

ARTICLES**GENDERED HARMS AND THE REGULATION OF
ARTIFICIAL INTELLIGENCE:
A COMPARATIVE ASSESSMENT OF EMERGING
LEGISLATIVE PRACTICE***Ramona Vijeyarasa*

ABSTRACT

The risk that Artificial Intelligence (AI) will magnify existing gendered harms and create new ones is relatively established among academic circles. Scholars highlight how AI replicates gender biases in the results of search engines or in the use of AI-driven technologies in employment or banking-related decisions when such technologies are designed, deployed, and used without due attention to gendered impacts. Yet, a question remains as to whether these gender perspectives are being incorporated into the AI-related laws emerging globally. At the time of writing, the race to regulate AI is intensifying, but too few initiatives pay attention to the gender-related challenges generated by AI systems. The vast majority of proposed or actual laws fail to adequately address gendered harms, if at all. In this article, I offer emerging global good practices to translate this gendered knowledge into legislation and seek to understand how an intersectional gender lens can be incorporated into domestic law. In Part II, I set out what is AI, what are its gendered implications and how do AI technologies replicate existing societal gender biases. I discuss the allocative harms of AI and the representative harms and elaborate upon an emerging but largely under-acknowledged harm, equality gaps in AI literacy. In Part III, I turn to the question of regulating AI with gender in mind. I seek to arrive at a better understanding of how

non-discrimination, equality, and bias can be incorporated into the laws governing AI. A comparative multi-jurisdictional study, I draw upon the legislative debates unfolding in the US, Japan, China, and Australia before turning to the more promising examples emerging from the EU, Canada, and Brazil. I conclude by considering how we might regulate better to achieve algorithmic fairness for a greater diversity of women.

ABSTRACT	114
I. INTRODUCTION	116
II. CONTEXT: UNDERSTANDING THE GENDERED IMPLICATIONS OF ARTIFICIAL INTELLIGENCE	122
A. <i>What is AI and What Makes it Intelligent?</i>	122
B. <i>An Overview: Scholarly Debates and AI’s Gendered Harms</i>	125
i. Allocative Harms.....	128
ii. Representative Harms	132
iii. Knowledge-Based Inequality	135
III. THE GLOBAL LANDSCAPE	137
A. <i>Method and Approach: Existing and Emerging AI Legislation</i>	137
B. <i>Weak Treatment of Gender-Based Harms</i>	139
i. The United States of America	139
ii. Japan	141
iii. China	142
iv. Australia	144
C. <i>Emerging Promising Practice</i>	146
i. The European Union.....	146
ii. Canada.....	151
iii. Brazil	154
D. <i>Comparative Lessons</i>	155
IV. CONCLUSION: FINDING THE WAY FORWARD IN THE GLOBAL LEGISLATIVE LANDSCAPE	157

GENDERED HARMS AND THE REGULATION OF ARTIFICIAL INTELLIGENCE: A COMPARATIVE ASSESSMENT OF EMERGING LEGISLATIVE PRACTICE

RAMONA VIJAYARASA*

I. INTRODUCTION

In 2018, Bradesco, one of Brazil’s largest banks, launched *Bradesco Intelligencia Artificial*, an Artificial Intelligence (AI)-powered chat bot that provides online help to the bank’s clients.¹ As with previous chatbots personified as women, BIA – also a common female Brazilian name – started to become the target of harassment. In 2020, Bradesco registered 95,000 morally or sexually offensive messages to BIA, including the use of explicit language about violence against women.² In the midst of both of these events, the United Nations Educational, Scientific and Cultural Organization (UNESCO) stressed that the harassment directed at virtual personal assistants, a highly feminized form of AI, was normalizing the verbal abuse women suffer in everyday life.³ In 2021, the concerns of both institutions converged with Bradesco and UNESCO launching “BIA against harassment.” The campaign helped change how the chat-bot responds to users when their language is “inappropriate”, asking them to “[p]lease, change the way you talk.”⁴

* Dr. Ramona Vijayarasa is an Associate Professor in the Faculty of Law at the University of Technology Sydney and the creator behind the Gender Legislative Index. She was the 2022 winner of the Australia and New Zealand Women in AI (Law) Award and 2nd Runner-up for the Woman in AI Innovator of the Year. Address: University of Technology Sydney, Faculty of Ultimo NSW 2007, Australia Email: ramona.vijayarasa@uts.edu.au. The author would like to thank Senior Research Fellow, José-Miguel Bello Villarino, at the University of Sydney’s Automated Decision-Making + Society Centre for his feedback on earlier drafts and her research assistant, Wendy Lam, for excellent support in undertaking background research for this article.

¹ CLEMENTINE COLLETT ET AL., UNESCO, THE EFFECTS OF AI ON THE WORKING LIVES OF WOMEN 62 (2022).

² *Id.* at 63.

³ EQUALS’ and UNESCO’s 2019 publication was entitled, “I’d blush if I could,” the response given by Siri, a female-gendered voice assistant used by hundreds of millions of people, when a human user would direct at “her” online abuse. See MARK WEST ET AL., I’D BLUSH IF I COULD: CLOSING GENDER DIVIDES IN DIGITAL SKILLS THROUGH EDUCATION 4 (EQUALS & United Nations Educational, Scientific and Cultural Organization, 2019), <https://doi.org/10.54675/RAPC9356> (last visited May 29, 2023).

⁴ COLLETT ET AL., *supra* note 1, at 63.

These efforts to teach members of the public that harassment against an AI-driven chatbot is “disrespectful and invasive to real women” became the subject of much academic debate.⁵ Yet, for the purpose of this article, BIA exemplifies one of its core arguments: it is not that the vast majority of AI-related harms originate with AI, but rather that AI replicates and reinforces the discrimination, inequality, and harm already suffered by a diversity of women in society. In other words, AI “mirrors” – and often magnifies – preexisting gender biases.⁶ Moreover, much of this gender-based harm is, to the surprise of some, relatively well known. The issue at hand, therefore, is *not* that the gendered implications of AI are vastly under-researched or even under-estimated. The challenge, instead, is how to build on the scholarship that has deepened our understanding of the gendered implications of AI⁷ and to bring this gendered lens to the task of legislating.

The gender-based harms of AI would have been well beyond the concerns of most women’s rights activists a decade ago. Today, however, the gendered implications of AI’s rapid deployment have been the subject

⁵ Joana Moreira Ferreira, *Harassment and Offenses to Conversational Agents: The Case of BIA, the Feminist Chatbot* (Oct. 12, 2022) (M.Sc. in Marketing dissertation, Universidad Católica Portuguesa), <https://repositorio.ucp.pt/handle/10400.14/39425>; Beatrys Rodrigues & André Peruzzo, *Real Harassment Virtual Robots? Implications of Online Harassment Geared at Virtual Assistant BIA*, *AOIR SELECTED PAPERS OF INTERNET RESEARCH* (2022), <https://spir.aoir.org/ojs/index.php/spir/article/view/13079> (last visited May 29, 2023); Mauriceia Rodrigues Barbosa, *Assédio Sexual Em Uma Escola Pública Do Município De Abaetetuba/Pa*, 9 *DIVERSIDADE E EDUCAÇÃO* 110 (2021); K.C. Santosh & Casey Wall, *AI and Ethical Issues*, in *AI, ETHICAL ISSUES AND EXPLAINABILITY—APPLIED BIOMETRICS* 1 (2022), https://link.springer.com/10.1007/978-981-19-3935-8_1 (last visited Aug 31, 2022); Denise Braga Sampaio et al., *Violência Contra a Mulher Na Perspectiva Dos Regimes de Informação: Uma Análise Sobre o Machismo Direcionado a Assistentes Digitais*, 14 *TENDÊNCIAS DA PESQUISA BRASILEIRA EM CIÊNCIA DA INFORMAÇÃO* 1 (2021), <https://revistas.ancib.org/index.php/tpbci/article/view/584> (last visited May 29, 2023).

⁶ Ardra Manasi et al., *Mirroring the Bias: Gender and Artificial Intelligence*, 26 *GENDER TECH. & DEV.* 295 (2022).

⁷ María López Belloso, *Women’s Rights Under AI Regulation: Fighting AI Gender Bias Through a Feminist and Intersectional Approach*, in *LAW AND ARTIFICIAL INTELLIGENCE: REGULATING AI AND APPLYING AI IN LEGAL PRACTICE* 87 (Bart Custers & Eduard Fosch-Villaronga eds., 2022), https://doi.org/10.1007/978-94-6265-523-2_5 (last visited Mar. 15, 2023).

of a notable body of scholarship by legal⁸ and computer science scholars.⁹ Numerous examples from the past 15 years could fill the pages of this article: from pornographic deep fakes¹⁰ to menstrual apps appropriated by the police to identify abortion users in the US.¹¹ It is unsurprising, therefore, that human rights scholars Elizabeth Coombs and Halefom Abraha described gender stereotypes as “part of AI’s fabric.”¹²

As many scholars from different disciplines weigh in on these regulatory debates, some suggest that the gravest concern that AI poses is its misuse by authoritarian regimes.¹³ In other writings with co-author José-Miguel Bello y Villarino, we have respectfully disagreed.¹⁴ Rather, it is the normalized, daily use of AI that will have far more persistent, widespread, and perhaps even largely undetected implications for women’s lives.¹⁵ Given the identified gender biases that AI risks

⁸ For example Ramona Vijeyarasa & José-Miguel Bello y Villarino, *Lessons and Consequences of the Failure to Regulate AI for Women’s Human Rights*, OPENGLOBALRIGHTS, <https://www.openglobalrights.org/lessons-and-consequences-of-failure-to-regulate-ai/> (last visited Sept. 20, 2022); José-Miguel Bello y Villarino & Ramona Vijeyarasa, *International Human Rights, Artificial Intelligence and the Challenge for the Pondering State: Time to Regulate?*, NORDIC J. OF HUM. RIGHTS 1, 194 (2022); ELIZABETH COOMBS & HALEFOM ABRAHA, GOVERNANCE OF AI AND GENDER: BUILDING ON INTERNATIONAL HUMAN RIGHTS LAW AND RELEVANT REGIONAL FRAMEWORKS (Andrej Zwitter & Oskjar J. Gstrein eds., 2022); *Artificial Intelligence and Gender Equality: Key Findings of UNESCO’s Global Dialogue*, UNESCO (Aug. 2020), <https://unesdoc.unesco.org/ark:/48223/pf0000374174> (last visited Aug 30, 2022) [hereinafter UNESCO]; Rachel Adams & Nóra Ní Loideáin, *Addressing Indirect Discrimination and Gender Stereotypes in AI Virtual Personal Assistants: The Role of International Human Rights Law*, 8 CAMBRIDGE INT’L L.J. 241 (2019); Prashant Chauhan & Gagandeep Kaur, *Gender Bias and Artificial Intelligence: A Challenge within the Periphery of Human Rights*, 8 HASANUDDIN L. REV. 46 (2022); López Belloso, *supra* note 7.

⁹ For example, Enrique Latorre Ruiz & Eulalia Pérez Sedeño, *Gender Bias in Artificial Intelligence*, in GENDER IN AI AND ROBOTICS: THE GENDER CHALLENGES FROM AN INTERDISCIPLINARY PERSPECTIVE 61 (Jordi Vallverdú ed., 2023), https://doi.org/10.1007/978-3-031-21606-0_4 (last visited June 12, 2023); Susan Leavy, *Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning*, in 2018 IEEE/ACM 1ST INTERNATIONAL WORKSHOP ON GENDER EQUALITY IN SOFTWARE ENGINEERING (GE) 14 (2018); Susan Leavy et al., *Data Power and Bias in Artificial Intelligence* (2020), <http://arxiv.org/abs/2008.07341> (last visited June 12, 2023).

¹⁰ Travis L. Wagner & Ashley Blewer, “The Word Real Is No Longer Real”: Deepfakes, Gender, and the Challenges of AI-Altered Video, 3 OPEN INFO. SCI. 32 (2019).

¹¹ Jeremy Kahn, *After Roe, Fears Mount over A.I.’s Ability to Identify Those Seeking Abortion*, FORTUNE (June 29, 2023), <https://fortune.com/2022/06/28/after-roe-v-wade-fear-of-a-i-surveillance-abortion/>.

¹² COOMBS & ABRAHA, *supra* note 8.

¹³ Eileen Donahoe & Megan MacDuffee Metzger, *Artificial Intelligence and Human Rights*, 30 J. DEMOCRACY 115 (2019).

¹⁴ Vijeyarasa & Bello y Villarino, *supra* note 8, at 2.

¹⁵ Vijeyarasa & Bello y Villarino, *supra* note 8.

replicating,¹⁶ a gender-responsive approach to regulating its design, deployment, and use is a must if nations choose to adopt a human rights-centered approach to how AI is governed.¹⁷ A failure to do so may mean a missed opportunity to get ahead of the technology as numerous jurisdictions are currently engaging in the “race to regulate.”¹⁸

In response, the central research question driving this article is, with the knowledge at hand, how can AI be regulated with women’s rights in mind? In this respect, this article sits within a wider body of scholarship that has, for decades, sought to bring a women’s standpoint to lawmaking.¹⁹ Several decades ago, scholars of gender and the law identified how gender “neutral” legislation was erasing women from legislative debates; the outcomes were “anything but neutral.”²⁰ These earlier scholars challenged legal “neutrality” across a range of domains, from the law’s response to rape and sexual violence²¹ to the likely effect of tax laws on women.²²

Here, I take this pioneering work and ask how to avoid the gender-neutral treatment of AI’s regulation and, contrastingly, bring this gender-responsive legislative approach to the regulation of new and emerging technologies. Gender-responsive lawmaking is understood as achieving legislation that responds to the specific needs of different sexes and considers different gendered perspectives on pivotal economic, social, and political issues.²³ With regard to the existing literature that examines the regulation of gender bias, much of this work tends to refer to existing

¹⁶ UNESCO, *supra* note 8.

¹⁷ Donahoe & Metzger, *supra* note 13.

¹⁸ JOSÉ-MIGUEL BELLO Y VILLARINO ET AL., STANDARDISATION, TRUST AND DEMOCRATIC PRINCIPLES: THE GLOBAL RACE TO REGULATE ARTIFICIAL INTELLIGENCE (2023), <https://www.ussc.edu.au/analysis/standardisation-trust-and-democratic-principles-the-global-race-to-regulate-artificial-intelligence> (last visited Nov. 30, 2023).

¹⁹ Margaret Thornton, *Feminist Jurisprudence: Illusion or Reality?*, 3 AUSTL. J.L & SOC’Y 5, 12 (1986); Katharine T. Bartlett, *Feminist Legal Methods*, 103 HARV. L REV. 829, 837 (1990); Susan Boyd & Elizabeth Sheehy, *Canadian Feminist Perspectives on Law*, 13 J.L & SOC’Y 283, 283 (1986); CAROL SMART, *FEMINISM AND THE POWER OF LAW* (Routledge 1989); Reg Graycar & Jenny Morgan, *The Hidden Gender of Law: A Public Talk*, 41 AUSTL. FEMINIST L.J 29 (2015).

²⁰ Ruth Halperin-Kaddari & Marsha A. Freeman, *Backlash Goes Global: Men’s Groups, Patriarchal Family Policy, and the False Promise of Gender-Neutral Laws*, 28 CAN. J. WOMEN & L. 182, 189 (2016), <https://www.utpjournals.press/doi/abs/10.3138/cjwl.28.1.182> (last visited Mar. 21, 2019).

²¹ Annabelle Mooney, *When a Woman Needs to Be Seen, Heard and Written as a Woman: Rape, Law and an Argument against Gender Neutral Language*, 19 INT’L J. SEMIOTICS L. 39, 62 (2006).

²² See generally EDWARD J. MCCAFFERY, *TAXING WOMEN* (Univ. of Chi. Press, 1997); Patricia A. Cain, *Taxing Lesbians*, 6 S. CAL. REV. L. & WOMEN’S STUD. 471 (1996).

²³ Ramona Vijayarasa, *Making the Law Work for Women: Standard-Setting through a New Gender Legislative Index*, 44 ALT. L.J. 275, 277 (2019).

legislation, such as the application of anti-discrimination legislation to AI.²⁴ By contrast, the existing scholarship pays significantly less attention to the potential for new legal provisions to be drafted and incorporated into AI-specific regulations to overcome such gender bias. Moreover, AI is an area requiring both an individual lens but also a group one that accounts for collective harm. This includes harms to transgender women, lesbian women, and women of color who suffered AI's intersectional harms. A focus on new and emerging AI-specific provisions that can respond to gendered harms is the most significant contribution of this article.

Moreover, I seek to go beyond this task by addressing an issue given far less attention in the scholarship. Can AI be deployed in a way that can even correct pre-existing inequality, including intersectional gender-based harm? In this respect, writing on an equity lens to AI, Carmina Ravanera and Sarah Kaplan refer to AI's "double-edged sword."²⁵ AI poses risks to women – and to equality achieved to date – but offers notable potentiality to advance equality, as the examples discussed in Part IV demonstrate. Yet at the same time, a vast majority of scholars remain skeptical of the "ICT for development" space, given the indisputable ongoing inequities that pervade access to technologies within countries but also in light of the global North-South divide.²⁶ Unless this "algorithmic divide" is bridged, many of the political, social, economic, cultural, educational, and employment opportunities provided by machine learning and artificial intelligence will be undermined.²⁷

In the following section (Part II) of this article, I provide the context to enable readers to better understand what is AI and what are the gendered implications of these new technologies. I offer an overview of the scholarly debates concerning the gendered harms that have existed since the study of AI emerged, although I pay particular attention to writing from the last 15 years. By and large, I categorize such literature into two types of AI harms: allocative harm and representative harm.

²⁴ Robert Bartlett et al., *Algorithmic Discrimination and Input Accountability Under the Civil Rights Acts*, 36 BERKELEY TECH. L.J. 675, 678 (2021); PAULINE T. KIM, AI AND INEQUALITY 15–17 (Wash. Univ. in St. Louis L. Stud. Rsch. Paper Series 2021); Maya C. Jackson, *Artificial Intelligence & Algorithmic Bias: The Issues with Technology Reflecting History & Humans*, 16 J. BUS. & TECH. L. 299, 314–315 (2021).

²⁵ CARMINA RAVANERA & SARAH KAPLAN, AN EQUITY LENS ON ARTIFICIAL INTELLIGENCE 2 (Inst. for Gender and the Econ., Rotman Sch. of Mgmt., Univ. of Toronto 2021).

²⁶ Jolynna Sinanan & Tom McNamara, *Great AI Divides? Automated Decision-Making Technologies and Dreams of Development*, 35 CONTINUUM 747, 748 (2021).

²⁷ Peter K. Yu, *The Algorithmic Divide and Equality in the Age of Artificial Intelligence*, 72 FLA. L. REV. 331, 331 (2020).

What becomes clear in these sections is that AI may not be the root cause of the problem but rather the challenge lies in how AI amplifies existing societal inequalities. Throughout this examination, several examples are offered that speak to intersectional gendered biases. That is, the presence of multiple and compounded biases, such as gender and race, gender and socio-economic status or gender and age.²⁸ Yet it is important to acknowledge at the very outset that there is, at present, inadequate attention paid to the intersectional experiences of the gendered harms of AI. One particularly overlooked experience of the gender-based harms of AI is that facing non-binary people, a point I touch on briefly in Part II.

In Part III, I turn to the key task at hand: regulating AI with women in mind. Through a comparative study of those countries that demonstrate a weak response to gendered bias in their regulatory approaches to those examples that offer global promise—including Canada, Brazil and the EU—we arrive at a better understanding of how non-discrimination, equality and bias can be embedded into legislation on AI to address some of the harms experienced by women as a collective. In concluding, I consider how we might regulate better to achieve algorithmic fairness for a greater diversity of women.

AI regulation is just emerging. By exploring the various ways in which AI impacts women's lives, both directly and indirectly, we can better appreciate the considerations that lawmakers in this field must undergo when drafting legislation. This is a widely shared goal,²⁹ including, it appears, among some platform owners.³⁰ While gender may not be on the minds of all when engaged in AI-related debates, this regulatory turning point offers an opportune moment to get the regulation of AI for women right from the very outset.

²⁸ Kimberle Crenshaw, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, 140 UNIV. CHI. L.F. 139, 140 (1989); Jennifer C. Nash, *Re-Thinking Intersectionality*, 89 FEMINIST REV. 1 (2008); JOANNE CONAGHAN, INTERSECTIONALITY AND THE FEMINIST PROJECT IN LAW 2 (Cooper, Davina eds., 2008).

²⁹ UNESCO, *supra* note 8, at 5.

³⁰ Johana Bhuiyan, *OpenAI CEO Calls for Laws to Mitigate 'Risks of Increasingly Powerful' AI*, THE GUARDIAN (May 17, 2023, 3:45 PM), <https://www.theguardian.com/technology/2023/may/16/ceo-openai-chatgpt-ai-tech-regulations>.

II. CONTEXT: UNDERSTANDING THE GENDERED IMPLICATIONS OF ARTIFICIAL INTELLIGENCE

A. *What is AI and What Makes it Intelligent?*

AI is a relatively old idea but a relatively new reality as part of our daily lives.³¹ At the time this article went to print, debates over whether states should regulate AI or leave the creators of AI-driven technology to self-regulate, what one scholar describes as the “Silicon Valley narrative,”³² remain robust discussions.³³ Yet it is worth taking a step back to understand what exactly we are trying to regulate. Intelligence is a human trait. AI attempts to reproduce that “human intelligence” with similar or better results.³⁴ AI involves the use of “computers to classify, analyse, and draw predictions from data sets, using a set of rules, which we often call algorithms.”³⁵ Within the concept of AI, there are many ways of training algorithms to use different data sets. You can have rules-based algorithms (“if x occurs, then y”). These tend to be static and do not change their outputs for the same queries. You can alternatively have systems that are capable of evolving such as those based on machine learning.³⁶ Hence the name: they evolve as they learn.

Deep learning is a sub-set of machine-learning that looks for patterns or connections in data that are less apparent and more difficult to calculate.³⁷ Not all deep learning is problematic; rather, deep learning advances have been deployed to solve complex problems.³⁸ Yet those same deep learning-powered advances have been deployed in much more concerning ways, including in the form of deepfake algorithms. A technology that was brought to the mainstream in 2017, deepfakes use AI to transfer or map an image onto an existing video.³⁹ Deepfake content may be created using two algorithms in competition with each other: one, the generator, creates the fake digital image while the other, the

³¹ J. McCarthy et al., *Dartmouth Summer Research Project on Artificial Intelligence* (1956), <https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth>.

³² CÉLINE CASTETS-RENARD, *AI AND THE LAW IN THE EU AND THE US* 384 (Florian Martin-Bariteau & Teresa Scassa eds., 2020).

³³ GLION HUMAN RIGHTS DIALOGUE 21 (Universal Rights Group ed., 2020).

³⁴ Santosh & Wall, *supra* note 5, at 2.

³⁵ UNESCO, *supra* note 8, at 4.

³⁶ *Id.*

³⁷ Santosh & Wall, *supra* note 5, at 4.

³⁸ Xizhao Wang et al., *Recent Advances in Deep Learning*, 11 INT'L J. MACH. LEARN. & CYBERNETICS 747, 747 (2020).

³⁹ Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALI. L. REV. 1753, 1757 (2019).

discriminator, decides if the content is real or artificial. The generator is then made aware of whether the discriminator correctly identified the deepfake as real or fake, allowing it to learn and improve its next deepfake.⁴⁰ According to Forbes, in 2019 alone, the number of “deepfakes” doubled from 7,964 to 14,679.⁴¹ The legal problems created by deepfakes are well-known.⁴² They can be highly convincing. They are viewed and understood to be real videos of the people depicted in them, inciting political deception, voter manipulation and commercial fraud.⁴³

Some viral deepfakes may be better known⁴⁴—Barack Obama swearing at Donald Trump or Mark Zuckerberg bragging about stealing users’ data.⁴⁵ Yet these forms of deepfakes have also been the object of specific regulation. For instance, in 2019, the State of California’s Governor signed a law making it illegal to create or distribute doctored videos, images, or audio of politicians within sixty days of an election, driven by concerns as to how they would sway voters.⁴⁶ Yet when the political and pornographic come together, women are the primary targets and may be less protected in certain jurisdictions. In 2016, the legal counsel of President Rodrigo Duterte of the Philippines used what was most likely a deepfake pornographic video of Senator Leila De Lima⁴⁷ to justify her imprisonment, shortly after she initiated a Senate Inquiry into the extrajudicial executions of thousands of Filipino drug offenders during Duterte’s tenure. Unsurprisingly, the use of AI technology in this

⁴⁰ THANH THI NGUYEN ET AL., DEEP LEARNING FOR DEEPFAKES CREATION AND DETECTION: A SURVEY 3–5 (2022).

⁴¹ Rob Toews, *Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared*, FORBES (May 25, 2020, 11:45 PM), <https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/>.

⁴² Tyrone Kirchengast, *Deepfakes and Image Manipulation: Criminalisation and Control*, 29 INFO. & COMM’NS TECH. L. 308, 308 (2020).

⁴³ *Id.* at 308–09.

⁴⁴ Andrew Ray, *Disinformation, Deepfakes and Democracies: The Need for Legislative Reform*, 44 UNSW L.J. 983, 986 (2021).

⁴⁵ Ian Sample, *What Are Deepfakes – and How Can You Spot Them?*, THE GUARDIAN (Jan. 13, 2020), <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them>.

⁴⁶ Kari Paul, *California Makes ‘Deepfake’ Videos Illegal, but Law May Be Hard to Enforce*, THE GUARDIAN (Oct. 7, 2019), <https://www.theguardian.com/us-news/2019/oct/07/california-makes-deepfake-videos-illegal-but-law-may-be-hard-to-enforce>.

⁴⁷ MATTHEW J. DAVIS & PER FORS, TOWARDS A TYPOLOGY OF INTENTIONALLY INACCURATE REPRESENTATIONS OF REALITY IN MEDIA CONTENT 291 (14th IFIP TC 9 Int’l Conf. on Hum. Choice & Comput., 2020).

case was considered by many commentators as retribution for her critique of his authoritarian rule.⁴⁸

A further challenge with AI is its end use when it comes to decision-making. AI may enable automated decision-making or be incorporated into decision support systems (DSS). Both have widespread implications for women as they rely on probabilistic algorithms to make inferences by learning existing patterns from pre-existing data.⁴⁹ These are prone to biases and systemic unfairness.⁵⁰ Using such an algorithm, an automated decision-making system may “automatically” grant or deny applications for a bank loan based on the available data. Alternatively, DSS might facilitate a bank manager’s decision-making about a particular application by providing information about the applicant’s security of employment and offering insights from similar applicants in comparable circumstances that have already been processed.

Research demonstrates that decision support systems for credit loan applications favor certain socio-demographic groups in a disproportional way, disadvantaging people living in certain areas, of specific ethnic backgrounds, and women.⁵¹ This becomes an obvious challenge for women if certain applicants—e.g., women of color or women sole-parents—are singled out as less likely to repay a loan because the system has learnt from the repeated denials of applications in the past of people sharing similar traits, such as sex, age or race.⁵² Here we begin to understand the ways in which AI may not necessarily be the root-cause of the problem but rather it is a historical one whereby AI may replicate and reproduce existing assumptions and biases pertaining to both the majority and marginalized groups if those biases are embedded in datasets used for design, testing and training of algorithms.⁵³

With this brief overview of the technology at hand, I now turn to the scholarly treatment of the gender-based harms that result from AI. While the gendered harms of AI are relatively known, here I present them

⁴⁸ BRITT PARIS & JOAN DONOVAN, DEEPFAKES AND CHEAP FAKES 27 (Data & Soc’y’s Rsch. Inst., 2019), <https://apo.org.au/node/259911> (last visited Feb. 2, 2023).

⁴⁹ For a fuller discussion of how DSS works and its potential use in the determination of applications for refugee status, see Bello y Villarino & Vijayarasa, *supra* note 8.

⁵⁰ Stefan Feuerriegel et al., *Fair AI*, 62 BUS. INFO. SYS. ENG’G. 379, 379 (2020).

⁵¹ *Id.*; see also ALGORITHMIC DISCRIMINATION IN EUROPE: CHALLENGES AND OPPORTUNITIES FOR GENDER EQUALITY AND NON-DISCRIMINATION LAW 33–34 (European Commission ed., 2021) [hereinafter ALGORITHMIC DISCRIMINATION].

⁵² HEIKKI HIILAMO, HOUSEHOLD DEBT AND ECONOMIC CRISES 109 (Edward Elgar Publ’g Ltd. 2018).

⁵³ Michael Fay & Lesley Williams, *Gender Bias and the Availability of Business Loans*, 8 J. BUS. VENTURING 363, 363 (1993).

to readers in a classified manner for clarity and to enhance comprehension of such harms. Two types of harms—allocative and representative harms—are explained before turning to a third under-explored harm: gaps in AI literacy. As is evident, there has been a palpable swell in the volume of literature addressing the gendered implications of AI in the last three to five years.

B. An Overview: Scholarly Debates and AI's Gendered Harms

There is a growing appreciation that AI affects women's lives in specific ways. Bringing such a gendered lens to the topic of AI allows us to challenge sexism and other forms of oppression evident in AI-driven technologies. Catherine D'Ignazio and Lauren Klein write of "data feminism" as a perspective that seeks to bring a new lens to how we value and evaluate data by rethinking binaries and hierarchies, embracing pluralism, and considering context, among other things.⁵⁴

Much of these gendered concerns lie with algorithmic bias.⁵⁵ The European Commission's Gender Equality Strategy 2020-2025 recognizes the risk that AI will intensify gender-based inequities through representative harm.⁵⁶ It identifies six challenges that algorithms pose for the advancement of more gender-equal societies. Put simply, (a) harmful human stereotypes may "infect" the algorithm which then reinforces them; (b) training algorithms on biased, incorrect, unrepresentative or unbalanced data can reproduce structural inequalities; (c) algorithms offer correlations without question ("gender" = "poor work performance"), reifying discrimination; (d) algorithms are often too hard to understand (even for computer scientists), leaving decision-making opaque; and (e) AI-based discrimination can spread at a much faster speed and on a much bigger scale than human discrimination.⁵⁷ Algorithmic-driven data may, for example, inaccurately reflect mass behavior, entail the use of incomplete supervision mechanisms, or provide insufficient protections for those impacted. These datasets encourage the replication of existing biases, including gender biases, because the datasets are too small or may only include part of the data.⁵⁸ More often, the data may simply be based on a biased

⁵⁴ CATHERINE D'IGNAZIO & LAUREN F. KLEIN, DATA FEMINISM 18 (MIT Press ed. 2020).

⁵⁵ AUSTRALIAN GOVERNMENT DEPARTMENT OF INDUSTRY, SCIENCE & RESOURCES, SAFE AND RESPONSIBLE AI IN AUSTRALIA: DISCUSSION PAPER 8 (2023) [hereinafter SAFE AND RESPONSIBLE AI].

⁵⁶ ALGORITHMIC DISCRIMINATION, *supra* note 51, at 7.

⁵⁷ *Id.* at 8.

⁵⁸ SAFE AND RESPONSIBLE AI, *supra* note 55, at 8.

reality, that is, based on “facts” that are already underpinned by societies’ existing gender biases. Existing gendered inequalities are, in turn, exacerbated.

Before setting out the literature on the allocative and representative harms of AI, it is important to draw attention to one group of individuals who are particularly impacted by algorithmic bias but largely overlooked in the literature: lesbian, gay, bisexual, transgender, and queer (LGBTQ+) communities. Many AI technologies continue to operate using the binaries of male and female.⁵⁹ The key problem lies with the reality that registration processes frequently provide for only binary choices of “male” or “female,” which for non-binary people is not simply difficult to answer but potentially offensive to complete. Indeed, there is a far greater complexity to the discussion on gender and AI than I am able to do justice to in this article in calling for a shift beyond binaries in the use of automated body scanners, facial recognition, or social media content filtering, just to name a few examples. Misclassification and misrecognition are inadequately addressed in existing privacy regulations;⁶⁰ no US state, as of 2022, has in place legislation to address the consequences of gender misidentification by AI systems.⁶¹ Canadian scholars too have pointed out, albeit in passing, the lack of accuracy of results for non-binary and gender diverse individuals when generating results from, for example, patient intake forms that retain the outdated “he” and “she.”⁶²

While Facebook has been commended for being “ahead of the curve” when it expanded its gender categories in 2014 from the standard two to over fifty choices, ranging from “Genderqueer” to “Neither,” behind the scenes Facebook continues to resort to gender binaries which make it easier, for instance, for paid advertisers to undertake marketing on Facebook; binaries remain the preferred choice of Facebook’s paying clients.⁶³ Here, I agree with the arguments of Sonia Katyal and Jessica Jung that the legal-oriented solutions that I propose in this article have their limitations and more needs to be done at the design stage.⁶⁴

⁵⁹ Sonia K. Katyal & Jessica Y. Jung, *The Gender Panopticon: AI, Gender, and Design Justice*, 68 UCLA L. REV. 692, 699 (2021).

⁶⁰ *Id.* at 761.

⁶¹ *Id.*

⁶² Bradley Henderson et al., *Artificial Intelligence in Canadian Healthcare: Will the Law Protect Us from Algorithmic Bias Resulting in Discrimination?*, 19 CAN. J. L. & TECH. 475, 483 (2022).

⁶³ D’IGNAZIO & KLEIN, *supra* note 54, at 100.

⁶⁴ Katyal & Jung, *supra* note 59, at 762–763 (Katyal and Jung see a notable limit to what law can do in this space. Rather than focus on regulation, they place a greater

Turning to the literature on the harms of AI for women, the UK's Information Commissioner's Office⁶⁵—an independent body focused on information rights—has offered us a dual classification method to understand AI's harm. Harms to an individual are either allocative harms or representational harms. Allocative harms result from decisions about how to allocate goods and opportunities among a group. Here, we can think of the way an AI system used in a recruitment process may disproportionately classify applications for male candidates as more suitable than female. Such a system potentially results in a loss of financial opportunities, livelihoods, and freedom of choice for women when compared to men.⁶⁶

The other type of harm is representational. Representational harm comes about when systems reinforce gendered subordination through stereotyping, under-representation, or denigration. The example offered here is where an AI system may take an open-source photo and assign to it a denigrating—for example, racist—trope.⁶⁷ Other scholars have described this two-part approach to AI's harms as how AI withholds opportunities on the one hand and the way AI imposes burdens on the other.⁶⁸

For the purposes of this article, I present here the scholarly debates from the last two decades, largely categorizing these discussions according to these two types of harms. Some debates remain as to whether allocative harms or representational harms are greater. For example, researchers at UN Women have called out an overemphasis on the computational factors related to the statistical representativeness of the data. In their view, certain types of representational harm are less concerning than the human and systemic institutional and societal

emphasis on an earlier stage and argue that gendered differences should be accounted for at the design-stage. That is, to these scholars, there is more value in design-oriented solutions. For instance, Katyal and Jung want to see AI-driven technologies deployed in a way that AI users can more easily self-determine their identity or where the AI-driven technologies acknowledge a plurality of genders beyond male and female).

⁶⁵ *Information Commissioner's Office (ICO)*, INFO. COMM'R OFF., <https://ico.org.uk/> (last visited Dec. 22, 2023).

⁶⁶ *What are the Accountability and Governance Implications of AI?*, INFO. COMM'R OFF., <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/what-are-the-accountability-and-governance-implications-of-ai/> (last visited Apr. 5, 2023).

⁶⁷ *Id.*

⁶⁸ ADAM LEON SMITH ET AL., *REGULATORY FRAMEWORKS RELATING TO DATA PRIVACY AND ALGORITHMIC DECISION MAKING IN THE CONTEXT OF EMERGING STANDARDS ON ALGORITHMIC BIAS* 5 (2018).

factors that deny women access and resources, i.e., allocative harms.⁶⁹ These tensions and debates are elaborated further below. I conclude this section by adding a rarely acknowledge third type of harm: inequality when it comes to AI literacy.

i. Allocative Harms

Allocative harms principally emerge from wrongful discrimination that results from automated or assisted decision-making. While human organizations can define the parameters of the decision in order to at least reduce and possibly remove human biases, in practice, a number of features of assisted or automated decision-making can create biases: a reliance on categorizations that replicate existing bias, emerging from the under- or over-representation of particular data or the biases of the individual algorithmic designers;⁷⁰ automation and predictions that lead to wrongful generalizations; and a lack of transparency in decision-making that undermines the ability to explain,⁷¹ reducing the room for accountability.

Prejudice in relation to job vacancies is a core subject of this literature. Algorithms are said to impede the opportunities afforded to women to advance.⁷² Amazon's AI-driven recruitment tool, developed around 2014, was scrapped back in 2018 when it became clear that it would teach itself that male candidates were better than female ones,⁷³ penalizing resumes with the word "women," whether that was "Women's Chess Club Champion" or a graduate from an "all-women" college.⁷⁴ Rather than helping to diversify the workforce, it was perpetuating men's dominance of the sector as the algorithm was trained on ten years of resumes that were primarily submitted by men. In other words, algorithmic bias is created and reinforced by a feedback loop. This is a typical trait of AI.⁷⁵ One study of over 60,000 ads published on Google

⁶⁹ ELEONORA LAMM ET AL., *THE GENDERED IMPACTS OF AI: POLICIES AND SAFEGUARDS TO REGULATE NEW TECHNOLOGIES, MITIGATE RISKS AND PROTECT RIGHTS* 4 (2022).

⁷⁰ Jackson, *supra* note 24, at 299.

⁷¹ Hugo Cossette-Lefebvre & Jocelyn Maclure, *AI's Fairness Problem: Understanding Wrongful Discrimination in the Context of Automated Decision-Making*, 3 *AI & ETHICS* 1255 (2022), <https://doi.org/10.1007/s43681-022-00233-w>.

⁷² Chauhan & Kaur, *supra* note 8, at 48.

⁷³ Julien Lauret, *Amazon's Sexist AI Recruiting Tool: How Did it go so Wrong?*, *MEDIUM* (Aug. 16, 2019), <https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e>.

⁷⁴ *Id.*

⁷⁵ Adriane Chapman et al., *A Data-Driven Analysis of the Interplay between Criminological Theory and Predictive Policing Algorithms*, in 2022 ACM CONF. ON

Ads for employment found that adjusting a user's gender to "female" brought up fewer instances of advertisements for high-paying jobs than for users selecting "male."⁷⁶

Moreover, such use of AI-driven technologies in employment decisions persists partly because of the dearth of academic literature, law, or judicial guidance on how to address the employment-related impacts of discriminatory algorithms.⁷⁷ Yet regulation can frequently play a role. For instance, AI-related legislation in an employment context can require, for example, notification to job applicants that an automated hiring system will be used in the recruitment process.⁷⁸ In fact, in July 2023, New York City put in place prohibitions on the use by employers of "automated employment decision tools" or AEDTs to screen a candidate or employee for employment-related decisions unless the tool has been subject to an audit for biases that has been conducted within the year prior and all notice requirements have been complied with.⁷⁹ "Independent auditors" cannot be employed by the organization or have a vested financial interest.⁸⁰

Arguably the New York approach needs replicating on a mass scale. A study published in 2020 of human resource professionals representing 500 mid-sized organizations from various industries in five different countries found that in 2020, 24% of businesses had already implemented AI for recruitment purposes, and 56% of hiring managers planned to adopt it in the following year.⁸¹ In fact, recruitment using AI tools saw a new spike during COVID-19. At that time, many human resource practitioners were unsatisfied with traditional methods, while lockdown restrictions saw an upturn in the value placed on AI-driven

FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 36 (June 20, 2022), <http://doi.org/10.1145/3531146.3533071>; Donghee Shin & Yong Jin Park, *Role of Fairness, Accountability, and Transparency in Algorithmic Affordance*, 98 *COMPUTS. IN HUM. BEHAV.* 277, 283 (2019).

⁷⁶ Amit Datta et al., *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination*, 1 *PROCEEDINGS ON PRIV. ENHANCING TECH.* 92,105 (2015).

⁷⁷ Natalie Sheard, *Employment Discrimination by Algorithm: Can Anyone Be Held Accountable?*, 45(2) *UNSW L.J.* 617, 619–20 (2022).

⁷⁸ *Id.* at 621.

⁷⁹ New York City Department of Consumer and Worker Protection, *Amendment to Title 6 of the Rules of the City of New York*, N.Y.C. RULES (2023), <https://rules.cityofnewyork.us/wp-content/uploads/2023/04/DCWP-NOA-for-Use-of-Automated-Employment-Decisionmaking-Tools-2.pdf>.

⁸⁰ *Id.*

⁸¹ Eleanor Drage & Kerry Mackereth, *Does AI Debias Recruitment? Race, Gender, and AI's "Eradication of Difference"*, 35 *PHIL. & TECH.* 1, 4 (Oct. 10, 2022), <https://doi.org/10.1007/s13347-022-00543-1>.

recruitment tools. HireVue, for instance, saw a 614% increase in their Japanese client's hiring activity during this time.⁸²

Yet the ability of algorithmic design to simply remove the category of “gender”—and similarly “race”—to avoid bias is far from a simple or appropriate solution. California-based startup Talent Sonar, for example, claims to deploy machine-learning algorithms to write job descriptions in a manner that improves gender diversity; the software hides applicants’ personal information like names and gender to reduce the unconscious biases of hiring managers.⁸³ Yet as some scholars note, a data point that serves as a proxy for a class of people, such as height and weight can be *de facto* discriminatory. Even when algorithms draw correlations from seemingly non-discriminatory data points, the outcome can be unfavorably discriminatory.⁸⁴ Hence, Drage and Mackereth note, “attempts to ‘strip’ gender and race from AI systems often misunderstand what gender and race are, casting them as isolatable attributes rather than broader systems of power.”⁸⁵ Questions remain today as to whether such tools live up to their claims that they are evolving to remove bias from the hiring processes strategically.⁸⁶

Access to financial resources is a further example commonly cited as reflecting AI’s allocative harms to particular groups of individuals. As Bartlett and colleagues point out, in the US, an individual’s creditworthiness reflects variables such as their cash flow and ability to pay back a loan, which legally is an acceptable business necessity for a bank to factor into its decision-making.⁸⁷ Yet when deployed at scale, the risk is the systematic penalizing of an entire minority group—such as women or people from minority neighborhoods. Such practices prove problematic under US anti-discrimination law.⁸⁸ The concern of these scholars is placed less on how to tune or fix a biased algorithm to achieve fairer outcomes and more on whether outcomes are the “result of structural inequalities requiring a more direct intervention to address their root cause.”⁸⁹

Moreover, other scholars raise concerns about the amplification of these harms that result from a lack of transparency. Latorre Ruiz and

⁸² *Id.* at 2.

⁸³ Nizan Geslevich Packin & Yafit Lev-Aretz, *Learning Algorithms and Discrimination*, in RESEARCH HANDBOOK ON THE LAW OF ARTIFICIAL INTELLIGENCE 88, 104 (Woodrow Barfield & Ugo Pagallo eds., 2018).

⁸⁴ *Id.* at 96–97.

⁸⁵ Drage & Mackereth, *supra* note 81, at 1.

⁸⁶ Drage & Mackereth, *supra* note 81, at 2–3.

⁸⁷ Bartlett et al., *supra* note 24, at 679.

⁸⁸ *Id.* at 680.

⁸⁹ *Id.* at 682.

Pérez Sedeño note that at times even the bank and finance staff do not know the reasons for the refusal when an AI-deployed system rejects a bank loan. Such a lack of transparency makes it difficult for individuals impacted by such decisions to identify whether they have suffered discrimination and, therefore, when they have a right to complain.⁹⁰ Moreover, the decisions go from individuals to mass impact relatively quickly. The scalability of AI means that a small discriminatory outcome by one bank if replicated in systems across several banks, can have significant and yet non-transparent discriminatory impacts overall.

The same can be said of other insidious uses of AI against women. AI's use to further undermine a woman's reproductive right to choose in the US probably best reflects its capacity to reinforce and amplify women's unequal positions in society. After the overturning of the constitutional right to an abortion in the US established in *Roe v. Wade* with the *Dobbs v. Jackson Women's Health Organisation* decision in June 2022,⁹¹ women across the US were warned to delete popular menstrual cycle apps and switch to web browsers that did not store their history.⁹² The 2022 arrest in the US State of Nebraska of a mother and her daughter, then aged 17, for an alleged abortion, tracked through chat history obtained from Facebook via a warrant, warned women of reproductive age what tech-based data could mean for their futures.⁹³

Some scholars of computing science have rightly argued that tackling AI-driven surveillance of abortion-seeking behavior removes only "a tile from the mosaic of ubiquitous surveillance."⁹⁴ In other words, the challenges go beyond the use of AI-based technologies in this case. There is a significant amount of amassed data involved in tracking women's reproductive decision-making: "search query data, text messages, license plate tracking, and easily purchased data from brokers."⁹⁵ In this case, the physical is entangled with the digital: physical photos taken of the number plate of a woman accessing a clinic may be used online to trace her address or to email her colleges about her abortion-related decision-making. In reality, these scenarios are complex

⁹⁰ Latorre Ruiz & Pérez Sedeño, *supra* note 9, at 71.

⁹¹ *Dobbs v. Jackson Women's Health Org.*, 142 S.Ct. 2228 (2022).

⁹² Kahn, *supra* note 11 (discussing that after *Roe*, concerns mount over A.I.-enabled surveillance).

⁹³ Johana Bhuiyan, *Facebook Gave Police Their Private Data. Now, This Duo Face Abortion Charges*, THE GUARDIAN (Aug. 10, 2022), <https://www.theguardian.com/us-news/2022/aug/10/facebook-user-data-abortion-nebraska-police> (last visited Aug. 17, 2022).

⁹⁴ Michela Meister & Karen Levy, *Digital Security and Reproductive Rights: Lessons for Feminist Cyberlaw*, in FEMINIST CYBERLAW 1, 7 (Oct. 31, 2022).

⁹⁵ *Id.*

and while AI-related regulations alone will not address the underlying challenges, law can assist in creating a consciousness of the differential experiences and impacts of AI-driven technologies on women. The challenge at hand is whether law can reach its optimum potential in better protecting women's rights.

ii. Representative Harms

AI technologies portray gender relations in particular ways, that is, there is a gender performance. In short, such technologies reify heteronormative gender roles and objectify women in the process.⁹⁶ This critique of AI is not new but rather can find a home in feminist theories on female representation since as early as the 1960s.⁹⁷ AI's gender biases may exist through stereotypical associations. The use of female voices in AI-powered virtual assistants—Amazon's Alexa, Apple's Siri, Microsoft's Cortana, and Google's Voice Assistant—was one of the earliest subjects of gender-specific AI literature, including among scholars of human rights and of gender and development.⁹⁸ While Google is the only one of these engines that does not adopt a female name, the female voice is the device's default. Meanwhile, in Nordic, Siri's name translates to "beautiful woman who leads you to victory," while Cortana's name is adapted from a character from Halo, the video game, that has a female avatar.⁹⁹ In the vast majority of cases, these technologies adopt a female voice. Scholars point out that the consequences are significant, not least because of the anticipated 8 billion AI-powered voice assistants in use by 2024.¹⁰⁰ That is more than one, on average, for every single person on the planet.¹⁰¹

These virtual assistants feed into a gendered stereotype that "she assists rather than directs; she pacifies rather than incites."¹⁰² Adams and Ni Loideáin go further and call out the obligation of States Parties under the Convention on the Elimination of All Forms of Discrimination against Women, the only standalone UN treaty dedicated to women's

⁹⁶ Manasi et al., *supra* note 6, at 297.

⁹⁷ Leavy, *supra* note 9, at 14.

⁹⁸ Manasi et al., *supra* note 6, at 295; Vijayarasa & Bello y Villarino, *supra* note 8.

⁹⁹ Manasi et al., *supra* note 6, at 298.

¹⁰⁰ *Id.*

¹⁰¹ Barry Elad, *Artificial Intelligence Statistics 2022: AI Usage and Voice Assistant Stats*, ENTERPRISE APPS TODAY (May 11, 2022),

<https://www.enterpriseappstoday.com/stats/artificial-intelligence-statistics.html>.

¹⁰² Rachel Adams et al., *Gender as Emotive AI and the Case of 'Nadia': Regulatory and Ethical Implications* 6 (Priv. L. Scholars Conf. 2021 Geo. Univ., Working Paper, 2021), <https://papers.ssrn.com/abstract=3858431>.

rights, to eliminate harmful gender stereotypes.¹⁰³ They also note the soft law obligations under the United Nations Guiding Principles on Business and Human Rights, to protect women from such discrimination at the hands of private actors.¹⁰⁴ Manasi and colleagues add to this discussion by indicating the contrasting use of male voices in instructing or teaching contexts. IBM's Watson, for example, uses a male voice to teach physicians about cancer treatment.¹⁰⁵ Meanwhile, with co-author José-Miguel Bello y Villarino, I have noted a risk of competing interests when it comes to the rights of different stakeholders when voice assistant technologies are deployed. For instance, Australia attempted to deploy an AI-driven voice assistant service, "Nadia", to help individuals living with a disability. Such technologies can enable marginalized groups to receive information on an equal basis with others – in this case, people with disabilities who would otherwise face barriers to accessing essential information. Yet, the technology itself risked reinforcing "assumptions associating the female gender with feelings of assurance, trust, safety, and placidity."¹⁰⁶ Its deployment was abandoned.

The use of "deepfakes" in pornography-related attacks are also a representative challenge with a gendered dimension.¹⁰⁷ Some estimates suggest that more than 19 out of every 20 deepfake videos on the internet in 2019 were pornographic.¹⁰⁸ Pornography has been the battle-ground for feminists for a near century.¹⁰⁹ While some scholars take the view that any pornography is harmful,¹¹⁰ the issue at hand is not whether one approves or disapproves of pornography. Rather, the harm that exists in this *non-consensual* form of gender-based online abuse originated well before AI and is now being replicated at a rapid pace through AI-driven technologies. This type of abuse distinctly affects women when compared to men or people of other genders.

¹⁰³ Adams & Loideáin, *supra* note 8, at 246–52.

¹⁰⁴ *Id.*

¹⁰⁵ Manasi et al., *supra* note 6, at 299.

¹⁰⁶ Vijayarasa & Bello y Villarino, *supra* note 8, at 204.

¹⁰⁷ HENRY AJDER ET AL., THE STATE OF DEEPFAKES: LANDSCAPE, THREATS, AND IMPACT 1 (2019), https://regmedia.co.uk/2019/10/08/deepfake_report.pdf; Wagner & Blewer, *supra* note 10, at 33. See also Amrita Khalid, *Deepfake Videos Are a Far, Far Bigger Problem for Women*, QUARTZ (Oct. 9, 2019), <https://qz.com/1723476/deepfake-videos-feature-mostly-porn-according-to-new-study-from-deeprace-labs/>.

¹⁰⁸ AJDER ET AL., *supra* note 107, at 1–2.

¹⁰⁹ RAMONA VIJAYARASA, SEX, SLAVERY AND THE TRAFFICKED WOMAN: MYTHS AND MISCONCEPTIONS ABOUT TRAFFICKING AND ITS VICTIMS 155 (Routledge 2016); Nadine Strossen, *A Feminist Critique of "the" Feminist Critique of Pornography*, 79 VA. L. REV. 1099 (1993).

¹¹⁰ Catharine A. MacKinnon, *Pornography as Trafficking*, 26 MICH. J. INT'L L. 993, 1008 (2005).

In some instances, gender-biased representation takes the form of language and gender-related associations. Natural language processing (NLP) is a form of AI that analyses human language to identify rules, classify results, and make predictions based on the patterns employed by humans in their use of language. Gender bias in NLP can take various forms. It can be both how women are represented—with certain gender biases embedded in word associations—as well as NLP’s ability to perform better on “majority genders,” i.e., men over women.¹¹¹ For instance, standard NLP may work better for a white male from California (as certain character traits may be considered as the NLP is developed) over a woman of Latino or Arabic descent.¹¹²

Word embedding is a natural language processing technique that manages to represent text data as vectors, allowing words with similar meanings to have a similar representation. Word embedding impacts, for instance, how Google News segments parts of its potential audience, delivering some content to some users and other content to others.¹¹³ Google image search results were found to produce more images of men in response to the gender-neutral keyword search “person” and “human” (in a nation’s dominant language).¹¹⁴ In other words, what should be a “gender-neutral” term and search that should produce an equal number of male and female images, nonetheless produced more images of male figures. Moreover, a further study from the same series in the United States on associations of “men” and “women” with particular professions again demonstrated that the search results do influence judgments and associations made by people between gender and work, even when it comes to new and previously unknown categories of occupations.¹¹⁵

NLP, therefore, risks propagating or even amplifying gender biases in text. Men are associated with “firefighters,” and women “nurses”; linguistic staging puts “son” before “daughter” and “Mr.” before “Mrs.,” while the press may describe a man by his behavior and a woman by her physical appearance and sexuality.¹¹⁶ A study conducted by

¹¹¹ Tony Sun et al., *Mitigating Gender Bias in Natural Language Processing: Literature Review*, CORNELL ARXIV (2019), <https://arxiv.org/abs/1906.08976v1> (last visited Sept. 21, 2022).

¹¹² Dirk Hovy & Shannon L. Spruit, *The Social Impact of Natural Language Processing*, PROCEEDINGS OF THE 54TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (VOLUME 2: SHORT PAPERS) 591, 593 (Aug. 2016), <https://aclanthology.org/P16-2096>.

¹¹³ Latorre Ruiz & Pérez Sedeño, *supra* note 9, at 68.

¹¹⁴ Madalina Vlasceanu & David M. Amodio, *Propagation of societal gender inequality by internet search algorithms*, 119 PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES 1, 1-2 (May 27, 2022).

¹¹⁵ *Id.* at 1–3.

¹¹⁶ Leavy, *supra* note 9, at 15.

Brazilian researchers in 2018 demonstrated similar problems with automated translations, where the machine translation is strongly biased towards male defaults, especially for fields such as STEM where sentences were incorrectly translated by Google Translate to give a male pronoun and not the correct female one.¹¹⁷ After the results of this study were released, Google Translate released a statement admitting the possibility of gender bias in its system.¹¹⁸

The above discussion is, of course, an abridged reflection of a bigger set of problems but provides a context for understanding the complex environment at hand. As discussed in Part III(B), methods to mitigate gender bias, such as in NLP, while nascent, do exist but the issue is how often these mitigating techniques are deployed.¹¹⁹ Therefore, many would say that the fear of AI for women and female-identifying individuals is self-explanatory. AI has in turn instigated valid concerns among law scholars,¹²⁰ and we have given significant consideration to the implications for women if AI is not properly regulated. However, less attention has been paid to what actually drives the problems that AI perpetuates. Revisiting the gendered harms, it becomes readily apparent that, in fact, it is not the AI that is the root of the cause. Rather, AI duplicates, reinforces, or replicates existing inequality that we, as humans, have created in our societies.

iii. Knowledge-Based Inequality

At the conclusion of this literature review, it becomes readily apparent that perhaps a third category of literature should be considered in the future: gender-based barriers to acquiring knowledge about how AI works. A scholar of media and communications, Massimo Ragnedda

¹¹⁷ Sinead O'Connor & Helen Liu, *Gender Bias Perpetuation and Mitigation in AI Technologies: Challenges and Opportunities*, *AI & SOC.* 1, 5–6 (2023), <https://doi.org/10.1007/s00146-023-01675-4>.

¹¹⁸ *Id.* at 5.

¹¹⁹ Sun et al., *supra* note 111, at 1.

¹²⁰ Vijayarasa & Bello y Villarino, *supra* note 8, at 196; Christoph Lutz, *Digital Inequalities in the Age of Artificial Intelligence and Big Data*, 1 *HUM. BEHAV. & EMERGING TECH.* 141, 142 (2019); Mike Zajko, *Artificial Intelligence, Algorithms, and Social Inequality: Sociological Contributions to Contemporary Debates*, 16 *SOCIO. COMPASS* (2022); Padmashree Gehl Sampath, *Governing Artificial Intelligence in an Age of Inequality*, 12 *GLOB. POL'Y* 21, 25 (2021); Katie Miller, *A Matter of Perspective: Discrimination, Bias, and Inequality, in AI*, in *LEGAL REGULATIONS, IMPLICATIONS, AND ISSUES SURROUNDING DIGITAL DATA* 182 (Margaret Jackson & Marita Shelly eds., 2020), <https://www.igi-global.com/chapter/a-matter-of-perspective/>; Mihail Caradaica, *Artificial Intelligence and Inequality in European Union*, 14 *EUROPOLITY – CONTINUITY AND CHANGE IN EUR. GOVERNANCE* 5, 5 (2020).

has pointed to such a third category of harm: the different levels of understanding that individuals have concerning how AI works and how algorithms influence our everyday lives.¹²¹ Women are under-represented in the AI industry. According to Stanford University's Human-Centered Artificial Intelligence 2022 AI Index, in 2020, women composed only around 20 percent of new doctoral students (PhDs) in AI and computing science in North American universities.¹²² This statistic draws from the Computing Research Association's database of more than 200 North American universities and the Association's efforts to collect data that also considers how men and women climb up and remain in the research pipeline.¹²³

The Computing Research Association's most recent data also shows that more than 31,000 undergraduates completed computer science degrees in 2020. That is an 11.6 percent increase from the number in 2019.¹²⁴ The field is evidently growing, yet women are not sufficiently present.

It is also worth noting that the gains of AI have been largely witnessed in countries in the global North. North America and China have enjoyed the greatest economic gains from AI.¹²⁵ Most countries leading the regulatory race – the US, EU, the UK and Canada – are also in the global North, alongside technologically-advanced and tech-producing economies, such as China, Japan and Singapore.¹²⁶ By contrast, thirty-three per cent of the world is not online (2.7 billion people);¹²⁷ global South citizens are vulnerable to data harvesting; and few nations in South are ready to deploy AI, for example, to advance public services. The World Economic Forum and UNESCO note the

¹²¹ Massimo Ragnedda, *New Digital Inequalities. Algorithms Divide*, in ENHANCING DIGITAL EQUITY: CONNECTING THE DIGITAL UNDERCLASS 61 (Massimo Ragnedda ed., 2020), https://doi.org/10.1007/978-3-030-49079-9_4 (last visited June 12, 2023).

¹²² DANIEL ZHANG ET AL., STAN. U. HUMAN-CENTERED A.I., THE AI INDEX 2022 ANNUAL REPORT 169 (2022), https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf.

¹²³ *Promoting Diversity in Computing through Evaluation and Research*, COMPUTING RSCH. ASS'N, <https://cra.org/cerp/> (last visited Feb. 2, 2023).

¹²⁴ ZHANG ET AL., *supra* note 122, at 165.

¹²⁵ World Economic Forum, *The 'AI Divide' Between the Global North and Global South*, WORLD ECON. F. (2023), <https://www.weforum.org/agenda/2023/01/davos23-ai-divide-global-north-global-south/> (last visited Nov. 20, 2023).

¹²⁶ BELLO Y VILLARINO ET AL., *supra* note 18.

¹²⁷ UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION, MEASURING DIGITAL DEVELOPMENT: FACTS AND FIGURES 2022 iii (2022), https://www.itu.int/hub/publication/d-ind-ict_mdd-2022/ (last visited Nov. 20, 2023).

widespread gaps between the North and South in access to AI's technical expertise and the data and infrastructure needed to support AI.¹²⁸

As a result, I consider AI literacy, or by contrast, AI illiteracy, a fundamental women's rights concern, given the over-representation of men in AI's design, deployment, and use, and one that has clear implications for global North-South inequalities as well. I return to this third category—knowledge and participation barriers—in Part III(C)iii and in the conclusion of this article.

III. THE GLOBAL LANDSCAPE

A. *Method and Approach: Existing and Emerging AI Legislation*

Most regulators acknowledge that regulation needs to be targeted, proportional, and in the case of AI, not stifle innovation. AI is regulated by both general laws (laws that apply across multiple industries) as well as sector-specific regulations. For instance, in many countries, general laws may be relevant to the governance of AI. This might include data protection and privacy law, consumer law, competition law, copyright law, corporations' law, online safety, discrimination law, administrative law, criminal law, and the common law of tort and contract.¹²⁹ Sectors may then have specific regulations that may affect AI's design, deployment and use, including therapeutic goods, food, motor vehicles, airline safety and financial articles.¹³⁰

Yet we are at a pivotal juncture in AI's increasing use in society where laws are needed to offer frameworks for how to mitigate risk. Guidance must be provided on what laws apply and how policy goals can be achieved.¹³¹ AI in almost all instances will be regulated under multiple pieces of legislation, increasing the possibility of duplication. Duplication becomes problematic when there are conflicts between regulatory systems and if the industry feels compliance is overburdensome.¹³² Within this large body of debate, this article focuses on AI-specific regulation, that is, laws enacted to respond to AI's concrete regulatory needs. In turn, this discussion does not cover certain issues, such as product liability, or the potential for AI's deployment to result in job

¹²⁸ World Economic Forum, *supra* note 125; WEST ET AL., *supra* note 3, at iii.

¹²⁹ SAFE AND RESPONSIBLE AI, *supra* note 55, at 10.

¹³⁰ *Id.*

¹³¹ *Id.*

¹³² *Id.* at 13.

replacement, and the consequent legislation related to employment protection.¹³³

In just the last few years, we have moved away from what was a previously largely unregulated landscape, described by Santosh and Wall as a “metaphorical Wild West.”¹³⁴ China, the US, the EU, and the UK have all invested heavily in the development of their AI sectors. Such investments have major implications for regulatory debates. They may, for instance, propel forward legislative proposals, but alternatively, such investments in innovation may be the very reason why the progress once made in enacting AI-related legislation has stalled. In the words of an Australia-based team of legal and AI scholars, “incentives, standards and hard regulation are intertwined with geopolitical, technological and value-driven interests.”¹³⁵

In this respect, it is important to acknowledge that globally, some nations appear to be favoring the setting of voluntary industry standards by experts over legislative protections.¹³⁶ Singapore and the US federally would fall within this category,¹³⁷ although important shifts have been made in recent times under the Biden administration in the US as discussed further below. Negotiated among technical experts “behind-the-scene,” voluntary industry standards risk obscurity for those individuals who may be most impacted. Nonetheless, legal analysts have been critiqued for ignoring such softer standard-setting approaches.¹³⁸

This section presents my analysis of all existing or emerging (as in, under review) AI-specific regulations across the globe. However, my primary filter has been the law’s impact on women and gendered harms. Here, I explore nations favoring voluntary standard setting and jurisdictions leaning towards tighter binding regulations, such as the EU, Canada, and Brazil.

My goal in the sections that follow is not only to provide readers with a comprehensive understanding of the state of regulation at the time of print but also to analyze the extent to which a gendered perspective has been incorporated in these regulatory approaches to exploit the potential to legislate AI with gender diversity in mind from the outset.

¹³³ Antonio Aloisi, *Regulating Algorithmic Management at Work in the European Union: Data Protection, Non-Discrimination and Collective Rights*, 40 INT’L J. COMPAR. LAB. L. & INDUS. RELS. (forthcoming 2024).

¹³⁴ Santosh & Wall, *supra* note 5, at 8.

¹³⁵ BELLO Y VILLARINO ET AL., *supra* note 18, at 3.

¹³⁶ José-Miguel Bello y Villarino, *Global Standard-Setting for Artificial Intelligence: Para-regulating International Law for AI?* 41 AUSTL. Y.B. INT’L L. ONLINE 157 (2023).

¹³⁷ CASTETS-RENARD, *supra* note 32, at 387.

¹³⁸ BELLO Y VILLARINO ET AL., *supra* note 18, at 12.

B. Weak Treatment of Gender-Based Harms

i. The United States of America

In October 2023, US President Biden shifted the landscape with an Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, what one AI expert described as “catapulting the US to the front of conversations about regulating AI.”¹³⁹ The Executive Order covers eight areas that include “equity and civil rights” and an intolerance of the “use of AI to disadvantage those who are already too often denied equal opportunity and justice.”¹⁴⁰ While congressional action will still be required to enshrine legislative protections in law,¹⁴¹ it is considered a sweeping and important step to protect people’s rights vis-à-vis the use of AI by the federal government, which has immediate application.¹⁴²

Concretely, there is a clear shift away from voluntariness, with the EO directing federal agencies to enforce civil rights protections to challenge algorithmic discrimination.¹⁴³ The US Federal Government is also seeking to model and promote responsible AI in its own practices.¹⁴⁴ The US is shifting towards an impact assessment model, “impact assessments” requiring agencies to mitigate risk before AI-driven systems are put to use. This includes an evaluation of the intended purpose of AI-driven technologies and their expected benefit; potential risks to a broad range of stakeholder groups; and quality and appropriateness of the data the AI model is built from.¹⁴⁵ Some requirements are yet to be detailed – the reporting requirements and

¹³⁹ Toby Walsh, *The US Just Issued the World’s Strongest Action yet on Regulating AI. Here’s What to Expect*, THE CONVERSATION (2023), <http://theconversation.com/the-us-just-issued-the-worlds-strongest-action-yet-on-regulating-ai-heres-what-to-expect-216729> (last visited Nov. 21, 2023).

¹⁴⁰ Joseph Biden, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, THE WHITE HOUSE 2(d) (2023), <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/> (last visited Mar 14, 2022).

¹⁴¹ Sorelle Friedler et al., *How the AI Executive Order and OMB Memo Introduce Accountability for Artificial Intelligence*, BROOKINGS (Nov. 16, 2023), <https://www.brookings.edu/articles/how-the-ai-executive-order-and-omb-memo-introduce-accountability-for-artificial-intelligence/> (last visited Nov 21, 2023).

¹⁴² *Id.*

¹⁴³ Biden, *supra* note 140, at 2(e).

¹⁴⁴ *Id.* at 2(h).

¹⁴⁵ *Id.* at 2(a).

whether existing loopholes and under-reporting will be avoided¹⁴⁶ – but the overall positive reception seems justified.

The Executive Order has created notable changes in the use of AI by the public sector in the US, although as will be seen, there remains a strong reliance standards. The significance of these changes becomes evident when one considers what came before it. On 1 January 2021, the *National Artificial Intelligence Initiative Act* became law in the US.¹⁴⁷ Underpinning the Act is the view that the US will lead the world in the “development and use of trustworthy artificial intelligence systems in the public and private sectors.”¹⁴⁸ Yet much of what was presented in the Executive Order issued by former US President Trump emphasized the need for American citizens to trust and have confidence in the use of AI and its decisions that would nonetheless remain weakly regulated.¹⁴⁹ The US’s approach at the time was centered on developing technical standards and promoting a harmonious technological world order based on these standards.¹⁵⁰ Achieving trust with this voluntary approach is dependent both on the robustness of the standards and on the belief of consumers that the standards are being—voluntarily—adhered to.

Even if the emphasis is on the creation of standards, not all is lost in gendered terms. Standards mean that “a practitioner can talk the same language as a regulator and both can talk the same language as a technical expert.”¹⁵¹ It therefore offers a starting point, although far away from where a gender-responsive approach may be. Gendered fairness can also be a pivotal part of what US citizens understand by “trust” in AI. That is, if AI’s designers and deployers want to foster an environment in which there is widespread trust in the system, one could argue that such trust can only be fostered on systems that have taken gendered harms into account. Yet this line of thinking, particularly in such a voluntary standards-based context, requires gender expertise from within in order to trigger and enable such considerations.

A strong interventionist approach at the state level has also been attempted in the US. In California, what has been described as “sweeping

¹⁴⁶ Sorelle Friedler et al., *supra* note 141.

¹⁴⁷ William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021, Pub. L. No. 116-283, 134 Stat. 3338, 4524–25 (2021).

¹⁴⁸ *Id.* at 5101(a)(2).

¹⁴⁹ Exec. Order No. 13,859, 84 Fed. Reg. 3967 (Feb. 14, 2019).

¹⁵⁰ NAT’L SEC. COMM’N ON A.I., FINAL REPORT 1 (2021), <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>.

¹⁵¹ STEFANO NATIVI & SARAH DE NIGRIS, EUR. COMM’N, JOINT RSCH. CTR., AI WATCH: AI STANDARDISATION LANDSCAPE: STATE OF PLAY AND LINK TO THE EC PROPOSAL FOR AN AI REGULATORY FRAMEWORK 10 (2021), <https://data.europa.eu/doi/10.2760/376602> (last visited July 27, 2022).

legislation” in the form of Civil Code Section 1708.86, established a private cause of action for the creation and intentional disclosure of sexually explicit material in the form of the deepfakes described above. While narrow in subject matter, it provides wide-reaching protections for women as a group most likely to be harmed by pornographic deepfakes. The momentum, however, appears to have been short-lived. New York State followed suit in 2020, enacting a law to addresses synthetic or digitally manipulated media, but legislation in Maryland, Massachusetts, Pennsylvania, and Florida stalled at the bill stage.¹⁵² Beyond these interventions, the US appears to offer little by way of exemplars for regulating AI with gendered concerns in mind.

ii. Japan

Japan currently relies on existing laws to address legal issues that are emerging with the design, deployment, and use of AI-driven technologies, with no current plans to modify such laws.¹⁵³ The Government’s overarching approach is one of trust in the systems’ designers to do what is in the best interests of Japanese society. Japan wants to create an AI-ready society¹⁵⁴ in which AI is perceived as a largely positive development and Japanese policy is driven by a concern to not suppress technological advancements through an overestimation of the risks associated with AI.¹⁵⁵ Its emphasis, therefore, is on non-binding documents. This includes the Social Principles of Human-Centric AI, which flags discrimination as a concern.

Japanese scholar Souichirou Kozuka argues that there are clear similarities between the Japanese and EU approaches to the regulation of AI and therefore, there is promise that both can be harmonized.¹⁵⁶ Yet we cannot ignore the fact that the Japanese approach is built on a trust in the designers and deployers of AI-driven technologies to “do the right thing” and an implicit assumption that all are collectively driven towards common societal good. Not all scholars agree that this is the right

¹⁵² Samuel Hodge, *Don’t Always Believe What You See: Shallowfake and Deepfake Media Has Altered the Perception of Reality*, 50 HOFSTRA L. REV. 51, 62–63 (2021).

¹⁵³ Tomoko Nambu, *Legal Regulations and Public Policies for Next-Generation Robots in Japan*, 31 A.I. & SOC. 483, 485 (2016).

¹⁵⁴ Souichirou Kozuka, *A Governance Framework for the Development and Use of Artificial Intelligence: Lessons from the Comparison of Japanese and European Initiatives*, 24 UNIF. L. REV. 315, 321 (2019).

¹⁵⁵ HIROKI HABUKA, JAPAN’S APPROACH TO AI REGULATION AND ITS IMPACT ON THE 2023 G7 PRESIDENCY 1 (Ctr. for Strategic & Int’l Stud. ed., 2023), <https://www.csis.org/analysis/japans-approach-ai-regulation-and-its-impact-2023-g7-presidency>.

¹⁵⁶ Kozuka, *supra* note 154, at 328.

approach, some arguing that the absence of a clear definition of AI in any law, regulation or standards document is problematic.¹⁵⁷ From a gender and discrimination perspective, Japan's approach simply appears as a soft law approach that is too soft. If Japan's standing in generative AI and large language models are to accelerate as Japan hopes, such an unregulated environment may be overwhelmed by the absence of structures to respond adequately and in a timely manner to claims of emerging harm. Japan risks a perpetuation of traditional gender roles if Japanese AI is trained without consideration to the gender biases evident in existing historical data in Japan.¹⁵⁸

iii. China

China has set an ambitious target of becoming an, or perhaps put more accurately, *the* AI superpower by 2030.¹⁵⁹ Yet, as some scholars suggest, China's ambitions to be an AI norm-setting power is likely to remain unrealized, particularly given AI's global governance architecture is one that will be largely built on "Westernised democratic values and a liberal global order."¹⁶⁰

China is a nation that will see notable impacts from the automatization of its workforce, with an estimated 100 million Chinese members of the workforce expected to experience employment-related changes by 2030.¹⁶¹ This will obviously have a notable impact on women in the Chinese workforce. In terms of regulatory design, China has been targeting specific uses of AI rather than AI as a whole, as we have seen in the emerging examples discussed immediately below. Yet this should not be taken as a suggestion of an overarching gap; China has a comprehensive body of AI-related laws. All such laws, regulations, and guidelines in China, including at the provincial and local levels, are aligned with the Government's 2017 *A Next Generation Artificial*

¹⁵⁷ 史生新保, *The Principal Japanese AI and Robot Law. Strategy and Research toward Establishing Basic Principles*, 3 情報法制研究 44, 47 (2018).

¹⁵⁸ Ming Hui, "An Awkward Photo-op": *Japan's Persistent Gender Problem*, THE INTERPRETER, THE LOWY INSTITUTE (Jan. 10, 2024), <https://www.lowyinstitute.org/the-interpreter/awkward-photo-op-japan-s-persistent-gender-problem> (last visited Jan 11, 2024).

¹⁵⁹ Jing Cheng & Jinghan Zeng, *Shaping AI's Future? China in Global AI Governance*, 32 J. CONTEMP. CHINA 794, 795 (2022).

¹⁶⁰ *Id.* at 806.

¹⁶¹ *Id.* at 799.

Intelligence Development Plan.¹⁶² This Development Plan acknowledges that AI is a disruptive technology with widespread influence that may cause a transformation to employment structures, impact legal and social theories and potentially violate personal privacy.¹⁶³ The answer, according to the Plan, lies in minimizing risk, and “ensur[ing] the safe, reliable, and controllable development of AI.”¹⁶⁴ (translation)

China is currently considering more comprehensive and general AI regulations and already has quite a spate of rules, regulations, and codes in place. This includes the Chinese Civil Code, which references privacy and personal information; a guideline for self-assessment of data collection through mobile apps; a white paper on the same issue; a guide to AI ethics in 2020 and a further guidance released in 2021; and a Whitepaper on Trustworthy AI produced that same year.¹⁶⁵

Nonetheless, here I focus on the existence or absence of specific regulations to address gendered harms. China has been regulating AI by targeting precise applications of AI, such as through the Internet Information Service Algorithmic Recommendation Management Provisions. Introduced on 1 March 2022, the Provisions focus on the use and impact of algorithmic recommendation systems and create transparency obligations which entail, among others, user notifications regarding the criteria for recommendations and clear indicators of algorithmically generated or synthetically-developed information; mechanisms for manual intervention that can override automated decision-making; and regulation of the use of the information, ranking, and presentation of recommendations that result from AI-driven technologies.¹⁶⁶

When it comes to gendered harms, on 28 May 2019, the Beijing Academy of Artificial Intelligence (BAAI) released the “Beijing AI Principles,” an outline to guide the research and development, implementation, and governance of AI. These are neither a form of regulation nor government-issued binding principles. They recommend the integration of ethical principles into the entire lifecycle of AI. This

¹⁶² State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan, Document (China), *translated in China's New Generation Artificial Intelligence Development Plan* (Rogier Creemers et al., 2017), NEW AMERICA, <https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>.

¹⁶³ *Id.* at 3.

¹⁶⁴ *Id.*

¹⁶⁵ Cheng & Zeng, *supra* note 159, at 802–03.

¹⁶⁶ Luca Belli et al., *AI Regulation in Brazil: Advancements, Flows, and Need to Learn from the Data Protection Experience*, 48 COMPUT. L. & SEC. REV. 105767, 105770 (2023).

includes fairness and justice and avoiding prejudice and discrimination.¹⁶⁷ However, their generic and broad-brush approaches to discrimination and the non-binding nature make them, at best, weak. There is no gender-specific provision or any provisions that acknowledge how minority groups may be at greater risk. Rather, the Beijing AI Principles simply require designers and deploys to, in general, avoid the use of AI to “harm human beings.”¹⁶⁸

iv. Australia

Australia adopted a voluntary and aspirational set of AI principles relatively early—an Artificial Intelligence Ethics Framework—under the Department of Industry, Science and Resources published in November 2019. They include an end goal of AI systems that foster societal well-being, fairness and contestability, that is, having a timely process to allow people to challenge the use or outcomes of the AI system as well as accountability.¹⁶⁹ Importantly, it perceives of an ideal in which an impacted person or community group, including on behalf of an affected environment, has time to challenge the use or outcomes of an AI system. It also seeks to enhance accountability, but again, in a non-binding sense, for example, through checklists.¹⁷⁰

Some of the areas of regulation in Australia that may prove most relevant to AI from a gender perspective do not presently contain specific provisions on AI. For instance, Australia’s Office of the eSafety Commissioner was established in 2015 to coordinate and lead the online safety efforts of government, industry, and the not-for-profit community in Australia. It is to date a relatively under-analyzed government intervention.¹⁷¹ Its remit was expanded from women’s safety to online

¹⁶⁷ Stephen Chan, *What You Need to Know about China’s AI Ethics Rules*, TECHBEACON, <https://techbeacon.com/enterprise-it/what-you-need-know-about-chinas-ai-ethics-rules>.

¹⁶⁸ Beijing Acad. A.I., *Beijing AI Principles*, 43 DATENSCHUTZ DATENSICH 656 (2019), <https://link-springer-com.ezproxy.library.sydney.edu.au/content/pdf/10.1007/s11623-019-1183-6.pdf>.

¹⁶⁹ *Australia’s AI Ethics Principles: Australia’s Artificial Intelligence Ethics Framework*, AUSTRAL. GOV’T, DEP’T INDUS., SCI. & RES. <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>.

¹⁷⁰ *Id.*; see also SAFE AND RESPONSIBLE AI, *supra* note 55, at 14.

¹⁷¹ See, for example, some analysis conducted of the e-Safety approach by Kerryann Walsh et al., *Best Practice Framework for Online Safety Education: Results from a Rapid Review of the International Literature, Expert Review, and Stakeholder Consultation*, 33 INT’L J. CHILD-COMPUT. INTERACTION 100474 (2022); CARMINA

safety for all Australians, with particular attention to women, children, and elderly Australians.¹⁷² While the Commissioner's strategy for the period 2022-2025 speaks to the potential harms caused by algorithmic biases,¹⁷³ the actual Online Safety Act 2021 itself does not reference AI. It empowers the Commissioner to issue a removal notice for abusive material online,¹⁷⁴ such as in response to disclosure without consent. Yet, the Act in its current form does not legislate directly to address deepfake videos or images.

Discussions were underway at the time of print about an expansion in the remit of the e-Commissioner to address online safety created or exacerbated by AI. A position statement released in August 2023 on generative AI acknowledges both unintentional and intentional harm, such as, the generation of child sexual abuse materials or blackmailing an individual using generated sexual content that appears to show a real adult.¹⁷⁵ There are also relatively legitimate expectations that a public consultation from June-July 2023 will allow Australia to benefit from the emerging promising practice discussed in the following section. Industry standards in Australia, such as those requiring cloud-based storage services like Apple iCloud, Google Drive and Microsoft OneDrive, as well as messaging services like WhatsApp, to rid their platforms services of unlawful content, such as child sexual abuse (including AI-generated ones), are in constant development.¹⁷⁶ If strengthened, the e-safety portfolio has the potential to rise to the level of the "trusted flagger" model discussed below.

However, to date, there exists no AI-specific legislative mechanism enacted to address gender discrimination in AI systems, creating a dependency on existing discrimination laws in areas such as racial discrimination, disability discrimination, age discrimination, and fair work.¹⁷⁷ Such a dependency requires a test case to determine the

MEOLA, *Helping Aussie Women Online: A Discourse Analysis of the Australian ESafety Commissioner Website*, in SOCIAL MEDIA IN LEGAL PRACTICE 130 (Vijay Bhatia & Girolamo Tessuto eds., 2020).

¹⁷² MEOLA, *supra* note 171, at 132.

¹⁷³ *ESafety Strategy 2022-25*, AUSTL. GOV'T: E-SAFETY COMM'R, <https://www.esafety.gov.au/about-us/who-we-are/strategy>.

¹⁷⁴ Online Safety Act 2021, No. 76, 2021, at 66–77 (2021) (Austl.), <http://www.legislation.gov.au/Details/C2021A00076>.

¹⁷⁵ *Generative AI: Position Statement*, AUSTL. GOV'T: E-SAFETY COMM'R, <https://www.esafety.gov.au/industry/tech-trends-and-challenges/generative-ai>.

¹⁷⁶ *Industry Codes and Standards*, AUSTL. GOV'T: E-SAFETY COMM'R (2023), <https://www.esafety.gov.au/industry/codes> (last visited Dec. 22, 2023).

¹⁷⁷ Finn Lattimore et al., *Austl. Hum. Rts. Comm'n, Using Artificial Intelligence to Make Decisions: Addressing the Problem of Algorithmic Bias: Technical Paper*, at 56

extent to which existing laws are up to the task of adequately protecting individuals and groups from AI-related harm.

C. *Emerging Promising Practice*

Given the above survey of legislative approaches, it may be natural to doubt the level of commitment to address, in law, gender-based harms that have emerged from or are replicated and magnified by AI. Gender appears to have been inadequately canvassed in legislative debates, leaving a reliance on either weak protection in AI-specific laws or the use of non-AI-specific legislation. Now, I turn to jurisdictions that offer more promise.

i. The European Union

In this section I consider two pieces of legislation emerging from the EU, one in force at the time of print and the second in its last steps towards legislative enactment after the political agreement formed on 9 December 2023.¹⁷⁸ I classify both as promising examples, although they are certainly not without their critiques.¹⁷⁹ The EU Digital Services Act has been in force since 2022. In response to the issue of deep fakes and harmful AI-driven online content, the EU Digital Services Act offers the figure of the trusted flagger, an officially recognized expert that acts as an independent industry whistleblower.¹⁸⁰ Given the particular harms facing women and girls from deepfakes, the figure of a trusted whistleblower offers a gendered response to an evident gender-based harm from AI.

Organizations are appointed in the role of “trusted flagger” if they meet predefined criteria, such as a specific expertise in illegal content, independence from the platforms, and based on the integrity of its activities. Government funding is received for the role. Trusted Flaggers

(2020),

https://humanrights.gov.au/sites/default/files/document/publication/final_version_technical_paper_addressing_the_problem_of_algorithmic_bias.pdf.

¹⁷⁸ *Artificial Intelligence Act: Deal on Comprehensive Rules for Trustworthy AI*, EUR. PARLIAMENT (Dec. 9, 2023), <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai> (last visited Jan. 11, 2024).

¹⁷⁹ Emma Carmel & Regine Paul, *Peace and Prosperity for the Digital Age? The Colonial Political Economy of European AI Governance*, 41 IEEE TECH. & SOC. MAG. 94, 94 (2022).

¹⁸⁰ Regulation 2022/2065, of the European Parliament and of the Council of 19 Oct. 2022, Digital Services Act, 2022 O.J. (L 277) 1, 61–62 (2022), <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>.

are required to publish reports at least once a year on their activities. Once content is flagged, the responsible entity—a platform such as Meta, or a cloud server such as Dropbox—would be required to remove the content. If the content is not removed, the entity must explain why.¹⁸¹

The second major legislative initiative from the EU that presents itself as an emerging good promise reached a provisional agreement in December 2023. Back in April 2021, the European Commission presented a proposal for a European Union Act on AI (EU AI Proposal), sometimes misleadingly described in short form as an “EU AI Act.” It is not yet law. Nonetheless, the provisional agreement requires its entry into force within two years, with the exception of some specific provisions. The EU AI Proposal, once enacted, would require the expansion of at least five other legal instruments.¹⁸² Moreover, by its nature, AI—characterized by opacity, complexity, dependency on data, and autonomous behavior¹⁸³—will inevitably impact a number of rights currently protected by the EU Charter of Fundamental Rights, including non-discrimination (Article 21) and equality between men and women (Article 23).

Since its first legislative steps, the EU AI Proposal has acknowledged the risk that AI systems “may perpetuate historical patterns of discrimination,” including against women.¹⁸⁴ The EU AI Act, as currently perceived, will identify three types of risks in three main ways that are distinct from the three categories of gendered harm noted above. Risk is understood as involving different degrees. First, it will ban those AI applications that demonstrate an “unacceptable risk” for “contravening Union values, for instance, by violating fundamental rights.”¹⁸⁵ Real-time, remote biometric identification systems, for example, would be banned.¹⁸⁶ Second, “high-risk” applications such as a CV-scanning tool that ranks job applicants, with the risk of applying the

¹⁸¹ *Id.* at Art. 6(1)(b).

¹⁸² *See id.* at 98. This would include the draft Digital Services Act (with provisions on recommenders and research data access); the draft Digital Markets Act (with provisions on AI-relevant hardware, operating systems and software distribution); the draft Machinery Regulation (revising the Machinery Directive in relation to AI, health and safety, and machinery); an announced product liability revision relating to AI; and a draft Data Governance Act (concerning data sharing frameworks).

¹⁸³ European Comm'n, *Proposal for a Regulation of the European Parliament and the Council: Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, 2021/0106 (COD) 3.5 (2021), <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>.

¹⁸⁴ *Id.* at 36.

¹⁸⁵ *Id.* at 5.2.2.

¹⁸⁶ *Id.* at 5.2.2.

inherent gender biases also discussed in Part II, would be subject to specific legal requirements, “a promising acknowledgement of the profound effects that AI-powered HR tools could and will have on employment and the workforce.”¹⁸⁷ AI providers would be obliged to provide meaningful information about their systems and what assessments of those systems have been conducted to ensure conformity with Union values. This includes, for example, risk management measures where identified risks are eliminated or reduced.¹⁸⁸ Training, validation, and testing of the AI are subject to particular requirements.¹⁸⁹ Record-keeping, transparency, provision of information to users, and human oversight are all discussed in the EU AI Proposal.¹⁹⁰

Lastly, applications not explicitly banned or listed as high-risk are largely left unregulated, barring a few transparency requirements, such as informing users when they are seeing deepfakes or communicating with a bot. The European Parliament’s proposed amendments to the initial EU Act AI Proposal were adopted in June 2023. The Parliament suggested text that made clear that the determination of an AI system as high-risk should take account of the fundamental rights protected by the European Charter.¹⁹¹ That proposed language went on further to highlight the need to take specific account of the rights of children and on the right to health, safety of people and environmental protection.¹⁹² Further, Parliament’s EU AI Proposal proposed a definition of “deep fake” that would be two-pronged. A deepfake would entail either manipulated or synthetic audio, image or video content that falsely appears to be authentic, or interestingly, falsely appears to be truthful.¹⁹³ In other words, an AI-created film will not be a deepfake if it is known to be a fake but *depicts* what is in reality the truth. The issue targeted in the proposed law is the generation of fake videos with a goal of incorrectly implying that they are true.

What makes the European Parliament’s proposed amendments to the initial EU Act AI Proposal stand-out is the response to the third type

¹⁸⁷ Drage & Mackereth, *supra* note 81, at 92.

¹⁸⁸ European Comm’n, *supra* note 183, at 2, Art. 9(4).

¹⁸⁹ *Id.* at 2, Art. 10.

¹⁹⁰ *Id.* at 2, Arts. 11–14.

¹⁹¹ Amendments Adopted by the European Parliament on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, Eur. Parl. Doc. P9_TA(2023), 0236 Amendment 56 (2023), https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html (last visited Aug. 10, 2023).

¹⁹² *Id.*

¹⁹³ *Id.* at 203.

of gendered-harm: gaps in knowledge. Several proposed amendments to the EU AI Proposal directly address the issue of AI literacy. It is rare for such a provision to be enacted into law—as opposed to public policy—but nonetheless, the Parliament’s amendments called for all Member States to promote measures for the development of a sufficient level of AI literacy, across sectors and to take into account the different needs of groups of providers, deployers, and affected persons concerned, including through education and training, skilling and reskilling programs, while ensuring proper gender and age balance, in view of allowing a democratic control of AI systems.¹⁹⁴ The onus is on providers to achieve such AI literacy, including teaching basic notions and skills about AI systems and their functioning, which includes “the different types of products and uses, their risks and benefits.”¹⁹⁵ Indeed, a sufficient level of AI literacy would be a requirement for compliance with the Act when passed.

Moreover, the inequalities with regard to who has access to decision-making roles is addressed head-on. The proposed text requires AI systems to be developed and used in a way that “includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by [European] Union or national law.”¹⁹⁶ Overall, this provision acknowledges that society’s most vulnerable may be those most affected by AI, and thereby creates an onus on providers to shrink current knowledge gaps.

Once passed within the two-year deadline, that is, before December 2025, the EU AI Proposal would be the foremost example of an AI law with large territorial scope, avoiding the risk of a patchwork of legislation across member states.¹⁹⁷ Part of the driver behind the EU AI Proposal is that a host of laws operating at the periphery of AI’s use are inadequate in minimizing harm. Existing laws are perceived as failing to fully meet the legal challenges of AI. For instance, AI’s use in data surveillance is increasingly common. Yet most people’s understandings of AI’s capabilities are too limited to meet the conditions of “informed consent.”¹⁹⁸ Moreover, data protection laws are driven at protecting an individual’s “personal information.” By contrast, AI such as machine-learning uses all sorts of databases to *infer* patterns and do inferential

¹⁹⁴ *Id.* at Amendment 214(1).

¹⁹⁵ *Id.* at Amendment 214(3).

¹⁹⁶ *Id.* at Amendment 213(e).

¹⁹⁷ CASTETS-RENARD, *supra* note 32, at 382.

¹⁹⁸ COOMBS & ABRAHA, *supra* note 8, at 7.

analysis that often ends up targeting sensitive characteristics such as gender.¹⁹⁹ Data protection laws for the purposes of AI fail to address the implications of the uses of non-personal data in a way that may affect the lives of individual people by drawing on gendered assumptions and conclusions from large datasets.

Laws in the fields of non-discrimination also appear insufficient and may fail to adequately address the problems generated by AI systems.²⁰⁰ For instance, the EU Directorate-General for Justice and Consumers is concerned about the impact of the exclusion of the media, advertising, and communication from the EU Directive on non-discrimination between men and women in the supply of goods and services.²⁰¹ Think here of the representative harms resulting from a Google search that returns mostly female images for a search of “nurse” and male pictures for a search of “doctor,” or advertisements for STEM careers that display more often to men than women, despite gender neutral target settings.²⁰² According to the EU Directorate-General for Justice and Consumers, neither would violate the directive, risking the reinforcement of structural inequality.²⁰³ In the case of the EU, protections provided elsewhere (e.g. Article 14(1)(a) of the Gender Recast Directive and Article 3(1)(a) of the Racial Equality Directive and the Framework Directive) help fill the gap, alongside protections offered from case law before the Court of Justice of the European Union.²⁰⁴

Nonetheless, it is relatively evident why AI-specific regulations in the form of the approved EU AI Proposal will fill gaps in existing laws, both big and small. The examples offered above illustrate that existing regulations alone may be limited in capturing instances of gender-based discrimination that AI reinforces. By contrast, the EU AI Act will offer good practice: there are numerous references to non-discrimination and in line with its emphasis on fundamental rights, a reference to “equality between women and men”²⁰⁵ and to “gender equality.”²⁰⁶ Yet it is unclear whether these frameworks alone will be enough to adequately minimize,

¹⁹⁹ *Id.*

²⁰⁰ DIRECTORATE-GENERAL FOR JUSTICE AND CONSUMERS (EUROPEAN COMMISSION) ET AL., ALGORITHMIC DISCRIMINATION IN EUROPE: CHALLENGES AND OPPORTUNITIES FOR GENDER EQUALITY AND NON DISCRIMINATION LAW 27 (2021), <https://data.europa.eu/doi/10.2838/544956> (last visited Aug 31, 2022).

²⁰¹ Council Directive 2004/113, art. 13, 2004 O.J. (L 338) 37, 38 (EU).

²⁰² DIRECTORATE-GENERAL FOR JUSTICE AND CONSUMERS (EUROPEAN COMMISSION) ET AL., *supra* note 200, at 59.

²⁰³ *Id.* at 58.

²⁰⁴ *Id.*

²⁰⁵ European Comm'n, *supra* note 183, at Explanatory Memorandum, Sec. 3.5.

²⁰⁶ *Id.* at Explanatory Memorandum, Sec. 1.2.

let alone eradicate the type of allocative and representative harms discussed above.

ii. Canada

Canada, while perhaps a smaller player, has found a place for itself in this debate and in my view, offers some of the most promising potential for global good practice on legislation for AI from a gendered standpoint.²⁰⁷ Canada also provides legislators a comparative example of how to regulate AI in the complex contexts that entail both Federal and Provincial law.²⁰⁸ As early as 2019, the Canadians introduced the Treasury Board Directive on Automated Decision-Making, a mandatory policy instrument which has applied to almost all federal government institutions since early 2019.²⁰⁹ It emphasizes impact assessments and transparency, including the likely impacts of AI on freedom, health, the economy, and environment.²¹⁰ More importantly, its concern lies with testing AI systems before deployment, particularly for impartiality and standardization.²¹¹

The standout feature of the Canadian Directive is the Gender-Based Analysis Plus, a quality reassurance requirement before launching into production of an AI-driven technology that could be used by the Canadian Federal Government.²¹² This Gender-Based Analysis Plus requirement entails compulsory testing for unintended biases. If an AI-driven technology meets the moderate, high, and very high-risk thresholds, the designers need to undertake Gender-Based Analysis Plus. This “plus” reflects going further than the gender impact assessment already required for procurement. Additional steps include an assessment of the impact of the automation on gender and/or other

²⁰⁷ Henderson et al., *supra* note 62, at 4.

²⁰⁸ FLORIAN MARTIN-BARITEAU & TERESA SCASSA, *ARTIFICIAL INTELLIGENCE AND THE LAW IN CANADA* 9 (LexisNexis Can., 2021), <https://papers.ssrn.com/abstract=3734675> (last visited June 22, 2023).

²⁰⁹ *Treasury Board Directive on Automated Decision-Making*, GOV'T CAN. (2019), <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>.

²¹⁰ *Algorithmic Impact Assessment Tool*, GOV'T CAN. (Apr. 25, 2023), <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>.

²¹¹ JOSÉ-MIGUEL BELLO Y VILLARINO, *Tale of Two Automated States: Why One-Size-Fits-All Approach to Administrative Law Reform to Accommodate AI Will Fail*, in *MONEY, POWER AND AI* 139 (Monika Zalnieriute & Zofia Bednarz eds., 2023).

²¹² *Directive on Automated Decision-Making: 6.3.6*, GOV'T CAN. (2019), <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592> (last visited June 16, 2023).

identifying factors but also naming what planned or existing measures are in place to address these identified risks in the future.²¹³

In November 2020, the Privacy Commissioner of Canada recommended a reform to the *Personal Information Protection and Electronic Documents Act* (PIPEDA). One week later, the federal government introduced Bill C-1156 to recast the Canadian federal data protection framework for the private sector.²¹⁴ Reforms were introduced in several provinces across Canada in the months that followed. The federal *Digital Charter Implementation Act*, passed in April 2023, requires the setting up of a tribunal specific to data and privacy protection²¹⁵ and probably most significantly, makes the legislation compatible with the proposed *Artificial Intelligence and Data Act* (AIDA). While it is still under debate,²¹⁶ Canada could soon be among the first nations to move to comprehensive AI-specific regulation. Importantly, the AIDA bill defines “biased output.” It seeks to *prevent* direct or indirect discrimination in the use of data, but *protects* the use of data designed to prevent, eliminate, or reduce disadvantage. In other words, it would prohibit the use of AI in instances when you can identify the types of allocative or representative harms discussed above *but* allows the use of AI for positive discrimination:

biased output means content that is generated, or a decision, recommendation or prediction that is made, by an artificial intelligence system and that adversely differentiates, directly or indirectly, and without justification, in relation to an individual on one or more of the prohibited grounds of discrimination set out in section 3 of the *Canadian Human Rights Act*, or on a combination of such prohibited grounds. It does not include content, or a decision, recommendation or prediction, the purpose and effect of which are to prevent disadvantages that are likely to be suffered by, or to

²¹³ *Gender-Based Analysis Plus (GBA Plus)*, GOV'T CAN. (Oct. 13, 2023), <https://women-gender-equality.canada.ca/en/gender-based-analysis-plus.html> (last visited June 16, 2023); Government of Canada, *supra* note 209, at Appendix C, 6.3.6.

²¹⁴ Henderson et al., *supra* note 62, at 499.

²¹⁵ House of Commons of Canada, *Bill C-27: An Act to Enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to Make Consequential and Related Amendments to Other Acts*, PARLIAMENT CAN., <https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading> (last visited Sept. 20, 2022).

²¹⁶ *Id.*

eliminate or reduce disadvantages that are suffered by, any group of individuals when those disadvantages would be based on or related to the prohibited grounds (*résultat biaisé*).²¹⁷

The terms “woman,” “women,” or “gender” do not appear explicitly in the Canadian legislation. However, two semi-hypothetical examples help explain how the provision may be implemented in practice. Take the example introduced earlier in this article concerning female bank applicants. When making a determination about credit worthiness, a bank may determine that a female applicant, particularly a woman of color or a sole head-of-household, is less likely to repay a loan within the time stipulated. A decision-support system could introduce the protected category of gender to determine that once the likelihood of loan repayments has been calculated, a certain percentage of all loan recipients must be female. Such an example meets the Canada’s AIDA definition of “a decision, recommendation or prediction, the purpose and effect of which are to prevent disadvantages that are likely to be suffered by,” in this case, women.

To offer another example, with co-author José-Miguel Bello y Villarino, I have elsewhere discussed the Spanish system VioGén, a women-centered decision support system that has helped predict the likelihood that a victim of domestic violence will become a victim of violence perpetrated by the same perpetrator.²¹⁸ In 2020, the Spanish Government compensated the family of Stefany González Escarramán who was murdered by a former partner against whom she had been denied a restraining order on the basis of an assessment generated by VioGén that determined she did not face a risk of re-victimization. That victim was an immigrant and woman of color. An immigrant may face greater challenges in accessing housing, making it harder to leave an unsafe home. Once again, the DSS may introduce the protected category of race in its algorithm, or map onto the expected output of race, in order to lower the threshold of risk to be met by an immigrant woman and/or woman of color in order to be considered at greater risk of some form of harm. The biased output would seek to prevent disadvantage against this sub-group of women, positively introducing racial bias, despite the risk of an incorrect prediction of a heightened risk of re-victimization. This example raises a notable challenge insofar as other stakeholders are impacted by the decision: at a minimum, the accused perpetrator, and

²¹⁷ *Id.* at Part 1, Sec. 5(1).

²¹⁸ Vijayarasa & Bello y Villarino, *supra* note 8, at 12–13.

possibly also other family members—including children—but shows the potential for AI to be deployed to positively discriminate to protect rights and reduce risk.

Some analysts of the EU AI Proposal suggest that, if enacted as proposed, the Act will deliver similar outcomes to the risk assessments that will result from the Canadian law if passed. While bias is not treated as fully or explicitly in the draft under discussion in the EU as in the Canadian bill, categories of discriminated against at-risk individuals may be applied at the stage of testing new AI technologies. This may be done with a view to examining for potential biases in those technologies' designs, outputs, or use. The EU approach would allow a developer to test for gender bias once an AI technology has been developed but before deployment.²¹⁹ It remains to be seen in the following two years how the EU AI Proposal will develop and whether the enacted legislation will treat gender as fully as in the Canadian experiences to date.

iii. Brazil

Brazil is one of the few jurisdictions that is not starting from scratch in terms of governing AI's design, deployment and use.²²⁰ In 2020, Brazil launched its national response to AI, opening its Brazilian Strategy for Artificial Intelligence (EBIA) to public consultation, accompanied by a parliamentary hearing on its bill on Artificial Intelligence. Moreover, the field is governed by a general data protection law.

Brazil is pursuing an AI-specific regulatory response. Brazilian scholars described earlier iterations of the bill as weak, based on “very timid efforts to the understanding of AI implications.”²²¹ Nonetheless, the nation's latest bill (Bill No. 2338/2023) aims to regulate AI as a whole, as opposed to specific applications of the technology. Moreover, some scholars and analysts consider Brazil's regulatory approach in relative alignment with the approved EU's AI Proposal. It is principles-based and categorizes AI's risks into different tiers. Still, there remain concerns that the 2023 bill contradicts—and at times waters down—existing protections already established under the general law, particularly in relation to non-discrimination and transparency.²²²

²¹⁹ European Comm'n, *supra* note 183, at 10(5).

²²⁰ Belli et al., *supra* note 166, at 10.

²²¹ *Id.* at 3.

²²² *Id.* at 11.

If passed, the proposed legislation would establish national ethical norms for responsible use of AI that include human rights, democratic values, equality, non-discrimination, plurality, and respect for labor rights. Both direct and indirect discrimination is defined in the bill to include a long list of bases of discrimination, including gender, sexual orientation, race, and socio-economic class.²²³ Discrimination and disproportionate impacts are set out in extensive detail in Article 12, including a prohibition on the implementation and use of AI systems that have a “disproportionate impact” due to personal characteristics such as geographic origin, race, color or ethnicity, gender, sexual orientation, socioeconomic class, age, disability, religion or political opinions.²²⁴ If enacted, the law would also guarantee individuals’ rights and protections related to AI, including an explanation of AI-assisted decisions, the ability to contest those decisions, and human participation in the decision-making process.²²⁵ In setting out a right to challenge AI-based decisions, affected people also have a right to request anonymization, blocking, or deletion of data that is unnecessary, excessive, or treated in violation of the legislation.²²⁶

The proposed law categorizes systems into risk tiers, with “excessive risk” systems being banned and “high risk” systems being subject to preliminary assessment and transparency obligations.²²⁷ AI agents would be required to undertake algorithmic impact assessments and the law would introduce a protective system of civil liability for providers or operators of AI systems. For instance, the supplier or operator of a “high risk” system would be held objectively liable for any resulting damages. A new regulator would monitor and enforce the law. There are, therefore, numerous strengths to the bill. While a grandiose statement, perhaps Brazil will meet the promise suggested by Brazilian scholars of law and technology that the “tropical giant has a remarkably relevant potential to become a regional leader in AI policymaking.”²²⁸

D. *Comparative Lessons*

Why look to other legal systems? In the admittedly dated but still helpful words of comparative law scholars, René David and John

²²³ RODRIGO PACHECO, PROJETO DE LEI NO. 2338, DE 2023: DISPÕE SOBRE O USO DA INTELIGÊNCIA ARTIFICIAL 4 (Senado Federal, 2023).

²²⁴ *Id.* at 12.

²²⁵ *Id.* at 5.

²²⁶ *Id.* at 9, Section 1.

²²⁷ *Id.* at 13–17.

²²⁸ Belli et al., *supra* note 166, at 13.

Brierley, “Comparative law is useful in gaining a better understanding of one’s own national law and in the work of improving it.”²²⁹ Merely cherry-picking good practice will not help us to arrive at the point of a gender-responsive AI law. Nonetheless, it appears worthwhile to extract some of the good practices from these nations and to place them on the relatively blank canvas that remains as many nations determine their regulatory approach to AI.

For one, the EU’s “trusted flagger” seems not only replicable but also very well aligned to the relatively large number of nations with roles similar to that of e-safety or online safety.²³⁰ While these roles are not without their critiques,²³¹ Australia, the UK and US have all made legislative efforts to tackle online harms such as cyber-bullying, sexual abuse and deepfake abuse, including the Office of the eSafety Commissioner in Australia, the UK’s Online Safety Bill and the US’ Violence against Women Reauthorization Act 2021,²³² and can be strengthened by learning from the EU. The binding nature of the “trusted flaggers” role, embedded in law, has received praise.²³³ In response to the critiques of the various types of “trusted flaggers” that exist—from those intended to target child abuse materials to hate speech and violations of intellectual property²³⁴—new regulations to deal with emerging technologies allow us to take lessons from these jurisdictions and adapt these models. For one, the EU’s approach has been to make the eligibility criteria for trusted flaggers explicit; there are performance reviews and oversight, both of which have been pinpointed as musts.²³⁵

Canada’s Gender-Based Analysis Plus has much to offer for the regulation of private actors but also as a guidance for the Government’s own behavior. Such gender assessments lay the ground for a preventative approach that is needed when it comes to regulating AI. Pre-emptively

²²⁹ RENÉ DAVID, *MAJOR LEGAL SYSTEMS IN THE WORLD TODAY: AN INTRODUCTION TO THE COMPARATIVE STUDY OF LAW* 3 (3rd ed. 1985).

²³⁰ *The Global Online Safety Regulators Network*, AUSTRALIAN GOVERNMENT: E-SAFETY COMMISSIONER, <https://www.esafety.gov.au/about-us/who-we-are/international-engagement/the-global-online-safety-regulators-network> (last visited Aug. 10, 2023).

²³¹ Bruce Baer Arnold, *Safe, but for Whom?: E-Safety Changes*, 166 PRECEDENT (SYDNEY, N.S.W.) 39, 39 (2021).

²³² COLLETT ET AL., *supra* note 1, at 58.

²³³ Ethan Shattock, *Self-Regulation 2.0? A Critical Reflection of the European Fight against Disinformation*, HARVARD KENNEDY SCHOOL MISINFORMATION REVIEW (May 31, 2021), <https://misinfocoreview.hks.harvard.edu/article/self-regulation-20-a-critical-reflection-of-the-european-fight-against-disinformation/> (last visited Aug. 10, 2023).

²³⁴ Naomi Appelman & Paddy Leerssen, *On “Trusted” Flaggers Special Issue: The Yale-Wikimedia Initiative on Intermediaries & Information & Yale Journal of Law and Technology White Paper Series*, 24 YALE J.L. & TECH. 452 (2022).

²³⁵ *Id.* at 472.

planning a response to future harms also offers a sense of protections against the discriminatory and rights-based impacts of AI spiraling out of control. Moreover, application of Gender Based Analysis Plus in other contexts—such as mining²³⁶ and the defense forces²³⁷—means AI regulation can learn and adapt. This includes in relation to lessons on how to operationalize intersectionality in practice.²³⁸ In short, tried and tested models can be adapted for AI's gain.

None of the examples presented are perfect in nature. Brazil has a proposed regulator in its draft bill; this is a gap that has been identified for Canada.²³⁹ Yet Canada has in place the foundations for such regulatory oversight, for instance, with the possibility of investing more in existing structures such as its Commissioners on Competition, Privacy, and Human Rights.²⁴⁰ Meanwhile concerns have been raised about the time involved in undertaking the in-depth gender assessments required under Canadian law. For some, this level of intervention undermines the Canadian government's goal to be agile.²⁴¹ Moreover, all three of the emerging promising examples of legislative practice reflect bills currently under debate. The world is still not where we need to be in terms of existing AI-laws that adopt such a gender perspective. Yet there is promise and we can arrive at even stronger practice by building on these models.

IV. CONCLUSION: FINDING THE WAY FORWARD IN THE GLOBAL LEGISLATIVE LANDSCAPE

There are a number of issues that sit at the periphery of debates on AI and gender that are beyond the scope of this discussion. I have not, for instance, explored in great depth the limitations of women's participation in AI's design that underpins a significant part of the

²³⁶ Dawn Hoogeveen et al., *Sex, Mines, and Pipelines: Examining 'Gender-Based Analysis Plus' in Canadian Impact Assessment Resource Extraction Policy*, 8 *EXTRACTIVE INDUS. & SOC'Y* 100921 (2021).

²³⁷ Rachael Johnstone & Bessma Momani, *Gender Mainstreaming in the Canadian Armed Forces and the Department of National Defence: Lessons on the Implementation of Gender-Based Analysis Plus (GBA+)*, 48 *ARMED FORCES & SOC'Y* 247 (2022).

²³⁸ Olena Hankivsky & Linda Mussell, *Gender-Based Analysis Plus in Canada: Problems and Possibilities of Integrating Intersectionality*, 44 *CANADIAN PUB. POL'Y* 303 (2018).

²³⁹ Teresa Scassa, *Regulating AI in Canada: A Critical Look at the Proposed Artificial Intelligence and Data Act*, 101 *CANADIAN BAR REV.* 29 (2023), <https://cbr.cba.org/index.php/cbr/article/view/4817> (last visited June 22, 2023).

²⁴⁰ *Id.* at 30.

²⁴¹ *Id.* at 29.

barriers to identifying and understanding the biases in AI. Women are simply under-represented in the sector, which also means in the sector's leadership.

It is also worth paying attention to the extent to which the potential use of AI for social good may be overlooked.²⁴² While beyond the scope of this discussion, we must acknowledge the possibility of harnessing AI for social justice for women. Recognizing the opportunity to positively deploy AI is necessary in part to avoid fueling the sense of hype or fear over what is too-often framed as AI's catastrophic and inevitable outcomes. From using Natural Language Processing to assess large volumes of legislation for whether or not existing or draft laws will improve women's lives,²⁴³ through to the use of AI to better assess the risk of reoffending in instances of domestic violence in Spain,²⁴⁴ there is some evident good in AI.

In 2022, the Australian Federal Police and Monash University launched a call, asking all people over the age of 18 to contribute childhood photographs of themselves for a crowdsourcing campaign. AI can be trained and can therefore learn what is a "safe" situation for a child and what is an "unsafe" situation.²⁴⁵ These photographs of a toddler playing in a sandpit; a 9-year-old winning her first award; and a somewhat disgruntled teenager unhappy with their Christmas present reluctantly smiling at the camera are aiding the fight against the online abuse of children.²⁴⁶ A collaboration with AiLecs Labs, this is a solid example of harnessing AI as a response to a technology-based, pervasive and grave risk to children.

²⁴² Nenad Tomašev et al., *AI for Social Good: Unlocking the Opportunity for Positive Impact*, 11 NATURE COMM'NS 2468 (2020); *AI for Social Good*, GOOGLE AI, <https://ai.google/social-good/> (last visited Feb. 2, 2023); *Applying AI for social good*, MCKINSEY (Nov. 28, 2018), <https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good> (last visited Feb. 2, 2023).

²⁴³ There is potential to scale the Gender Legislative Index, and apply Natural Language Processing, to develop a global database of gender-responsive laws. See Ramona Vijayarasa, *Gender Legislative Index*, GENDER LEGIS. INDEX (2019), <https://www.genderlawindex.org/>.

²⁴⁴ Julie Van Hoey et al., *Profile Changes in Male Partner Abuser After an Intervention Program in Gender-Based Violence*, 65 INT'L J. OFFENDER THERAPY COMPAR. CRIMINOLOGY 1411 (2021).

²⁴⁵ AiLECS Lab, *My Pictures Matter*, MONASH UNIV. & AUSTL. FED. POLICE, https://mypicturesmatter.org/?utm_source=Article&utm_medium=Article&utm_campaign=TCITP22&utm_term=&utm_content=TCITP (last visited Sept. 14, 2022).

²⁴⁶ Royce Kurmelovs, *100,000 Happy Pictures: A New Tool in the Cyber 'Arms Race' against Child Sexual Abusers*, THE GUARDIAN (July 24, 2022), <https://www.theguardian.com/technology/2022/jul/25/pictures-of-happy-children-to-fight-child-sexual-abusers-ailecs-lab-afp-australian-federal-police> (last visited Sept. 14, 2022).

This project is one of many emerging that demonstrate AI's potential to fill a gap where other responses have proved inadequate in the past. Rwandan women farmers have been using Buy from Women (an open-source, cloud-based platform set up by UN Women) to predict product levels and crop yields.²⁴⁷ It has been expanded into eight countries. A map of new users' land plots upon registrations is helping these women farmers forecast yields, connect to supply chains and gain access to market prices.²⁴⁸

The list of potentially "good tools" is endless. Brazil-based Think Eva is tracking harassing emails, texts and comments. NexLP's #MeTooBots monitor and flag communications between colleagues and detect bullying and sexual harassment in company documents, emails and chats.²⁴⁹ While seemingly positive interventions given the scope of gender-based harassment, this level of discretionary surveillance—too often unbounded—may leave some wondering who decides what is monitored, how, when that monitoring takes place and who controls the findings. Yet the positive possibilities of AI for women deserve at least a little of our applause.

Yet law is needed to foster AI's use for the good as well as to minimize and ideally eradicate harm. Growing recognition of the degree of such harms has contributed to a "data" versus "Goliath" understanding of AI and the industry behind it.²⁵⁰ In order to avoid a sense of panic, particularly given AI's rapid pace of development, this article presents a careful and deliberate attempt to understand AI's harms from a gender perspective and the role of law in overcoming them.

Moreover, with caution, I note that ensuring women's voices are heard in the development of AI and machine learning technologies could help solve the problem of bias or at least partly mitigate the impact by incorporating different perspectives in the evaluation of data and the impact of gender in the contexts of technology's use. Increasing the presence of women will not resolve the ways in which AI, by design, builds upon datasets that themselves reflect gender biases. Yet the current statistics in terms of women's participation in AI workforces—recalling the figure above that places women as only one-fifth of

²⁴⁷ UN Women launches the "Buy From Women" Innovative Platform in Rwanda, UN WOMEN: AFRICA, <https://africa.unwomen.org/en/news-and-events/stories/2016/07/un-women-launches-the-buy-from-women-innovative-platform-in-rwanda> (last visited Sept. 20, 2022).

²⁴⁸ *Id.*

²⁴⁹ COLLETT ET AL., *supra* note 1, at 58.

²⁵⁰ BRUCE SCHNEIER, DATA AND GOLIATH: THE HIDDEN BATTLES TO COLLECT YOUR DATA AND CONTROL YOUR WORLD 1 (1st ed. 2016).

computer science doctoral-degree graduates²⁵¹—show we are too far at the other extreme for an increase in women’s representation not to make *some* difference. Even computer scientists acknowledge that this is generally not happening given male over-representation, alongside a lack of racial diversity.²⁵² Susan Leavy calls out the over-representation of men in these designs of AI technologies as “quietly undo[ing] decades of advances in gender equality.”²⁵³ Leavy also posits that the leading thinkers in the field who focus on bias in AI are also primarily female,²⁵⁴ suggesting that those who are potentially affected by bias are more likely to see, understand, and attempt to resolve it.²⁵⁵ This is an argument commonly made²⁵⁶ and shows that part of the solution may be far greater participation of women in AI’s design, deployment, use, and leadership.

In fact, in the regulatory battle to minimize the harms of AI, women, and women’s organizations may be readier than we think. The Feminist Principles of the Internet—seventeen principles organized into five clusters—emerged from a meeting in Malaysia in 2014 of 50 activists and advocates working on sexual rights, women’s rights, violence against women, and the Internet.²⁵⁷ The Principles offer clear frameworks for understanding the gender of AI.²⁵⁸ For instance, women need access, unconditionally, meaningfully, and equally. Women’s movements need to use the space (of the internet) and if we extrapolate, the tools of AI, for feminist resistance and movement building. AI is a space of privatization and profit, whereas a feminist vision calls for “alternative forms of economic power” with a commitment to free and open-source software.²⁵⁹ To name another concept from the Feminist Principles of the Internet that proves relevant for AI, women need to maintain full control over their data—if that is possible for any human—as we challenge surveillance and attempts to control or restrict women’s bodies, speech, and activism.

The Terms-We-Serve-With (TWSW) is another feminist approach from which we can learn, in this case, a refusal to accept the present-day

²⁵¹ ZHANG ET AL., *supra* note 122.

²⁵² Latorre Ruiz & Pérez Sedeño, *supra* note 9, at 64.

²⁵³ Leavy, *supra* note 9, at 1.

²⁵⁴ For instance, Buolamwini and Gebu (n 58); Leavy (n 66); Adams and Loideáin (n 32), just to name a few.

²⁵⁵ Leavy, *supra* note 9, at 14.

²⁵⁶ Jackson, *supra* note 24, at 316.

²⁵⁷ *Feminist Principles of the Internet*, ASS'N FOR PROGRESSIVE COMM'NS, <https://feministinternet.org/> (last visited Sept. 14, 2022).

²⁵⁸ *Id.*

²⁵⁹ UNESCO, *supra* note 9, at 11.

take-it-or-leave-it approach to terms-of-service agreements.²⁶⁰ Its primary goal is a more meaningful and engaging form of consent and comes about as feminist activists come to terms with the realities of algorithmic systems in our lives. They try to re-perceive such engagement as relational and less one-directional.²⁶¹ For instance, there is an attempt at co-constitution of user agreements, between the user and supplier.²⁶² There is an attempt at more informed refusal mechanisms that allow for opting out and for users to complain, dispute, or otherwise contest AI decisions and outcomes on the individual or collective level. Mediation is a key principle and seeks to be disclosure-centered.²⁶³

Gendered harms are therefore not inevitable, and AI may be deployed to correct such harm. The newly emerging regulation of AI presents an opportunity to grasp the good in emerging laws—think here of a clear articulation of biased output and the model that Canada offers that other nations can build upon. There is also much to gain in the deployment of novel technologies for good. Finally, a nuanced investment in accurately raising public awareness and correcting public misconceptions will be a must in the years to come.

²⁶⁰ Bogdana Rakova et al., *Terms-we-Serve-with*, <https://termsweservewith.org> (last visited June 21, 2023).

²⁶¹ Bogdana Rakova, *A New Framework for Coming to Terms with Algorithms*, MEDIUM (May 24, 2023), <https://points.datasociety.net/a-new-framework-for-coming-to-terms-with-algorithms-97c74d9667do> (last visited June 21, 2023).

²⁶² *Id.*

²⁶³ *Id.*