

Old Dominion University

ODU Digital Commons

Cybersecurity Undergraduate Research

2021 Fall Cybersecurity Undergraduate
Research Projects

Tackling AI Bias with GANs

Noam Stanislawski

William & Mary

Follow this and additional works at: <https://digitalcommons.odu.edu/covacci-undergraduateresearch>



Part of the [Artificial Intelligence and Robotics Commons](#)

Stanislawski, Noam, "Tackling AI Bias with GANs" (2021). *Cybersecurity Undergraduate Research*. 10.
<https://digitalcommons.odu.edu/covacci-undergraduateresearch/2021fall/projects/10>

This Paper is brought to you for free and open access by the Undergraduate Student Events at ODU Digital Commons. It has been accepted for inclusion in Cybersecurity Undergraduate Research by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

Tackling AI Bias with GANs

Noam Stanislawski, Junior, William & Mary

Research Mentor:

Dr. Iria Giuffrida, Professor of the Practice of Law, William & Mary Law School, and
Visiting Faculty for Business Law, Raymond A. Mason School of Business

Contents

Abstract	2
Part 1 - Introduction	3
1. <i>What is AI?</i>	3
2. <i>What is bias in AI?</i>	6
3. <i>Why should we care about bias embedded within AI systems?</i>	8
Part 2 - Exploring Possible Solutions to Mitigate the Risks of Bias	10
1. <i>Technical Solutions - GANs</i>	10
2. <i>Traditional Sociological Solutions</i>	13
Part 3 - Conclusion.....	16

Abstract

Throughout the relatively short history of artificial intelligence (AI), there has been a significant concern surrounding AI's ability to incorporate and maintain certain characteristics which were not inherently modeled out in its coding. These behaviors stem from the prominent usage of neural network AI, which can inherit human biases from the input data it receives. This paper argues for two possible avenues to combat these biases. The first is to rethink the traditional framework for neural network projects and retool them to be usable by a Generative Adversarial Network (GAN). In a GAN's zero-sum game, two network techniques can combat discriminatory beliefs or incorrect values in manners unlike traditional networks, while not necessitating a completely new algorithm for neural network systems already proven effective. GAN technology is one approach for helping to solve the bias issue but confronting the humans behind the AI is just as important. Incorporating humanistic techniques such as unconscious bias training and participatory design into AI development further promote equitable AI by fostering communication between others. AI biases are merely reflections of human biases in a technological form, and any "bad" output data stems from bad output humanity has generated from itself. There cannot be a perfectly unbiased AI model, as there are no perfectly unbiased humans, and the influences of economies, politics, and other vested interests ensure this to an even larger degree.

Keywords— artificial intelligence, neural network, generative adversarial network, bias, machine learning

Part 1 - Introduction

1. *What is AI?*

The field of artificial intelligence (AI) has continued to grow and improve for over half a century.¹ As an ever-expanding field, AI systems are increasingly being used in commercial contexts where they have been revealed to be carrying certain biases.² In general, these biases are evident within the most common form of AI found within the American workplace: machine learning (ML) algorithms. ML is a branch of computer science dedicated to the study of how algorithms can be developed to solve ever-more-complex problems, up to the point of imitating human learning.³ A neural network is a specific machine learning algorithm which is roughly based on the way in which the human brain is said to work.⁴ These artificial neural networks process data by means of algorithms that are interconnected through different layers. The more fine-tuned the

¹ Rockwell Anyoha, *The History of Artificial Intelligence*, Science in the News, August 28, 2017, <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>

² For instance, Amazon's Recognition software, a facial recognition tool that has been used by law enforcement, has been found to carry gender and racial biases. See, Joy Buolamwini, *Response: Racial and Gender bias in Amazon Rekognition — Commercial AI System for Analyzing Faces*, Medium, January 25, 2019, <https://medium.com/@Joy.Buolamwini/response-racial-and-gender-bias-in-amazon-rekognition-commercial-ai-system-for-analyzing-faces-a289222eeced>. This tool even misidentified 28 members of Congress with convicted criminals. Approximately, 40% of those Congress people were minorities, even those minorities make up only about 20% of Congress people. See, Jacob Snow, *Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots*, ACLU blog, July 26, 2018, <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28>. Amazon has since suspended its sale of Recognition to law enforcement. See, Nick Statt, *Amazon bans police from using its facial recognition technology for the next year*, The Verge, June 10, 2020, <https://www.theverge.com/2020/6/10/21287101/amazon-rekognition-facial-recognition-police-ban-one-year-ai-racial-bias>.

³ IBM Cloud Education, *Machine Learning*, July 15, 2020, <https://www.ibm.com/cloud/learn/machine-learning>.

⁴ Janelle Shane, *Neural Networks, Explained*, Physicsworld, July 9, 2018, <https://physicsworld.com/a/neural-networks-explained/>.

layers are with each other and with the data that they process, the better the quality of the output.⁵ When a neural network has many inner layers, it can be referred to as a “deep learning” network.⁶ Each node is assigned a certain “weight,” which is a value by which the node processes and multiplies the data input that it receives.⁷ At the outset, a neural network model does not have a properly-weighted system since all the weights are likely to have been assigned randomly.⁸ However, as the model is trained on data sets, the relative weights are adjusted and fine-tuned, one bit at the time.⁹ This means that the early outputs from a neural network are likely to be grossly inaccurate, but as the system “learns” from the data it is training on, the more accurate the output is going to be because those weights are being adjusted by the system itself. To put it another way, “[n]eural networks rely on training data to learn and improve their accuracy over time.”¹⁰ The importance of data in training of neural networks will be addressed in section 2.

Neural networks surround us constantly, on- and off- line, and have only continued to grow more useful in a range of fields. There are many frameworks of AI that employ neural networks, and one in particular is attracting the interest of computer scientists and researchers alike. In June 2014, the Generative Adversarial Network (GAN) was created

⁵ *Id.*

⁶ IBM Cloud Education, *Neural Networks*, August 17, 2020, <https://www.ibm.com/cloud/learn/neural-networks>.

⁷ Larry Hardesty, *Explained: Neural networks*, MIT News, April 14, 2017, <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>.

⁸ *Id.*

⁹ *Id.*

¹⁰ *Id.*

by researchers at the Université de Montréal.¹¹ GANs are unique in that there are two neural networks within the system: a *generative* network and an *adversarial* network.¹² The generative network first generates data (or outputs) based on a dataset it has been given. It then passes the generated data to the adversarial network, which then determines whether every piece of data is legitimate or is to be disregarded. GANs are said to embody the application of game theory and zero-sum games to neural networks. There are two players in the game, the two networks working together and against one another. If the generative network's data is wholly perfect, it has “won” the zero-sum game.¹³ If the adversarial network “wins,” all of the input data from the generative network is invalidated. In essence, the goal of GANs is for the two networks to balance each other out.¹⁴

¹¹ Ian J. Goodfellow et al., *Generative Adversarial Nets*, NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems, Vol 2, pp. 2672–2680, December 2014, <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>

¹² *Id.*

¹³ Lina Faik, *A Game Theoretical Approach for Adversarial Machine Learning*, Towards Data Science, May 3, 2020, <https://towardsdatascience.com/a-game-theoretical-approach-for-adversarial-machine-learning-7523914819d5>. It is important to note, however, that there is an emerging literature that disagrees with the notion of GANs being a form of zero-sum games. See Farzan Farnia and Asuman Ozdaglar, *Do GANs always have Nash equilibria?*, Proceedings of the 37th International Conference on Machine Learning, Online, PMLR 119, 2020, <http://proceedings.mlr.press/v119/farnia20a/farnia20a.pdf>.

¹⁴ Farzan Farnia and Asuman Ozdaglar, *Do GANs always have Nash equilibria?*, Proceedings of the 37th International Conference on Machine Learning, Online, PMLR 119, 2020, <http://proceedings.mlr.press/v119/farnia20a/farnia20a.pdf>.

2. *What is bias in AI?*

In many sci-fi movies, AI is presented as the solution to many human biases and conflicts.¹⁵ An AI system does not have feelings or thoughts, instead only generating based upon the data it was initially fed. In principle, this should lead to less-biased decisions. However, it is important to stress that machine learning algorithms wholly rely on their input data, and on how the algorithms have been designed by human scientists. The imperfections of humanity cannot be removed from AI systems. Humans hold biases towards other humans, places, emotions: bias is an integral part of how we think.¹⁶

In order to understand how deeply impactful implicit biases are within popular culture, consider the following example. In the 1920s, the first ever Band-Aid was invented by an employee of Johnson & Johnson.¹⁷ Cheap and absorbent, the Band-Aid could cover up small cuts extremely effectively and prevent infections so easily that their popularity exploded.¹⁸ Soon variations of Band-Aids came out, using different sizes, shapes, and beautiful different colors to utilize a Band-Aid as an accessory. In fact, Band-Aid iconography was so popular in mainstream culture at certain points that

¹⁵ Omar Mubin et al, *Reflecting on the Presence of Science Fiction Robots in Computing Literature*, ACM Transactions on Human-Robot Interaction, Vol. 8(1), March 2019, <https://dl.acm.org/doi/10.1145/3303706>.

¹⁶ Stanford Encyclopedia of Philosophy, *Implicit Bias*, First published February 26, 2015; substantive revision July 31, 2019, <https://plato.stanford.edu/entries/implicit-bias/>.

¹⁷ Johnson & Johnson, *Stick With It: 18 Fun Facts About the History of BAND-AID® Brand Adhesive Bandages*, <https://www.jnj.com/our-heritage/18-facts-about-the-history-of-band-aid-brand-adhesive-bandages>.

¹⁸ Kat Eschner, *Get Stuck on Band-Aid History*, Smithsonian Magazine, October 10, 2017, <https://www.smithsonianmag.com/smart-news/get-stuck-Band-Aid-history-180965157/>.

many began to wear Band-Aids purely for fashion.¹⁹ While Band-Aids could now be in any color, the default choice was a pale, peach tone matching Caucasian skin. Only in June 2020, over 100 years since the invention of the Band-Aid, Johnson & Johnson announced they would create Band-Aids designed for black and brown skin tones.²⁰ After creating white skin Band-Aids for over a century, Johnson & Johnson finally realized the value in diversifying Band-Aid color for all Americans. The unconscious bias of the white consumer market along with Johnson & Johnson is crucial to comprehending how many implicit biases in consumer products play out throughout modern American history.

Band-Aid as an implicit bias study demonstrating discriminatory design is simple to understand.²¹ Although it does not have the consequence of physically or professionally harming individuals, it is emblematic of larger cultural spheres that give preference for those within certain culturally defined groups. Consider the case of Amazon's job application AI tool. In 2014, Amazon unveiled a new system to find top talent within the hundreds of thousands of applicants to their job positions.²² In order for the AI to learn proper application selection, Amazon fed it the data of their current—and successful—

¹⁹ Kayleen Shaefer, *It's No Boo-Boo: Bandages as Fashion Accessories*, The New York Times, August 6, 2008, <https://www.nytimes.com/2008/08/07/fashion/07SkinTwo.html>.

²⁰ Chauncey Alcorn, *Band-Aid will make black and brown flesh-toned bandages*, CNN Business, June 12, 2020, <https://www.cnn.com/2020/06/12/business/black-band-aids/index.html>.

²¹ D.E. Wittkower, *Principles of Anti-Discriminatory Design*, 2016 IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS), May 2016, https://digitalcommons.odu.edu/philosophy_fac_pubs/28/.

²² Jeffrey Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women*, Reuters, October 10, 2018, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>.

employees. There was little media coverage of this tool while it remained in development. However, once deployed to assist in the hiring process, it became noticeable that the AI's results were problematic. Amazon, like most tech companies, is a predominantly white and male institution. As a result and unbeknownst to the human trainers, the AI tool selected applicants who fit those same characteristics. The AI would specifically penalize an applicant's resume if it included the word "women" on it anywhere.²³ In another model created by a separate Amazon team, their AI favored more "harsh, masculine" wording such as "executed" and "captured".²⁴ Amazon employees tried to continue to devise methods to counterbalance the AI's bias, but at a certain point, it became evident that the machine could develop new methods indefinitely to perpetrate the bias that it had learnt from the data it had been trained on. In other words, the bias was so embedded in the system that no amount of fine-tuning the relative weights of the neural networks would have removed the bias from the system.

3. Why should we care about bias embedded within AI systems?

With the surge in usage of AI systems in vastly different fields, the issue of bias is gaining critical importance. Using a set of sophisticated algorithms to choose a resume is one problem, but this example was merely the first ten yards of an incredibly large field of dubious and problematic AI applications. AI evaluation of human subjects is an

²³ Isobel Asher Hamilton, *Amazon built an AI tool to hire people but had to shut it down because it was discriminating against women*, Business Insider, October 10, 2018, <https://www.businessinsider.com/amazon-built-ai-to-hire-people-discriminated-against-women-2018-10>.

²⁴ See Jeffrey Dastin, footnote 22 above.

extremely delicate subject, with one small error in a training set or algorithm derailing the entire platform.²⁵ There have been AI beauty contests,²⁶ AI applications for determining recurring felons,²⁷ and even AI chatbots that have talked amongst the world's population on social media.²⁸ Within each of these cases, biases towards certain races, ethnicities, and gender identities became the norm with the AI system. This shows the inability to separate human biases from AI systems, but since AI is only increasing its role in our societies, do we need to worry or are there some possible solutions?

A PWC study from January 2021 of over 1,000 American company executives further reinforced the fact that AI will only be more important in America's professional future.²⁹ Eighty six percent of these executives, which span over multiple large-scale sectors of the American economy, agreed that AI is a mainstream technology in 2021. In addition, over half of them responded that they have now accelerated their AI adoption

²⁵ Andre Ye, *AI Algorithms Are Much More Fragile Than You Think*, Medium.com, July 12, 2020, <https://medium.com/swlh/machine-learning-algorithms-are-much-more-fragile-than-you-think-25fdb3939fee>.

²⁶ Sam Levin, *A beauty contest was judged by AI and the robots didn't like dark skin*, The Guardian, September 8, 2016, <https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people>.

²⁷ Karen Hao, *AI is sending people to jail—and getting it wrong*, MIT Technology Review, January 21, 2019, <https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/>.

²⁸ Oscar Schwartz, *In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation The bot learned language from people on Twitter—but it also learned values*, IEEE Spectrum, November 25, 2019, <https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation>. See, generally, Daniel James Fuchs, *The Dangers of Human-Like Bias in Machine-Learning Algorithms*, Missouri S&T's Peer to Peer, Vol. 2(1), May 2018, <https://scholarsmine.mst.edu/cgi/viewcontent.cgi?article=1030&context=peer2peer>.

²⁹ PWC, *AI Predictions 2021— How to navigate the top five AI trends facing your business*, October 2020, <https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html>.

plans due to the COVID-19 pandemic.³⁰ Whether one supports the widespread adoption of AI or not, businesses and governments will continue to integrate into their plans for the foreseeable future. With this accelerated adoption, the concern for biases within their systems may grow, and the impacts on our society are largely unknown.

Part 2 - Exploring Possible Solutions to Mitigate the Risks of Bias

In the first section, the field of AI and its relationship to bias was established. The issues, both due to technical definitions of AI systems and human factors are clear. While it is an impossible task to eliminate bias completely from a given system, there are multiple strategies that can be implemented in order to mitigate the possible risks. This section explores two possible avenues to help mitigate possible biases: using a GAN based AI framework to push towards increasingly equitable technology and addressing human biases on a community wide scale so that those with power can consider the implications of their work before it is used.

1. *Technical Solutions - GANs*

As briefly mentioned above, generative adversarial networks (GANs) are a specific framework of machine learning. When compared to other machine learning algorithms, GANs are flexible in their adaptation to different types of data.³¹ Compared to three other common machine learning frameworks (deep directed graphical models, deep undirected graphical models, and generative autoencoders), GANs stand out as not

³⁰ *Id.*

³¹ Kabir Doshi, *An Introduction To Generative Adversarial Networks (GANs)*, A42 Labs, January 20, 2021, <https://www.a42labs.io/an-introduction-to-generative-adversarial-networks>

falling to the “Helvetica problem.”³² This occurs when the generator has placed too high a probability value upon a single value, which would eliminate the diversity of the network.³³ As long as the discriminator updates together with the generator this issue should not occur, as the discriminator (when working correctly) can act as a diversifying force in the network.³⁴ GANs are formidable in the face of a wide variety of situations and their applications are still just being explored.

Utilizing GANs in a scenario where other networks fail is not a new idea. Mkhuseli Ngxande, Jules-Raymond Tapamo, and Michael Burke wrote in their paper of February 2020 about the utilization of GANs as a technique to eliminate bias consciously.³⁵ By analyzing the imagery of thousands of drivers who are awake, tired, falling asleep, or completely asleep, their GAN and neural network had learned to determine the state of tiredness of a given face with a high probability (0.7-0.99).³⁶ Besides their success in determining drowsiness, their research is impactful as they specifically attempted to mitigate racial biases found in previous data sets. As this research has come out of South Africa, there was a special importance put on differentiating the skin tones of black

³² Ian J. Goodfellow et al., *Generative Adversarial Nets*, NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems, Vol 2, pp. 2672–2680, December 2014, <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>

³³ *Id.* at page 7.

³⁴ *Id.*

³⁵ Mkhuseli Ngxande et al., *Bias Remediation in Driver Drowsiness Detection Systems Using Generative Adversarial Networks*, IEEE Access, Vol. 8, pp. 55592-55601, 2020, <https://ieeexplore.ieee.org/document/9042231>.

³⁶ *Id.* at page 55598.

people, as many traditional facial recognition models tend to suffer with this exact issue.³⁷

The datasets utilized by Ngxande, Tapamo, and Burke included faces of people who are of European, Asian, and African descent.³⁸ Their framework focused upon running each face through a neural network, which would perform well on some faces but not others.³⁹ The data points that confused the network would be given as input data for the GAN, which would, in turn, send data to the neural network to fine tune its algorithmic process.⁴⁰ With this method, the network's chance at a correct guess increased a tremendous amount, with it producing a highly confident result on almost any image by the seventh visualization generation.⁴¹

Although GANs have high potential applications in reducing bias, this study demonstrates that eliminating biases with GANs still places a large emphasis on the input dataset. One of the reasons GANs flourish in image processing is the uniformity of any given input and output data. Researchers at Arizona State demonstrated how GANs can seemingly fall into the same issues as other machine learning algorithms in regards

³⁷ Mutale Nkonde, *Automated Anti-Blackness: Facial Recognition in Brooklyn, New York*, Harvard Kennedy School Journal of African American Policy, Vol. 20, 2019, <https://pacscenter.stanford.edu/wp-content/uploads/2020/12/mutalenkonde.pdf>. See also, Damien Patrick Williams, *Fitting the Description: Historical and Sociotechnical Elements of Facial Recognition and Anti-Black Surveillance*, Journal of Responsible Innovation, Vol. 7(1), 74-83, 2020, <https://www.tandfonline.com/doi/full/10.1080/23299460.2020.1831365>.

³⁸ Mkhusele Ngxande et al., at footnote 35.

³⁹ *Id.*

⁴⁰ *Id.* page 55595.

⁴¹ *Id.* page 55598.

to maintaining biases against traditionally discriminated groups.⁴² Due to GANs relying on game theory and a min-max function, there can be cases where the function does not produce results or data with high accuracy, as the function may not always converge. This failure is seen on a Snapchat face filter, where they employed a GAN system to those with darker skin and in order to produce “beautiful” results, their skin would be lightened.⁴³ Because of this instability, a GAN framework by itself cannot be a complete solution to a biased AI system, but the framework employed of modifying results with competing networks could be a possible move in the right direction.

2. *Traditional Sociological Solutions*

In the average American workplace, it is becoming more popular that employees must go through unconscious bias training (UBT). This is primarily done to create a more inclusive workplace and addresses biases coworkers may have towards each other due to race, gender, ethnicity, religion, or other social backgrounds.⁴⁴ This generally correlates to a decrease in the prevalence of both conscious and unconscious biases throughout the workplace, but does not lead to the elimination of implicit biases altogether.⁴⁵ In addition, there is the possible downside of UBT backfiring, as some see

⁴² Niharika Jain, *A Study on Generative Adversarial Networks Exacerbating Social Data Bias*, A Thesis Presented in Partial Fulfillment of the Requirements for the Degree Master of Science (2020), <https://proxy.wm.edu/login?url=https://www.proquest.com/dissertations-theses/study-on-generative-adversarial-networks/docview/2407359001/se-2?accountid=15053>.

⁴³ *Id.* page 27.

⁴⁴ Doyin Atewologun, Tinu Cornish & Fatima Tresh, Unconscious bias training: An assessment of the evidence for effectiveness 4–45 (2018), https://cstar.ie/equality/t4media/ub_an_assessment_of_evidence_for_effectiveness.pdf.

⁴⁵ *Id.* page 7.

biases or stereotypes as unchangeable after having gone through UBT.⁴⁶ Nonetheless, UBT is seen as a standard business practice to promote diversity and inclusion in the workforce today.

While UBT may suffice as a bare minimum bias deterrent in a small workplace environment, it is seen as largely ineffective for taking down institutional and structural inequities that plague large social institutions.⁴⁷ One key example is noted within the realm of academia. In general, “women remain disproportionately less likely to receive faculty appointments, obtain leadership positions, earn comparable wages, receive grant funding, and are more likely to leave the academy [academia] prematurely.”⁴⁸ UBT places an emphasis on one’s individual biases and thoughts on others, but it is not a viable strategy in order to tackle institutions that have been entrenched for centuries in fields dominated by those at the top of traditional power structures (generally Caucasian men).⁴⁹

Unconscious bias training may be enough for some companies, but in order to tackle the biases within an AI system, one cannot just have a single training approach. There are professional technical institutions which have codes of ethics (IEEE, ACM),⁵⁰ but these guidelines are either ineffective (because there are no consequences for those

⁴⁶ *Id.* page 23.

⁴⁷ Cheryl Pritlove et al., *The good, the bad, and the ugly of implicit bias*, 393 *The Lancet* 502–504 (2019), <https://www.thelancet.com/action/showPdf?pii=S0140-6736%2818%2932267-0>.

⁴⁸ *Id.* page 502.

⁴⁹ *Id.* page 503

⁵⁰ Ayanna Howard & Jason Borenstein, *The ugly truth about ourselves and our robot creations: The problem of bias and social inequity*, 24 *Science and Engineering Ethics* 1521–1536 (2017), <https://link.springer.com/content/pdf/10.1007/s11948-017-9975-2.pdf>.

who disregard them), or not followed by corporations implementing AI systems, as these issues are still widespread. As such, the burden of ethics falls upon those with direct power as AI engineers and developers. The concept of participatory design, creating systems by involving those who may be directly affected by it, would be a crucial social component to utilize when designing an AI which could fall victim to social biases.⁵¹ This design philosophy has done very well when creating AI or robotics which interact with the elderly, as some medical robots have shown great success when being designed with direct input coming from older people who require help.⁵² Even within a small sample size of a single elderly home, the heterogeneity of the community and different opinions of those in need helped to design a user product which worked for everyone.⁵³

UBT and participatory design are two methodologies that companies and innovators can employ in order to reduce social implicit biases on a small-community wide level. The work to help promote marginalized communities within the larger scale of the American workforce is a daunting task, but small steps such as these could lead to improved communication and the creation of positive work environments for the future.

⁵¹ Tone Bratteteig and Guri Verne, Does AI make PD obsolete? exploring challenges from artificial intelligence to participatory design, in Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial, 2018, Vol 2, pages 1–5, <https://doi.org/10.1145/3210604.3210646>.

⁵²Jia Zhou & Gavriel Salvendy, *The Benefits of Involving Older People in the Design Process*, in Human aspects of it for the aged population: First international conference, ITAP 2015, held as part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015; proceedings (2015), <https://link.springer.com/content/pdf/10.1007%2F978-3-319-20892-3.pdf>.

⁵³ *Id.* page 5.

Part 3 - Conclusion

Implicit biases are found to be an issue prevalent within AI systems and with the humans who design their algorithms and collect the data upon which AI systems are trained. Current machine learning frameworks, particularly standard neural networks, have been accused of reinforcing the stereotypes of those who have power within a given company, but also within American society at large. A GAN based framework, which utilizes a discriminatory network, can provide added benefit in deterring the traditional weaknesses of ML systems. However, it is still a fragile and new technology which can suffer the same fate as other systems have. This is tied to the source of biases themselves: humans. AI algorithms can only output what their given algorithm tells them to, and that is all generated from the brains of software engineers. AI biases are only emblematic of the biases which existed before they were born, and to judge their morality is to judge the morality of the society which they were born from. The hope is that multidisciplinary research will continue its efforts to find possible solutions to mitigate these risks while promoting innovation.