

Submission in Response to the Parliamentary Inquiry into Digital Transformation of Workplaces

Our Organisation

Submission by the [Trustworthy Digital Society \(TDS\)](#). Questions related to this submission can be addressed to TDS@unsw.edu.au.

TDS is a UNSW-UTS collaboration bringing together voices from academia and industry to facilitate the creation of an equitable, inclusive, and sustainable, digitally enabled society.

The authors of this submission are all researchers and academics who have expertise across several digital-technology-oriented disciplines, including computer science, information science, information and communication technologies, human-computer interaction, enterprise architecture, business architecture, software engineering, big data, data processing, information and data access, technology standards, regulation, and governance, security and privacy, technology ethics, research training, research ethics, educational technologies, learning analytics, citizenship education, digital citizenship, data science, communication sociology, digital and social media, artificial intelligence, and natural-language processing. All authors are affiliated with the [UTS Centre for Research on Education in a Digital Society \(UTS: CREDS\)](#).

UTS: CREDS is a cross-faculty research group comprising researchers who explore the dynamic relationship between technology and learning – across formal, informal, and professional education contexts throughout the lifespan. UTS: CREDS research interrogates the new ways in which technologies enhance learning, and the changing learning needs of a digital society.

This submission represents the views of the authors, not the position of UTS, UNSW, or any of their individual units.

Overview of Submission

Our submission stresses the crucial role of the Australian Government in taking a proactive and multi-faceted approach to managing and regulating the impact of digital transformation of Australian workplaces.

Our recommendations suggest clear regulatory frameworks that promote transparency and accountability by developers, technology companies, private organisations, public sector institutions, and the Federal, state and territory governments in Australia. This can be achieved through the following steps:

- safeguarding data privacy;
- ensuring equity of access and fairness;
- promoting justice-oriented pedagogies in STEM education;
- safeguarding against harm to vulnerable populations;
- protecting intellectual property;
- respecting Indigenous data sovereignty and self-determination;
- supporting workforce training and transitions;
- fostering ethical practices in the digitalisation process;
- encouraging ethical innovation;
- enhancing worker rights;
- prioritising employee wellbeing; and,
- raising public awareness.

It is also essential for Australian entities to collaborate internationally toward ethical governance and equitable access to digital transformation and to ensure that existing and emerging global digital technologies are used responsibly and for the benefit of current and future employers and employees everywhere.

Our Recommendations

The authors thank the House Standing Committee on Employment, Education and Training for the opportunity to respond to the inquiry into Digital Transformations of Workplaces, particularly with respect to the rapid development and uptake of automated decision-making and machine learning techniques in the workplace.

Many of our recommendations address more than one Term of Reference (ToR), and the ToR reference number is appended in parentheses to the themed recommendations below.

1. Regulatory environment for AI, and public trust in, and understanding of, AI

Recommendation 1.1: AI regulation and frameworks can only build trustworthiness if they are clearly communicated and explained to the community, including communities who have historically been excluded or disadvantaged by transitions to new workplace technologies, i.e. older workers, migrant communities, Aboriginal and Torres Strait Islander communities, people with disability. Public communication campaigns and tailored training schemes for different industries are essential. *[ToR b, d, e, and f]*

Recommendation 1.2: Federal, state and territory governments to work with creative industry organisations and peak bodies in matters of intellectual property and use of creative outputs, to protect workers and individuals in the creative industries. *[ToR d, e, and f]*

Recommendation 1.3: The benefits of automated decision making and AI should not be evaluated solely on their ability to improve the efficiency of specific tasks, but rather consider them in relation to long-term implications and risks, rooted in human factors. Monitoring systems to foster responsible use, and enforce governance of emerging technologies, should be in place. Policies must be in place to monitor and detect cyber bullying, hate speeches, disinformation or misinformation patterns with a view to avoid any mental, physical and reputational damages to all involved parties such as employer, employee and customer. *[ToR c, d, e, and f]*

Recommendation 1.4: Research the feasibility, funding and establishment of a body responsible for the independent quality assurance of Regulatory Technology (Regtech). This body should have oversight from an appropriate regulatory body (for example, the FairWork Commission). *[ToR b]*

2. Consolidate and improve workplace regulation

Recommendation 2.1: There are several existing regulations in the Australian workplace relations system and more are emerging, in particular, in the related emerging technology areas of data and AI. We recommend that the workplace relations system be reviewed with a view to consolidate and simplify regulations before making automation or digitalisation decisions. Data collection is one of the most resource-intensive tasks for effective and efficient decision making. We recommend that regulators investigate and use relevant data automation software technologies to collect quality data for supporting evidence-based decision making with regards to the review and improvement of the workplace regulations, observability over their compliance, handling errors and non-compliance. *[ToR a and b]*

Recommendation 2.2: Create laws specifically addressing the use of Automated Decision Making (ADM) and Machine Learning (ML) in the workplace, covering areas such as data privacy, transparency, accountability, and fairness. Tailor regulations to different industries to address unique challenges and ensure relevant protections. This process should establish evaluation criteria for decisions that may be used in decision governance, supporting employees to understand their rights and obligations, and in monitoring and regulation (see recommendations 1.3 and 1.4). *[ToR b, c, and d]*

Recommendation 2.3: Data is core to decision making. Employers need to make sure data is fit for decision making, and risks of bias and error are understood and monitored. Further, it is important to ensure that appropriate resources and tools are in place for data quality, personal privacy and safety management. Employees must have the necessary training and skills to utilise data technologies for continuously keeping the data fit for the purpose of decision automation. *[ToR e]*

3. Support AI use in appropriate contexts while protecting vulnerable populations

Recommendation 3.1: Identify clear areas for the use of automated decision making (e.g. simple decisions, complex decisions, complicated decisions), while ensuring humans are in the loop and are accountable for the ultimate regulatory compliance decisions and outcomes. *[ToR b and c]*

Recommendation 3.2: Employers to use data-informed policies, processes and systems (e.g. compliance monitoring, tracking) and periodical reviews to improve their increasingly digital workplace, particularly in areas such as recruitment, onboarding, work allocation, pay, professional development, and career progression, where workers may be most negatively impacted (e.g. health, safety, discrimination, harassment) to understand the range of enablers, barriers, and strategies for quality work. *[ToR c, d, and e]*

Recommendation 3.3: Defining clear boundaries for use of AI systems in workplaces, in line with other jurisdictions (e.g. EU-style law); enforcing restrictions, or even bans on the use of high-risk AI technologies (i.e. facial recognition technologies) where it is likely that use may discriminate against or profile vulnerable population groups. *[ToR c, d, and f]*

Recommendation 3.4: Access to AI technologies can address inequality in the workplace, for example in the use of assistive technologies to overcome barriers to workforce participation for people with disability. However, workplaces need evidence-based policies to ensure better decision-making regarding ethical, fair, accountable and equitable use of AI. *[ToR c and f]*

Recommendation 3.5: Review and strengthen anti-discrimination and workplace laws to protect vulnerable workers for whom the training of AI models on biased data has been shown to increase discriminatory decisions and outcomes in the workplace. The use of AI technologies has enhanced systemic discrimination against employees and potential employees on the basis of race, gender, disability, sexuality, class and geography [as further discussed in 1.3 (T or f.)], and requires a strengthening of legislation to mitigate against these harms. *[ToR c]*

Recommendation 3.6: Provide support for SMEs to have fair and equitable access to AI technologies and solutions that will increase workplace efficiency and productivity. *[ToR d, e, and f]*

4. Indigenous Data Sovereignty and Self-Determination

Recommendation 4.1: Extend the Framework for the Governance of Indigenous Data (Australian Government, May, 2024), to all Australian workplaces (public, private and not-for-profit organisations). The framework, which currently applies to the APS, calls for improved governance, legislative and policy changes to ensure 'ownership and control over Indigenous data across all phases of the data lifecycle, including creation, collection, access, analysis, interpretation, management, dissemination and reuse'. The framework should be adapted to emerging AI technologies and implemented across the public and private sector to address Closing the Gap Priority Reform 4. *[ToR d and f]*

Recommendation 4.2: Federal, state and territory governments to prioritise consulting with Indigenous people and organisations with expertise in matters of information governance, ICIP, and cultural collections, to understand the concerns and actions required during digital transformation, especially in the cultural heritage industries. *[ToR d and f]*

5. Lifelong Learning

Recommendation 5.1: Regulators and employers to introduce professional development programs for employees about the safe and responsible use of Artificial Intelligence (AI), Generative AI (GAI) and Robotic Process Automation (RPA) technologies and embed awareness and use-case scenarios in their regulatory policy compliance software-based training. Learning opportunities around AI and RegTech should be supported for those: designing and building tools; implementing and using tools; regulating, evaluating and monitoring tools; and those whose work may be impacted or displaced by tools, particularly those from marginalised communities. *[ToR b, d, and e]*

Recommendation 5.2: Funding toward skills training programs and monitoring of impacts on communities and industries for workers experiencing job displacement as a result of automation and AI, with associated communications supporting the value of lifelong learning. *[ToR c, d and e]*

Recommendation 5.3: To address gender equity concerns associated with AI biases, we recommend educational programs to encourage the involvement of more women and girls in STEM careers, and rewarding proactive hiring practices in the AI industry. *[ToR f]*

Recommendation 5.4: Initiatives for STEM education should draw on justice-oriented pedagogies, particularly for research and development in the digital technology domain, highlighting the important intersections of *social* and *technical or scientific* concerns in learning about AI and its applications in society. *[ToR a, d, e and f]*

Context for our recommendations

ToR a) the benefits for productivity, skills development, career progression and job creation in Australia

Background

Organisations in Australia and globally have recognised the need to invest in digital capabilities and adopt new technologies to stay competitive in an increasingly digital landscape. Past developments have had some impact on productivity and jobs, coupled with economic growth and the need to adapt to a changing landscape, but, the rise of Generative AI (GenAI) brings unprecedented potential for transforming work in enterprises in the near future. GenAI will hence be one key area of focus along with other automated decision-making technologies for this ToR because of the human capital it may impact. 44% of working hours across industries could potentially be impacted by GenAI, with large sectors such as banking undergoing significant transformation (Ghosh et al., 2023).

While it is difficult to say if some jobs would benefit more from automation than others, there are certainly individual tasks performed by workers that are prone to automation. For instance, legal research – the process of scanning for past cases and laws can become quicker through AI supported information retrieval, saving many hours of manual labour for legal professionals (Takyar, 2024). Generative code writing and debugging can cut down programming hours for software developers by half (Kalliamvakou, 2022). Auto transcription services can speed up time-consuming logistical tasks across industries, such as generating transcripts for note taking and summarising key action items. Administrative burden can be reduced for teachers through streamlined processes, freeing time for student interaction.

Activities of transactional nature are more likely to be affected, where tasks requiring little human supervision can be delegated to automated decision making through machine learning and GenAI algorithms. We already see many organisations using chatbots for transactional activities such as helping book a ticket, and logging an IT issue, before re-directing users to a human agent for further help, if needed. This cuts down time for both the user who otherwise has to wait for a long period of time to get connected to a customer service representative, as well as for the service rep to direct the customer to readily available information for repetitive tasks. Collected data logs from past interactions can continue to improve the services to provide more accurate and helpful responses in the future.

Implications

Routine tasks can be handled more efficiently this way for low-stakes applications, allowing humans to focus on more complex issues requiring their intervention. However, there is no easy way to define what is considered low-stakes in an organisational context. For instance, the diminishing social interaction might negatively impact customers who prefer interacting with a human being instead of a chatbot. This is important for companies that require long-term relationships and trust to be built with customer accounts for high customer satisfaction and loyalty. The temporary spikes in productivity and profit might indeed not be desirable if they lead to diminishing returns in the long run. In addition, productivity gains observed at task/firm levels may not always transfer to the economy level because of adoption challenges and other complexities, commonly known as Solow's paradox (Capello et al, 2022). It is imperative to move beyond anecdotal evidence to gather data on the impact on jobs and the economy, particularly to understand possible negative consequences on workers.

Beyond transactional tasks, there is also scope for creative tasks to be semi-automated with humans in the loop. The “generative” part indicating the ability of AI to create outputs of a creative nature is a distinctive element of GenAI, which can help accelerate innovation across industries when used ethically and without infringing on existing copyright and intellectual property. It can democratise creative expression by empowering a broad audience with diverse backgrounds to participate in brainstorming, rapid prototyping, writing, and other creative endeavours. However, new literacies of AI and digital media, new professional code of ethics, new knowledge and expertise (e.g. prompt engineering for large language models), and critical thinking skills need to be developed. AI use can be more suited for certain tasks to produce higher-

quality results creating a ‘jagged technological frontier’ for knowledge workers (Dell’Acqua et al., 2023), and requires skilful navigation of this frontier to augment human capabilities. Ongoing training and professional learning about and with these tools is a central component to developing trust for adoption and positive use.

Organisations and educational institutions should provide opportunities for re-skilling and training the future workforce to be adaptable to changing needs. With the rapid pace of evolving technologies (E.g. Advanced voice and multimodal capabilities that did not exist in initial text-based GenAI models a year ago are now publicly available), skill development is *not* a one-shot challenge. Continuous learning is key to keeping up with emerging trends and best practices, which leads us to emphasise the need for lifelong learning skills to be cultivated. Tools such as TRACK can help individuals monitor current skills, mapping them to personalised career pathways (Kitto, 2020), and can be used beyond educational contexts by keeping track of job advertisements and skills gaps for different sectors and industries. Devising methods to materialise the concept of a ‘Skills Passport’ by overcoming challenges of data gathering and portability across systems is one step forward toward the direction of lifelong learning.

ToR b) the role of business software and regulatory technology ('Reg Tech') in improving regulatory compliance in the workplace relations system, including their use by regulators, and accountability for errors resulting in non-compliance;

Background

The Australian workplace relations system includes several laws, policies and processes that govern the employment relationships between the employee and employer. This includes Workplace Relations Act 1996, The Fair Work Act & Regulations 2009, Work Health and Safety Act & Regulations 2011. Furthermore, there are specific discrimination laws such as Age Discrimination Act 2004, Disability Discrimination Act 1992, Racial Discrimination Act 1975 and Sex Discrimination Act 1984. This system as a whole intends to foster fair, responsible & safe use of resources, organisations of work, pay and career progression without any discrimination.

Implications

There are three main types of parties in the workplace system: regulator, employer and employee. All these parties need to collaborate to ensure the improvement of compliance in a complex digital transformation involving several parties, laws, policies and processes. The improvement of compliance in this complex system is an arduous task, in particular, for the SMEs (e.g. approx. 95% of the Australian Economy). This situation is further complicated due to the current context of an increasingly digital, federated, hybrid and remote work environment that involves the use of data and digital technologies such as AI, Generative AI (GAI), Robotic Process Automation (RPA) etc. Traditional manual regulatory audit and compliance of such a complex work environment will not be effective and efficient, thus these parties can benefit from the use of contemporary business software and regulatory technology for automating the monitoring and decision making in improving the regulatory compliance ensuring fair, responsible & safe work environments.

ToR c) the risks, opportunities, and consequences for the nature of work, including effects on hiring, rostering, work intensity, job design, wage setting, monitoring, surveillance and job quality;

Background

Vial (2019), based on a review of 285 works on digital transformation, formulated a definition for Digital Transformation as “*a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing, communication, and connectivity technologies.*” and describes it as an evolution of IT-enabled transformation.

Some of the implications of this digital transformation include the automation of routine tasks toward improved efficiency, and the use of data-driven decision making and machine learning techniques within organisational processes. Work or employment is a major pathway that influences human flourishing, particularly its effect on an individual’s sense of meaning and purpose. Meaningful work can be defined as, “people’s subjective experience that their jobs, work, or careers are purposeful and significant, that their

work is harmoniously and energetically synergistic with the meaning and purpose in their broader lives, and that they are enabled and empowered to benefit the greater good through their work” (Steger, 2016, p. 60). Meaningful work is said to be created or maintained through several mechanisms, including authenticity, self-efficacy, self-esteem, purpose, belongingness, transcendence, and cultural and interpersonal sensemaking (Rosso et al., 2010). These ideas can be considered facilitators of meaningful work, which can provide the foundation for greater wellbeing overall.

Implications

Although automation does not in itself undermine the dignity of human labour, it transforms the value of human skills. Some skills are rapidly undervalued as a consequence of such digital transformation, and result in the need for active government interventions such as for welfare, (re)training, and other resourcing. Simultaneously, workplaces may be slower in recognising and rewarding skills which have become newly valuable. Automated decision making might be used to rapidly identify and readjust wages and redesign educational opportunities; however, that can also be a hurdle to human innovation and human flourishing.

Some of the following outcomes of digital transformation may have unintended and unforeseen consequences as follows, and require evaluation, and consideration:

- Enhanced productivity tools can lead to higher expectations and increased work intensity, potentially causing stress and burnout for employees.
- Remote work technologies can blur the lines between work and personal life, leading to longer working hours, reduced downtime, and poor work-life balance.
- Automation of routine tasks can lead to job displacement, particularly for positions perceived as low skilled.
- The mismatch between the skills workers possess and the skills needed for new, technology-driven roles can cause a skills gap that will have a flow on effect on education systems and also welfare systems.
- The gig economy and freelance platforms can lead to wage variability and insecurity, exacerbating income inequality.
- Increased employee monitoring can lead to more performance-based pay structures, which may not always be fair or transparent, and may disadvantage the most vulnerable populations.
- Continual digital monitoring technologies can infringe upon employee privacy and autonomy.
- Digital surveillance can lead to distrust between employees and management, harming workplace culture.
- Over-automation can make work feel dehumanised, reducing job satisfaction, engagement, and productivity.
- Continuous connectivity and heightened performance expectations can increase stress levels and contribute to mental ill-health and burnout.

ToR d) the effects of these techniques on the scope of managerial prerogative, labour rights, ability for workers to organise, procedural fairness, equality, discrimination, and dignity at work

Background

Regtech has the potential to have significant effects upon the scope of managerial prerogative, labour rights, the ability for workers to organise, procedural fairness, equality, discrimination and dignity at work. In some cases, these effects will be negative; there are also possibilities for more positive effects. The brief descriptions below will outline possible cases and extant research in the field; before presenting this discussion, it should be noted that this is still a relatively new area, and hence there is still limited research.

Increasingly, there is use of Regtech, especially in the form of algorithmic decision making (ADM) to automate processes like pay alterations (due to, for example, personal or carers’ leave or parental leave), entrance into countries (through the use of SmartGates, such as are in place in Australia), debt recovery (as in the Robodebt or more correctly, the Online Compliance Intervention system matter) and in criminal justice and sentencing (Ng & Gray, 2022).

Implications

Harms arising from data deficits, poor stakeholder engagement, and lack of training: While there are claimed productivity benefits from these techniques, they also raise the risks of wage theft and adverse financial and mental health effects. This has proven to be the case in the Robodebt scandal that occurred in Australia, where almost 500,000 people received incorrect or inflated debt notices due to the use of ADM. This had significant effects upon both the people who received the notices and the welfare staff (Rinta-Kahila et al., 2022). In some cases, it led to suicide (Graycar & Masters, 2022). According to Rinta-Kahila et al., (2022), this failure was largely due to the system operating with insufficient data, caused in part by relevant stakeholders being excluded from the design processes for the system. In addition, there was no proper testing or piloting, and crucially, no human oversight to check information (Glenn, 2017). In fact, many of the staff required to use the system had received no training or information about it before it was launched. Such a flawed implementation led to significant loss of trust on behalf of the users.

Exclusion of marginalised groups: The Robodebt example also serves to highlight another concern related to the use of Regtech techniques. Those individuals who are most likely to be interacting with the institutions of government (e.g. Centrelink, ATO, the NDIS) are often drawn from those groups who are likely to be experiencing some form of marginalisation: low socio-economic status, non-English speaking backgrounds, Indigenous Australians, people with disabilities and other vulnerable groups (Goggin & Soldatić, 2022). While new technological tools (such as ADM, for example) are often presented as liberatory forces for people in these groups (see Alper, 2017), there is only mixed evidence that this is the case (Lazar & Stein, 2017). In some cases, this has led to increased exclusion, as ADM often leads to 'digital-by-default' policies (Al-Muwil, 2019). In any case, there have been concerns raised about bias and discrimination of marginalised groups as a result of artificial intelligence more generally (Whittaker et al., 2019; Bennett & Keyes, 2020). This is likely to be a situation with ADM too, but the potential for harm is likely to be greater in governmental policy arenas.

Public trust in ADM: Concerns about the utility of Regtech and ADM are front-of-mind in questions about the trustworthiness of these techniques. According to Araujo et al. (2020), there are strongly held concerns about the fairness and utility of decisions made via these techniques - even though participants in that particular study rated the decisions made by ADM as good as or better than those made by human experts!

Data provenance in ADM: There are also concerns regarding the provenance of various models used in the formation of ADM. This is often described as the digital supply chain, and is not as transparent as it should be. This means that there is both the potential and likelihood for the exploitation of workers (see, for example, Hockenberry, 2021).

Approaches to address risks

A dignity lens in the design process: Application of a dignity lens (Felstead et al., 2023) in the design process of ADM may help to address these concerns. Applying a dignity lens to the design process places the protection of human dignity at the centre of any large scale design project.

Trust frameworks for ADM: More broadly speaking, the development and implementation of a trust framework is also an important consideration (Andrews et al., 2022). Digital tools offer great potential but in order to mitigate some of the challenges, agency of the users and consumers needs to be central.

Learning and trustworthy systems, a key connection: There is a connection between understanding of tools, techniques and their application - and trusting them to do the job they are designed to do; learning is at the heart of that connection. Procedural fairness, understanding of outputs provided by AI (however transparent or explainable), and ability to intervene in AI decisions, all rest on knowledge and skills, alongside provision of clear evidence of the purposes and validation for systems. Agency will be developed through transparency, training and education, which will serve to increase understanding and trust in these tools. There is thus a role for education and training in the inception, design, implementation, and evaluation of Regtech. This means that another recommendation that is central to this submission is the attention that needs to be paid to education and training in these different tools - both for the designers and users, but also those who are affected by Regtech - the 'consumers'.

ToR e) appropriate safeguards or regulatory interventions to guide responsible implementation in the workplace, including the digital skills and resources necessary for employers to appropriately utilise these technologies.

Background

A decision is a choice between two or more alternatives and the allocation of irrecoverable resources. Decision making is a complex undertaking and involves several elements such as the decision maker, a frame or viewpoint, alternatives to choose from, preferences, information and logic. For instance, a decision maker, such as an individual or group, may have a certain view or mindset such as considering to change the work mode, alternatives to choose from could be in-person, online or hybrid working modes, preference could be to first focus on permanent employees, whereas information could be in the form of their current work patterns, and finally logic could be a mechanism that is used to make the actual decision. There is an increasing interest, noise and hype around the adoption of AI, GAI and RPA in the workplace. For instance, decision making can be augmented or automated using AI, GAI, data, rules and heuristics for improving employee efficiency and productivity.

Implications

There is a growing interest in decision automation using the emerging technologies (e.g. AI, GAI). However, a safe and responsible use of such technologies must be assured via appropriate regulations, policies, procedures, standards, guiding principles, frameworks and training. As noted earlier, there are several existing and new emerging regulations, policies etc., and compliance to these may not make it easy for the employees and employers thinking to adopt decision automation technologies. Furthermore, it is also important to understand the decision hierarchy (e.g. policy, strategic, tactical and operational level decisions) and quality.

Challenges to consider include:

- How would you determine the quality of a decision?
- How would you separate a good decision from a bad decision?
- Which part(s) of the decision making can be safely automated?
- Who is accountable and responsible for machine intelligence or decisions?
- What is the quality of the data, which is being used as an input to the decision-making process?
- Do employees have appropriate skills or motivation to learn new technologies?

ToR f) the effects on gender equality, job security, small businesses, Closing the Gap and disadvantaged and vulnerable cohorts of workers

Background

While some research has asserted that AI is a neutral technology, capable of making better decisions than humans owing to the inevitable bias in human decision-making, this has been strongly contested by tech industry insiders, human rights campaigners, and academics, who have argued that, as AI is trained on large amounts of private and public data, it cannot help but reproduce biases relating to gender, race, religion, class, age, sexuality, geography, physical attributes and abilities that exist in the broader society (Noble, 2018; Crawford, 2021). This leads to biased outcomes that, given the adoption of these technologies at scale, may broaden social inequalities and make accountability of biased decision making more difficult.

Implications

Even before the risks of AI were being debated and regulations hastily formulated, media scholar, Saffiya Umoja Noble argued that the search algorithms and databases that assist us in our working and everyday lives were not “neutral” or objective but reflected the values of their makers, and the machines that they built for commercial use to sort information and populations into different forms of hierarchy, credibility, trust and humanity. The examples Noble highlights show that queries regarding Black, Latin and Asian women entered into Google’s search engine returned results pointing to pornographic and other demeaning materials, reflecting broader social patterns of sexism and racism. This, she argued, underlined a “corporate

logic of either wilful neglect or a profit imperative...” from platforms whose algorithms were reproducing these harmful stereotypes (Noble, 2018, p. 5). This has been joined by more recent examples, where queries seeking representations of different professions which were sent to AI image generation models, such as Open AI’s DALLE-2 and Stable Diffusion, returned results where low skilled workers were represented as women and people of colour, while “experts” and professionals were represented as white and male (Lamensch, 2023; Nicoletti & Bass, 2023). These reports have also been accompanied by warnings that biased AI text to image models are being rapidly adopted across a range of industries and business products (Nicoletti & Bass, 2023). Biased AI isn’t just fuelling harmful stereotypes, but also leading to unfair treatment and workplace discrimination. This has been illustrated through examples such as Amazon’s AI CV screening tool, introduced in 2018 to increase efficiencies in recruitment, which was trained on its own employee data of the previous 10 years. As the employment practices of the company had skewed towards the hire of male employees the tool was found to discriminate against women and had to be abandoned (Council of Europe, 2023, p. 10). This reflects growing concerns that societal biases toward women, people of colour and other groups historically discriminated against in terms of wage inequality, hiring and recruitment, and stereotypes about ability and capacity to undertake certain roles, will become exacerbated and further baked into workplace decision-making with increasing reliance on AI in the workforce (Smith & Rustagi, 2021).

These risks and measures to mitigate against them have been highlighted in Australia’s AI Ethics Principles and Framework (Department Industry, Science and Resources), which has as one of its foundational principles to ensure that: “throughout their lifecycle, AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups” (p.10). The recent EU AI Act serves as the first set of regulatory guardrails for AI, where the use of some forms of AI have been banned because of its potential to lead to biased and discriminatory outcomes in the areas of welfare provision, education, the Criminal Justice system, and of course workforce participation. And while these harms are found to impact on vulnerable workers most often, they also impact businesses, with research conducted for the World Economic Forum showing that 62% of organisations who participated in the research experienced lost revenue as a result of using biased AI models, and 61% citing lost customers, 43% lost employees, 35% experiencing lawsuits and other damages and 6% citing reputational and brand damage (Datarobot, 2022).

In the following examples we consider existing and future opportunities for inclusion of vulnerable workers owing to adoption of AI technologies and systems in workplaces, as well as risks of AI bias and discrimination relating to vulnerable populations.

Gender equality: As businesses increase their reliance on AI to decide on hiring and recruitment, wage setting, promotion and other decisions that determine job and career advancement opportunities, there have been concerns that already existing gender inequalities in workplaces will worsen. In a study conducted by Berkeley Haas Center for Equity, Gender and Leadership on mitigating bias in artificial intelligence, researchers detected “unfair allocation of resources, information, and opportunities for women manifested in 61.5 percent of the systems identified as gender-biased, including hiring software and ad systems that deprioritized women’s applications” (Smith & Rustagi, 2021). One of the main problems identified (UN Women, 2024) is that the data AI models are trained on are themselves biased, and that a lack of more diverse datasets and gender inclusive teams conceptualising and developing AI, is further compounding this problem. To address the challenge, studies have recommended the involvement of more women and girls in STEM careers, and rewarding proactive hiring practices in the AI industry. It is also recognised that that the current “governance deficit” in relation to bias and discrimination generated by AI systems needs to be addressed with the development of multi stakeholder governance models that prevent the use of AI systems exhibiting gender bias (UN Women, 2024), or to set and enforce “accuracy standards” (European Network of National Human Rights Institutions, 2023).

Aboriginal and Torres Strait Islander communities: Emerging technologies and data-driven processes of governance have had damaging repercussions for Aboriginal and Torres Strait Islander communities, with the use of census data and other available data often leading governments to create decisions and laws for

Indigenous Australians without their consent or consultation. The results have had negative impacts on Aboriginal rights to self-determination and agency over decisions that impact community, while leading to discriminatory policies and widening inequalities in the experiences of Indigenous and non-Indigenous Australians. The latter has paved the way to forms of redress such as the Closing the Gap agreement (2020). Alongside this, leaders in the Indigenous Data Sovereignty movement have addressed emerging AI technologies and worked with government to design a Framework for the Governance of Indigenous Data (2024). The framework calls for: i) capability building to ensure Aboriginal people and communities can benefit from AI technologies, and practice self-determination in decision-making affecting communities (p. 6), and ii) legislative and policy changes to ensure ‘ownership and control over Indigenous data across all phases of the data lifecycle, including creation, collection, access, analysis, interpretation, management, dissemination and reuse’ (p. 6-7).

Groups who have been unfairly discriminated against on the basis of age, disability, sex, pregnancy and marital status, sexual orientation, gender identity and intersex status, as well as groups who may be vulnerable to uptake of AI systems in future workplaces: In the Australian Human Rights Commission’s Final Report on Human Rights and Technology (2021) advancements in AI are identified as bringing opportunities for innovation, inclusivity and accessibility, with a section of the report dedicated to outlining the opportunity AI provides for people with a disability, for whom text to speech models, captioning technologies and smart assistants are viewed as facilitating inclusion and rights to work (p. 14). On the other hand, it highlights evidence that AI can be used in ways that may increase risk of harm to populations who are already subject to discrimination and unfair treatment in the workplace, as outlined in the Introduction to this section. Addressing these opportunities and risks the report has made a number of recommendations, including formalising Accessible AI as an enabling right, similar to the right to education, and setting up provisions within workplaces and through the NDIS scheme to make accessible technologies more widely accessible, law reform and the establishment of an AI Safety Commissioner to provide oversight with regard to bias in AI systems.

Workers who experience job displacement: One of the greatest expected impacts of AI on human rights and inequality is the expected job displacements driven by AI, GAI and RPA technologies. As these job displacements are expected to impact low skilled workers and workers in specific industries, it is expected that some groups who already experience precarity and income inequality (migrants, women, low socio-economic groups) as well as new and emerging vulnerable groups will experience a growing income gap and reduced opportunities for social mobility (Marr, 2023; see also OECD, 2024, pp. 4-6). On the other hand, it can also reduce humans being employed to undertake dangerous jobs and jobs that infringe human dignity.

Small businesses: By concentrating ownership of AI technology among a small number of corporations and firms, inequality in terms of affordability and access of AI systems to improve workplace productivity, innovation and efficiency is likely to result, with SMEs finding themselves unable to compete. To address this, it is likely that policies that promote open access or inclusive AI, or government subsidy schemes will be needed to even the playing field and help halt another emerging technology divide (Marr, 2023).

Authors (in alphabetical order)

- Professor Asif Gill, Head of Discipline (Software Engineering), School of Computer Science, UTS
- Dr. Keith Heggart, Director, Centre for Research on Education in a Digital Society, UTS
- Associate Professor Amelia Johns, Digital and Social Media Discipline, School of Communication, UTS
- Associate Professor Simon Knight, Transdisciplinary School, UTS
- Associate Professor Bhuvan Narayan, Director - Graduate Research, Faculty of Arts and Social Sciences, UTS
- Dr. Antonette Shibani, Transdisciplinary School, UTS

References

- Al-Muwil A, Weerakkody V, El-haddadeh R, et al. (2019) Balancing digital-by-default with inclusion: a study of the factors influencing e-inclusion in the UK. *Information Systems Frontiers* 21(3): 635–659.
- Alper M (2017) *Giving Voice: Mobile Communication, Disability, and Inequality*. Cambridge, MA: MIT Press.
- Andrews, P., de Sousa, T., Haefele, B., Beard, M., Wigan, M., Palia, A., ... & Jacquet, A. (2022). *A Trust Framework for Government Use of Artificial Intelligence and Automated Decision Making*. Preprint <https://arxiv.org/abs/2208.10087>
- Araujo, T., Helberger, N., Kruijkemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, 35, 611-623.
- Australian Government. *The Australian Government's interim response to safe and responsible AI consultation*. <https://www.industry.gov.au/news/australian-governments-interim-response-safe-and-responsible-ai-consultation>
- Australian Government (2024) *Framework for the Governance of Indigenous Data* <https://www.niaa.gov.au/sites/default/files/documents/2024-05/framework-governance-indigenous-data.pdf>
- Australian Human Rights Commission (2021) *Final Report: Human Rights and Technology* <https://humanrights.gov.au/our-work/technology-and-human-rights/publications/final-report-human-rights-and-technology>
- Bennett CL, Keyes O (2020) What is the point of fairness? Disability, AI and the complexity of justice. *ACM SIGACCESS Accessibility and Computing* 125: Article 5. <https://dl.acm.org/doi/10.1145/3386296.3386301>
- Capello, R., Lenzi, C., & Perucca, G. (2022). The modern Solow paradox. In search for explanations. *Structural Change and Economic Dynamics*, 63, 166-180. DOI: 10.1016/j.strueco.2022.09.013
- Closing the Gap Partnership (2020) *National Agreement on Closing the Gap* <https://www.closingthegap.gov.au/sites/default/files/files/national-agreement-ctg.pdf>
- Council of Europe (2023) Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination. <https://edoc.coe.int/en/artificial-intelligence/11649-study-on-the-impact-of-artificial-intelligence-systems-their-potential-for-promoting-equality-including-gender-equality-and-the-risks-they-may-cause-in-relation-to-non-discrimination.html>
- Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (1st ed.). Yale University Press. <https://doi.org/10.2307/j.ctv1ghv45t>
- Datarobot (2022) *DataRobot's State of AI Bias Report Reveals 81% of Technology Leaders Want Government Regulation of AI Bias* <https://www.datarobot.com/newsroom/press/datarobots-state-of-ai-bias-report-reveals-81-of-technology-leaders-want-government-regulation-of-ai-bias/>

- Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (24-013).
- Department of Industry, Science and Resources (n.d) Australia's AI Ethics Principles
<https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>
- EU Artificial Intelligence Act (2024) *The EU Artificial Intelligence Act* <https://artificialintelligenceact.eu/>
- European Network of National Human Rights Institutions (2023) *The EU Artificial Intelligence Act: Up-to-date developments and analyses of the EU AI Act* <https://artificialintelligenceact.eu/>
- Felstead, C., Stockdale, R., & Scheepers, H. (2023). A dignity perspective on the potential harm of AI technologies: The case of Robodebt. *ACIS 2023 Proceedings*. 100.
<https://aisel.aisnet.org/acis2023/100>
- Ghosh, B., Wilson, H. James., & Castagnino, T. (2023) GenAI Will Change How We Design Jobs. Here's How. *Harvard Business Review*. <https://hbr.org/2023/12/genai-will-change-how-we-design-jobs-heres-how>.
- Glenn R (2017) *Centrelink's automated debt raising and recovery system. A Report about the Department of Human Services' Online Compliance Intervention System for Debt Raising and Recovery*. Report by the Acting Commonwealth Ombudsman, Richard Glenn, under the Ombudsman Act 1976, Report No. 02/2017. https://www.ombudsman.gov.au/data/assets/pdf_file/0022/43528/Report-Centrelinksautomated-debt-raising-and-recovery-system-April-2017.pdf
- Goggin, G., & Soldatić, K. (2022). Automated decision-making, digital inclusion and intersectional disabilities. *New Media & Society*, 24(2), 384-400.
- Graycar, A., & Masters, A. B. (2022). Bureaucratic bastardry: robodebt/debt recovery, AI and the stigmatisation of citizens by machines and systems. *International Journal of Public Policy*, 16(5-6), 333-344.
- Hockenberry, M. (2021). Redirected entanglements in the digital supply chain. *Cultural studies*, 35(4-5), 641-662.
- Kalliamvakou, E. (2022). Research: quantifying GitHub Copilot's impact on developer productivity and happiness. *The GitHub Blog*. <https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/>. Retrieved 5 Jun 2024
- Kitto, K. (2020) Using data to help students get on TRACK to success.
<https://lx.uts.edu.au/blog/2020/02/24/using-data-to-help-students-get-on-track-to-success/>
- Kitto, K. (2024) Will a Skills Passport ever get me through the lifelong learning border? : Two critical challenges facing personalised user models for lifelong learning. In Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '24), July 01–04, 2024, Cagliari, Italy. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3627043.3659564>
- Lamensch, M. (2023, June 14). *Generative AI tools are perpetuating harmful gender stereotypes*. Centre for International Governance Innovation. <https://www.cigionline.org/articles/generative-ai-tools-are-perpetuating-harmful-gender-stereotypes/>
- Lazar J, Stein MA (eds) (2017) *Disability, Human Rights, and Information Technology*. Philadelphia, PA: University of Pennsylvania Press.
- Marr, D. (2023) The 15 Biggest Risks of Artificial Intelligence. *Forbes*.
<https://www.forbes.com/sites/bernardmarr/2023/06/02/the-15-biggest-risks-of-artificial-intelligence/?sh=2092baef2706>

- Ng, Y. F., & Gray, S. (2022). Disadvantage and the automated decision. *Adelaide Law Review*, 43, 64
- Nicoletti, L. & Bass, D. (2023) *Bloomberg - Are you a robot?* <https://www.bloomberg.com/graphics/2023-generative-ai-bias/>
- Noble, S. U. (2018). *Algorithms of oppression: how search engines reinforce racism*. New York University Press. <https://doi.org/10.18574/9781479833641>
- OECD (2024) Using AI in the workplace: Opportunities, risks and policy responses, https://www.oecd-ilibrary.org/science-and-technology/using-ai-in-the-workplace_73d417f9-en;jsessionid=aq07ekapga1lh1NrWvPATs2KIQNnblk0c-8hcM40.ip-10-240-5-60
- Rinta-Kahila, T., Someh, I., Gillespie, N., Indulska, M., & Gregor, S. (2022). Algorithmic decision-making and system destructiveness: A case of automatic debt recovery. *European Journal of Information Systems*, 31(3), 313-338.
- Rosso, B. D., Dekas, K. H., & Wrzesniewski, A. (2010). On the meaning of work: A theoretical integration and review. *Research in Organizational Behavior*, 30, 91–127. <https://doi.org/10.1016/j.riob.2010.09.001>
- Smith, G., and Rustagi, I. (2021) *When Good Algorithms go Sexist: Why and how to advance AI Gender Equity (SSIR)*. (C) 2005-2024. https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity
- Steger, M. F. (2016). *Creating Meaning and Purpose at Work*. In *The Wiley Blackwell Handbook of the Psychology of Positivity and Strengths-Based Approaches at Work* (pp. 60–81). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118977620.ch5>
- Takyar, A. (2024) AI for legal research: Applications, architecture, benefits, tools and development. <https://www.leewayhertz.com/ai-for-legal-research>
- Think Digital. 2024. *Generative AI in the Public Sector: Review of Existing Employee Guidelines. Environmental Scan & Analysis*.
- UN Women (2024) *Artificial Intelligence and Gender Equality* <https://www.unwomen.org/en/news-stories/explainer/2024/05/artificial-intelligence-and-gender-equality>
- Vakil, S., & Higgs, J. (2019). It's about power. *Communications of the ACM*. 62(3), 31–33.
- Vial, G. (2021). Understanding digital transformation: A review and a research agenda. *Managing digital transformation*, 13-66.
- Whittaker M, Alper M, Bennett CL, et al. (2019) *Disability, Bias, and AI*. New York: AI Now Institute at NYU. Available at: <https://ainowinstitute.org/disabilitybiasai-2019.html>