2019). *Government Use of Artificial Intelligence in New Zealand*. NZ Law Foundation. https://www.cs.otago.ac.nz/research/ai/AI-Law/NZLF%20report.pdf

GitHub. (n.d.). *Microsoft ResponsibleAIAccelerator Financial Services Documentation*. Retrieved 21 July 2023 from https://github.com/microsoft/ResponsibleAIAccelerator/blob/main/documentation/financialServicesExample.md

Guo, S. (2023). No Priors. Episode 103. Emad Mostaque. https://sarahguo.com/blog/emadmostaque

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I., & Koedinger, K. R. (2022). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, *32*, 504-526. https://doi.org/10.1007/s40593-021-00239-1

Hutson, M. (2023). Conflicting Visions for AI regulation. *Nature*, *620*, 260-263. https://doi.org/https://doi.org/10.1038/d41586-023-02491-y

IDEO U. (n.d.). *What is Design Thinking?* https://www.ideou.com/blogs/inspiration/what-is-design-thinking

Krolikowski, K. A., Bi, M., Baggott, C. M., Khorzad, R., Holl, J. L., & Kruser, J. M. (2022). Design thinking to improve healthcare delivery in the intensive care unit: Promise, pitfalls, and lessons learned. *Journal of critical care*, *69*, 153999.

New Zealand Government. (2020). *Algorithm charter for Aotearoa New Zealand*. data.govt.nz. https://www.data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

Oliver, G., Cranefield, J., Lilley, S., & Lewellen, M. J. (2024). Understanding data culture/s: Influences, activities, and initiatives: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology*, *75*(3), 201-214. https://doi.org/https://doi.org/10.1002/asi.24737

Ouyang, C.-H., Chen, C.-C., Tee, Y.-S., Lin, W.-C., Kuo, L.-W., Liao, C.-A., Cheng, C.-T., & Liao, C.-H. (2023). The Application of Design Thinking in Developing a Deep Learning Algorithm for Hip Fracture Detection. *Bioengineering*, *10*(6). https://doi.org/10.3390/bioengineering10060735

Renz, A., & Vladova, G. (2021). Reinvigorating the Discourse on Human-Centered Artificial Intelligence in Educational Technologies. *Technology Innovation Management Review*, *11*, 5-16. https://doi.org/10.22215/timreview/1438

Royal Australasian College of Surgeons. (2023). *College of Surgeons supports adding an equity lens to patient prioritisation*. Retrieved 19 June from https://www.surgeons.org/News/media-releases/College-of-Surgeons-supports-adding-an-equity-lens-to-patient-prioritisation

Taiuru, K. (2020). *Treaty of Waitangi/Te Tiriti and Māori Ethics Guidelines for: AI, Algorithms, Data and IOT.* http://www.taiuru.Maori.nz/TiritiEthicalGuide

West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating Systems: Gender, Race and Power in AI*. https://ainowinstitute.org/discriminatingsystems.pdf