BIAS IN ARTIFICIAL INTELLIGENCE (AI)

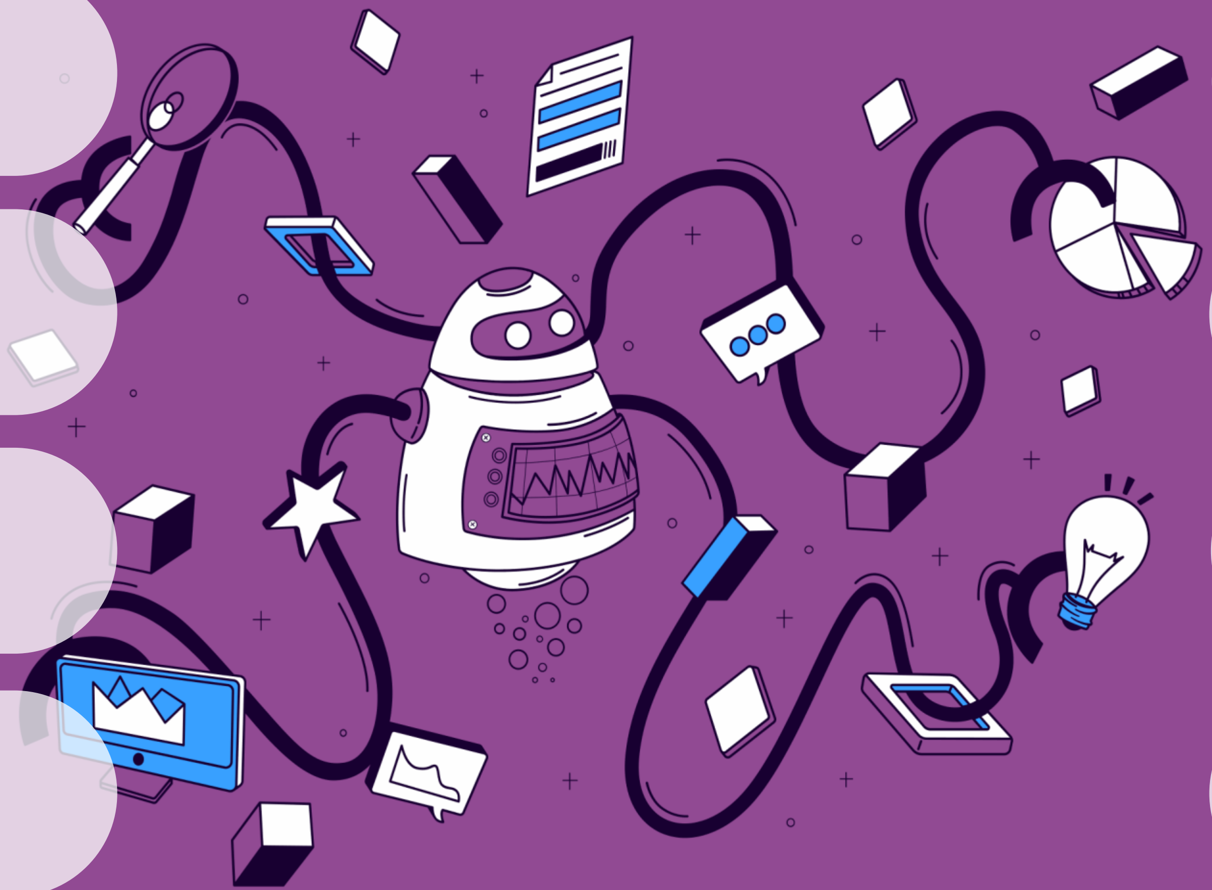
David Newey, Deputy Chief Information Officer,
The Royal Marsden NHS Foundation Trust

BIAS IN HEALTHCARE

AI USE IN HEALTHCARE AND BIAS EXAMPLES

BIAS IN CHOOSING MODELS

ELIMINATING BIAS



INEQUALITY IN HEALTHCARE

TYPES OF BIAS IN AI

DESIGNING FOR EQUALITY

BEING AN ALLY IN THE DIGITAL WORLD

Biases in Healthcare

Research Representation

Sexual Identity	Geographic Location	Socioeconomic Status
Obesity	Age	Ableism
Education	Sex and Gender	Racial bias

Inequality In Healthcare

Maternal deaths for various ethnic groups in comparison to White mothers (***BMA, 2021***)	
Black	5x more likely
Asian	2x more likely
Likelihood of medication receipt for Non-White patients (*Dehon et al., 2016*)	
Analgesic Medicine	22-30% less likely
Narcotic Analgesics	17% to 30% less likely

Waiting times are usually longer for Non-White Patients. Similarly, they usually have less likelihood of admission (Shah et al., 2015)

Hospitalisation was the highest for South Asian communities, with the highest being men and women of Pakistani origin.

Some medical professionals were more likely to dismiss chronic pain in women than in men (Samulowitz et al, 2018)

LGBTQIA+ Medical Experiences (***Samulowitz et al., 2018***)		
Percentage of medical students with implicit bias to LGBTQIA+ peoples	›80% of medical students	
Common ailments for LGBTQIA+ patients due to avoidance of medical assistance	❑ Anal Cancer ❑ Cardiovascular Disease ❑ Asthma	❑ Substance Misuse ❑ Obesity ❑ Suicide

Use Of Artificial Intelligence In Healthcare

Rapid diagnosis of cancer - MALIMAR (Machine Learning in Myeloma Response)

Study using machine learning[1] to read whole-body MRI scans in myeloma patients to find evidence of cancer.

Intelligent symptom checker using chatbots

- ❑ *https://www.buoyhealth.com/*

Radiology reporting and analysis

- ❑ *https://www.nanox.vision/*

Oculus Muscular Degeneration Prediction

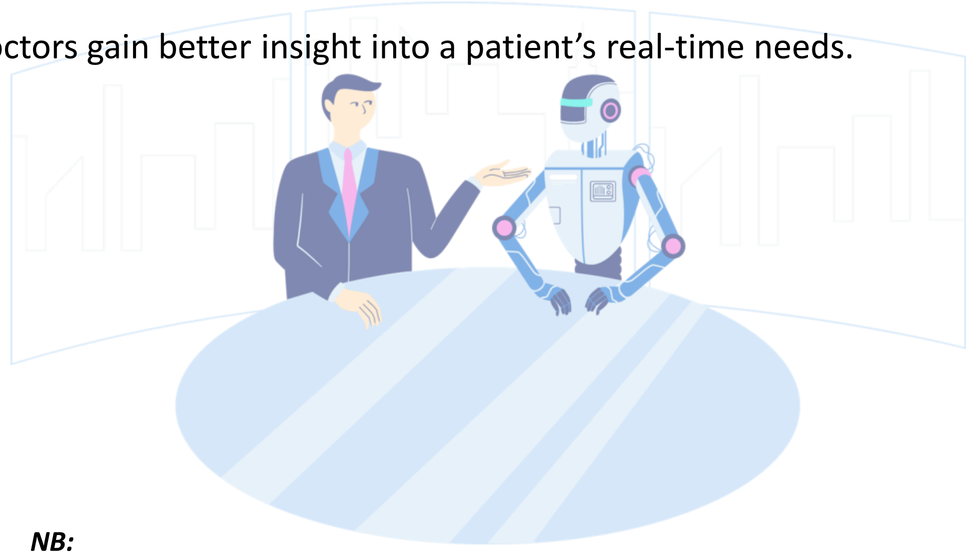
Predicts whether people with age-related macular degeneration will develop the more serious form of the condition in their 'good eye'. ***(Digitalhealth.net, 2020)***

Unstructured Medical Data Analysis with AI

Deep learning platforms like ***Enlitic*** can be used to analyse large, unstructured medical data. This includes:

- ❑ Radiology Images.
- ❑ Blood Tests.
- ❑ EKGs.
- ❑ Genomics.
- ❑ Patient medical history.

These enable doctors gain better insight into a patient's real-time needs.



NB:

1. A type of AI in which computers are taught how to do things independently

Types Of Biases In AI

IB

Implicit Bias

Unconscious prejudice formed against person(s) that is not easily noticed by the owner of such prejudice.

SB

Sampling Bias

A statistical problem where the sample data may be **skewed towards specific sections** of the data.

TB

Temporal Bias

We can build a machine-learning model that eventually becomes obsolete due to **future events not factored into the model**. .

O

Over-fitting to training data

When AI models accurately predict values from the training dataset but **cannot predict new data accurately**, thus unrepresentative of general population.

E

Edge Cases & Outliers

Outliers are **data points not within the data's normal distribution**. Edge cases are errors (missing/incorrect datasets) or noise (additional, irrelevant datasets that could impact the machine learning process negatively).

Source: Omowole, A. (2021) *Research shows AI is often biased. Here's how to make algorithms work for all of us :* https://www.weforum.org/agenda/2021/07/ai-machine-learning-bias-discrimination/ .

Examples Of Bias In Artificial Intelligence

1 Amazon

In 2014, a team of software engineers at Amazon were building a program to review the resumes of job applicants. Unfortunately, in 2015 they realized that the system discriminated against women for technical roles (Dastin, 2018)

2 Post Discharge

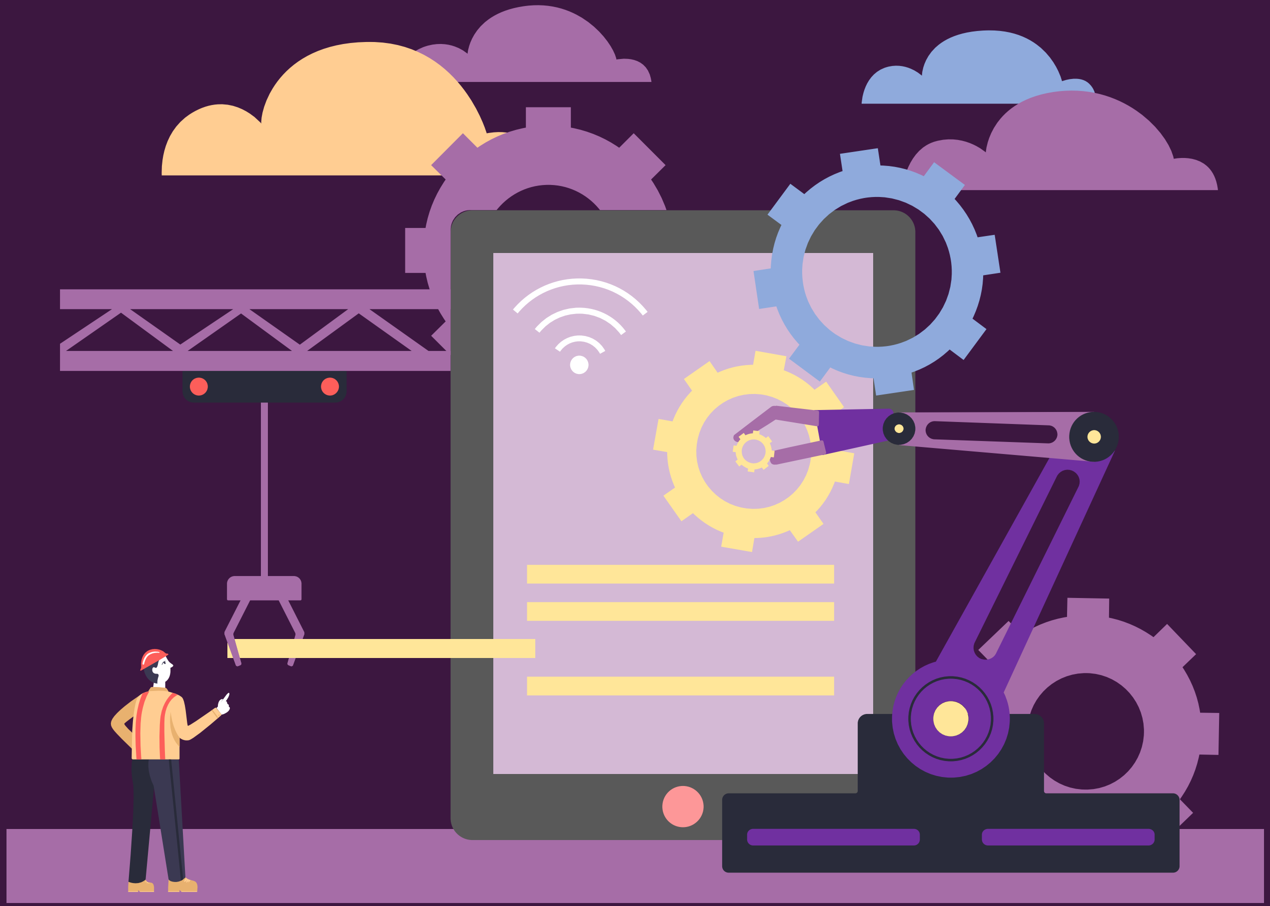
One university considered using AI to direct case management resources to patients for early discharge, until leaders recognized that doing so would preferentially benefit wealthy white patients and disadvantage poorer African-Americans (Rajkomar et al, 2018)

3 Health Insurance

A commercial algorithm to guide resource allocation in healthcare was found to be profoundly biased against black patients (Obermeyer et al, 2019)

4 Facial Recognition

In 2019, San Francisco legislators voted against the use of facial recognition, believing they were prone to errors when used on people with dark skin or women (BBC, 2019)



Examples of Bias in Artificial Intelligence

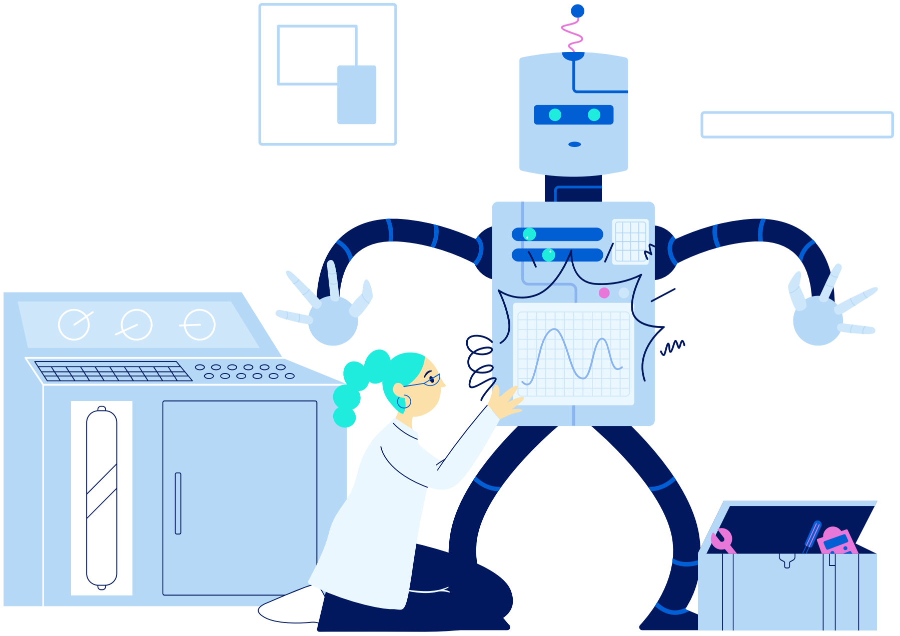
Pre-existing bias in datasets results in training biased models; under sampling of marginalized populations results in worse predictive accuracy in comparison to more privileged, represented populations. (***Forde et al,2021***)

This skin cancer prediction model showed that in striving to achieve overall accuracy, male accuracy was considerably greater

Table 1: Algorithms resulting in different model accuracy–fairness trade-offs for skin cancer risk prediction as presented in Srivastava et al. (2019). The authors asked test subjects to choose between the presented models and found that subjects preferred A_1 (which has highest overall accuracy).

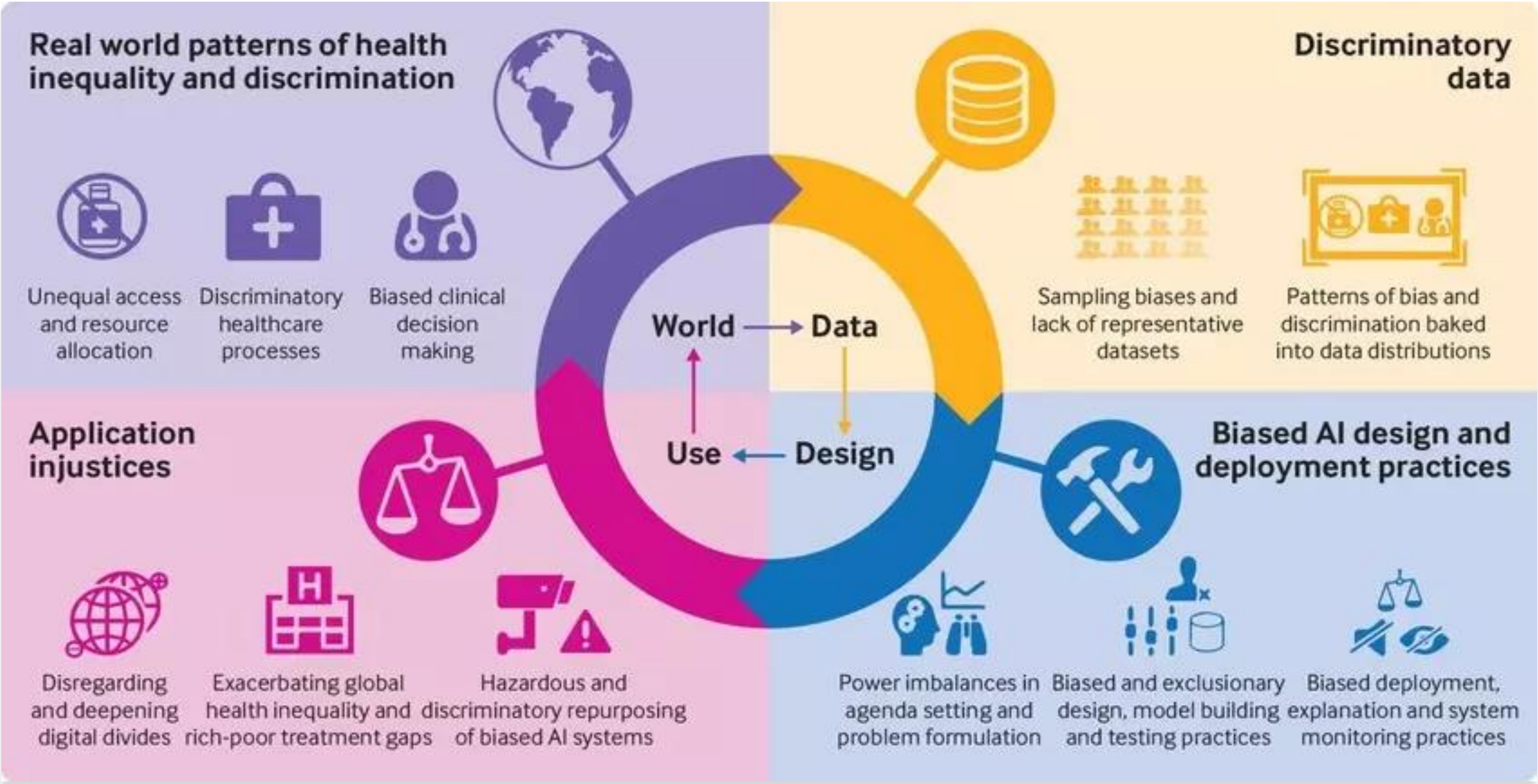
Algorithm	Overall Accuracy
A_1	94%
A_2	91%
A_3	86%

Bias In Choosing Models



- During model selection for solving a particular task, the model developer compares differences in the performance of several learned models trained under various conditions, such as different optimizers or hyperparameters.
- This procedure is not a strictly computational; rather, the metrics used to distinguish between models are subject to human interpretation and judgement (Adebayo et al., 2018; Jacobs & Wallach, 2019).
- Human preferences, often geared toward the particular application domain, ultimately play an important role in choosing the model to deploy.
- A Columbia University study found that "the more homogenous the [engineering] team is, the more likely it is that a given prediction error will appear." This can create a lack of empathy for the people who face problems of discrimination, leading to an unconscious introduction of bias in these algorithmic-savvy AI systems. ***(Wiggers, 2020)***

Designing For Equality



Source: Omowole, A. (2021) *Research shows AI is often biased. Here's how to make algorithms work for all of us*: https://www.weforum.org/agenda/2021/07/ai-machine-learning-bias-discrimination/ .

Best Practice To Eliminate Bias

- In October 2021, the US Food and Drug administration published a set of guidelines to advise on best practice when developing AI algorithms ***(US Food and Drug Administration, 2021)***.
- **Key aspects include:**
  - Clinical Study Participants and Data Sets Are Representative of the Intended Patient Population.
  - Training Data Sets Are Independent of Test Sets.
  - Testing Demonstrates Device Performance During Clinically Relevant Conditions. Considerations include the intended patient population, important subgroups, clinical environment and use by the Human-AI team, measurement inputs, and potential confounding factors.
  - Deployed Models Are Monitored for Performance and Re-training risks are managed.
- Inclusive design emphasizes inclusion in the design process (***Omowole, 2021***).
- The AI product should be designed with consideration for diverse groups such as gender, race, class, and culture.
- Foreseeability is about predicting the impact the AI system will have right now and over time.

Being An Ally In The Digital World

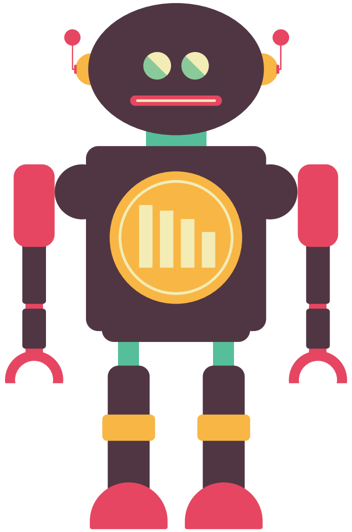
- AI poses a threat to societal harmony, & equality
  - Facebook
  - Twitter
  - Alexa/Siri
- AI will increasing be used to augment & then replace humans processes
- There is an urgent need to address this issue now before algorithms become pervasive.

Being An Ally In The Digital World

To be an ally in the digital world is not just about designing for inclusive UX, but ensuring diversity in the design and use of AI algorithms. Its about stopping digital prejudice before it becomes by actively:

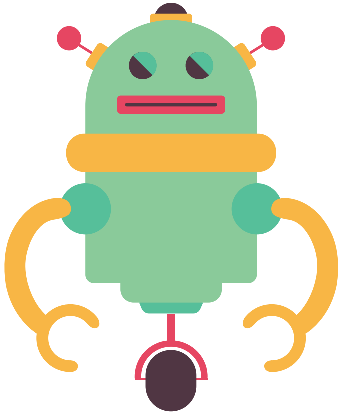
A

Raising awareness amongst under-represented teams.



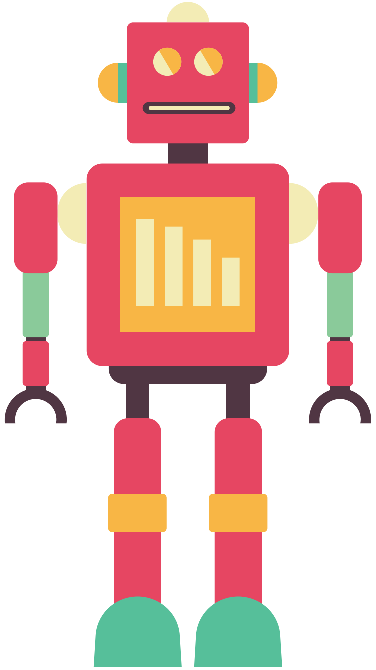
B

Recruiting and training a diverse team



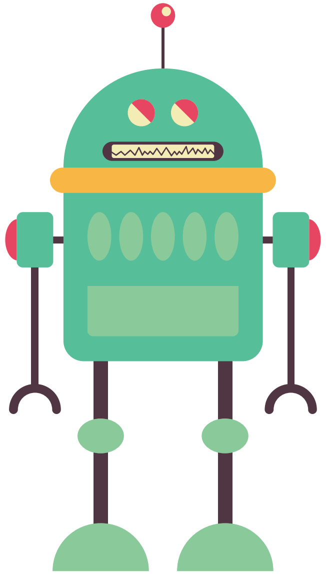
C

Providing challenge and oversight around training and testing of algorithms.



D

Ensuring governance is in place to undertake regular performance reviews against algorithms.



Digital Ethics Charter

https://www.ethicscharter.co.uk

References

Biases in healthcare: An overview. (2021) Available from: https://www.medicalnewstoday.com/articles/biases-in-healthcare .

British Medical Association. (2021) *A missed opportunity BMA response to the Race Report.*

Dastin, J. (2018) *Amazon scraps secret AI recruiting tool that showed bias against women.* Available from: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G .

Dehon, E., Weiss, N., Jones, J., Faulconer, W., Hinton, E., Sterling, S. & Choo, E. K. (2017) A Systematic Review of the Impact of Physician Implicit Racial Bias on Clinical Decision Making. *Academic Emergency Medicine.* 24 (8), 895-904. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/acem.13214. Available from: doi: 10.1111/acem.13214.

FitzGerald, C. & Hurst, S. (2017) Implicit bias in healthcare professionals: a systematic review. *BMC Medical Ethics.* 18 (1), 19. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28249596. Available from: doi: 10.1186/s12910-017-0179-8.

Forde, J. Z., Cooper, A. F., Kwegyir-Aggrey, K., De Sa, C. & Littman, M. (2021) Model Selection's Disparate Impact in Real-World Deep Learning Applications. Available from: https://arxiv.org/abs/2104.00606.

Lee, D. (2019) *San Francisco is first US city to ban facial recognition.* Available from: https://www.bbc.co.uk/news/technology-48276660 .

Morris, M., Cooper, R. L., Ramesh, A., Tabatabai, M., Arcury, T. A., Shinn, M., Im, W., Juarez, P. & Matthews-Juarez, P. (2019) Training to reduce LGBTQ-related bias among medical, nursing, and dental students and providers: a systematic review. *BMC Medical Education.* 19 (1), 325. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31470837. Available from: doi: 10.1186/s12909-019-1727-3.

Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science (American Association for the Advancement of Science).* 366 (6464), 447-453. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31649194. Available from: doi: 10.1126/science.aax2342.

Omowole, A. (2021) *Research shows AI is often biased. Here's how to make algorithms work for all of us.* Available from: https://www.weforum.org/agenda/2021/07/ai-machine-learning-bias-discrimination/ .

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G. & Chin, M. H. (2018) Ensuring Fairness in Machine Learning to Advance Health Equity. *Annals of Internal Medicine.* 169 (12), 866-872. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30508424. Available from: doi: 10.7326/M18-1990.

Sabin, J. A., Riskind, R. G. & Nosek, B. A. (2015) Health care providers' implicit and explicit attitudes toward lesbian women and gay men. *American Journal of Public Health (1971).* 105 (9), 1831-1841. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26180976. Available from: doi: 10.2105/AJPH.2015.302631.

Samulowitz, A., Gremyr, I., Eriksson, E. & Hensing, G. (2018) "Brave Men" and "Emotional Women": A Theory-Guided Literature Review on Gender Bias in Health Care and Gendered Norms towards Patients with Chronic Pain. *Pain Research & Management.* 2018 6358624-14. Available from: https://dx.doi.org/10.1155/2018/6358624. Available from: doi: 10.1155/2018/6358624.

Shah, A., Zogg, C., Zafar, S., Schneider, E., Cooper, L., Chapital, A., Peterson, S., Havens, J., Thorpe, R., Roter, D., Castillo, R., Salim, A. & Haider, A. (2015) Analgesic Access for Acute Abdominal Pain in the Emergency Department Among Racial/Ethnic Minority Patients. *Medical Care.* 53 (12), 1000-1009. Available from: https://www.jstor.org/stable/26418082. Available from: doi: 10.1097/MLR.0000000000000444.

US Food and Drug Administration. (2021) *Good Machine Learning Practice for Medical Device Development: Guiding Principles.* US Food and Drug Administration.

Wiggers, K. (2020) *Study finds diversity in data science teams is key in reducing algorithmic bias.* Available from: https://venturebeat.com/2020/12/09/columbia-researchers-find-white-men-are-the-worst-at-reducing-ai-bias/ .