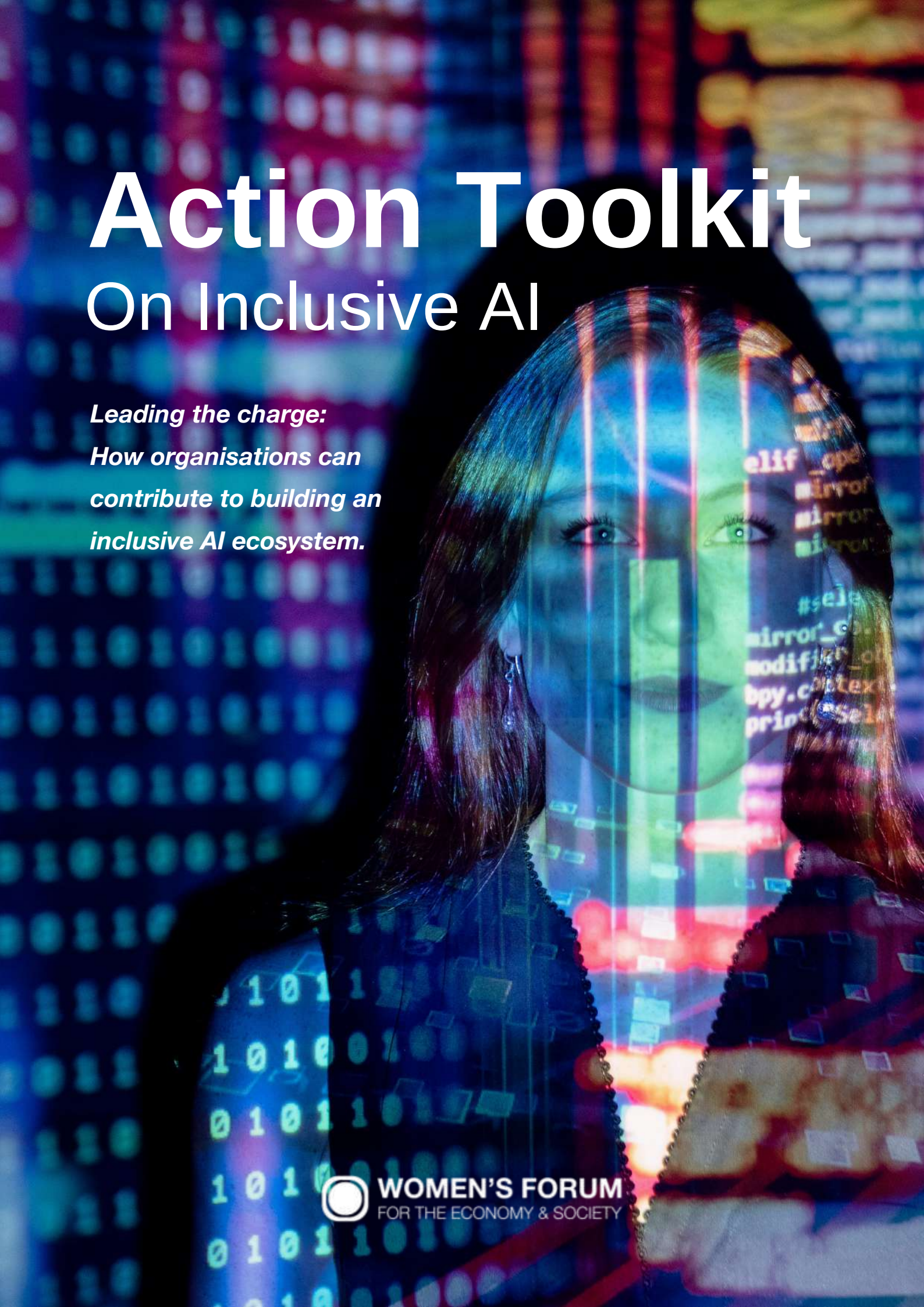


Action Toolkit
On Inclusive AI

Leading the charge: How organisations can contribute to building an inclusive AI ecosystem.

WOMEN'S FORUM
FOR THE ECONOMY & SOCIETY

TABLE OF CONTENTS

INTRODUCTION

Welcome to the Inclusive AI toolkit of the Women4AI Daring Circle of the Women's Forum for the Economy & Society. The Daring Circles are unique cross-industry working groups, which brings together an ecosystem of partners to develop tangible proposals and solutions to today's most pressing issues.

The Women4AI Daring Circle advocates for making the design and development of AI truly inclusive. Its aim is to encourage organisations of all types and sizes to take tangible action to create an environment where AI is designed by diverse teams and governed appropriately, so that AI can become a powerful asset to right existing disparities and shape a more inclusive world.

The Daring Circle is led by Microsoft in collaboration with L'Oréal, AXA, BNP Paribas, Publicis Groupe, Bayer, Lenovo, Société Générale and ExxonMobil. UNESCO participates as an Institutional Partner, Oxford Internet Institute as an Academic Partner, Shearman & Sterling and PwC, as Knowledge & Insight Partners.

This toolkit is the result of two years of research work by the Women4AI Daring Circle of the Women's Forum for the Economy & Society. It has been developed in partnership with PwC, and the Oxford Internet Institute. The recommendations put forward are based on a body of research including a survey, a series of expert interviews, focus groups, a deep review of literature and consultation with partners.

This instrument has been conceived for organisations who want to use AI in an ethical, inclusive way. The toolkit offers practical guidance on what Inclusive AI means, how to advocate for it internally, and how to achieve it. It is created for C-suite executives, technologists, HR managers, board members, developers, engineers and anyone who wants to change practice, policy, strategy and attitudes within their organisation towards ethical, inclusive AI. For organisations to truly achieve progress towards inclusive AI, the toolkit maintains that is imperative to embed Diversity and Inclusion from the onset of their digital transformation.

This toolkit is designed to support organisations who are at different stages of their Inclusive AI journey and the sections contained within are designed to provide support for those different stages. The 'Defining Inclusive AI' and 'The Case for Inclusive AI' sections provide a definition of Inclusive AI and offer advice on making the case for it within your organisation. These sections articulate what inclusive AI means in theory and in practice, and how to make an effective case for moving towards it. In the 'Organisational Change for Inclusive AI' section, the toolkit outlines how to build a strategy for inclusive AI, identifies three implementation gaps towards it, and offers recommendations to overcome them. This section is mainly intended for a non-technical managerial audience. In the 'How to implement Inclusive AI (for technical audiences)' section, the toolkit provides recommendations and an impact assessment questionnaire centered around building Inclusive AI. It explains how to implement Inclusive AI with regards to risk, bias mitigation and representation in development

processes. This toolkit features several examples of Daring Circle partner and non-partner initiatives working towards Inclusion in AI. These initiatives have been selected based on their exemplary characteristics as well as alignment, applicability and relevance to the content of the toolkit.

Hardwiring fairness and inclusiveness into organisations will require new strategies, resources, and critical reflection on how business as usual addresses issues of bias and identity. We recommend that you use this resource in the spirit of challenge and hope.

A snapshot of what's happening in the AI ecosystem:

- The Tech She Can Charter is a commitment by organisations to work together to increase the number of women working in technology roles in the UK.
- The Centre for Equity, Gender and Leadership, Berkeley Haas has released a playbook Mitigating Bias in AI which educates readers about the reasons for bias, its impacts, challenges and strategic plays to address bias.
- The international charter for inclusive AI commits companies that want to help combat bias and stereotypes by promoting diversity and responsible use in the development of artificial intelligence.
- The WEF's Global Future Council on Artificial Intelligence for Humanity is a diverse group comprising experts from business, civil society, academia and government. The council will develop policy and governance solutions to promote greater inclusion of underserved communities in the AI development and governance ecosystem.
- For simple explainers around Artificial Intelligence, check out the A-Z of AI project, led by Professor Gina Neff at the Oxford Internet Institute in collaboration with Google.

Special Mention

Cercle InterElles & The Women's Forum for the Economy and Society - Women & AI Charter

The Women's Forum for the Economy & Society is proud to be collaborating with the Cercle InterElles group as partner of the "Women & AI" Charter to advance the agenda of gender inclusion in the Artificial Intelligence Sector.

The "Women & AI" Charter is a pledge for Companies to build accountable and gender fair AI. This charter consists of a set of seven fundamental principles that enable companies to combat the risks of discriminatory cognitive biases, during development or while using AI-based solutions or devices.

Cercle InterElles also has a Women & AI workgroup which has developed an action plan.

This plan has two main objectives:

1. Raise awareness on this topic within our companies and the technology industry, supporting them in taking action to create accountable and gender fair systems and become a model for the future.
2. Work with the French ecosystem of ethics and AI to ensure that the topic of increasing women in AI development teams is discussed and resolved in companies, research and public authorities.

Cercle InterElles is a professional society that groups the diversity networks of 15 large companies in the technology industry. This network of networks has worked for the last 20 years to support an increase in the number of women in STEM careers.

"AI is holding out a mirror to humanity and pointing out where we should act. We must be brave and look into that mirror."

Maria Axente, Responsible AI and AI for Good Lead, PwC[1]

1'Quote taken from the Women's Forum virtual session "What can inclusive AI do for our economies and societies?" that occurredon the 18th of March 2021.'

EXECUTIVE SUMMARY

This toolkit offers practical guidance on how to use AI in an ethical, inclusive way. It outlines what inclusive AI looks like, how to advocate for it within your organisation, and the steps necessary to achieve inclusivity consistently.

Inclusive AI is defined as "AI which accounts for needs of different groups, including minorities, marginalised and underrepresented groups. The application of Inclusive AI addresses the issue of bias and discrimination with the aim of reducing inequalities, including representation, accessibility and interpretability."

Systems and Processes

In order to inculcate values of inclusive AI into organisations, certain changes must be made to systems and processes, beginning with how teams are built. Design and development teams should contain a mix of disciplines and identities (gender, racial, religious, sexuality, nationality, disability, neurodiversity, generational, and others). The teams and datasets used in the AI design process should be as close as possible to the populations likely to use it or be affected by it. These characteristics should be measured. If progress is slow, quotas should be considered for representation.

There must be systematic mechanisms and checks to detect, avoid and mitigate biased data, processes and outcomes at every stage of algorithm development.

Organisations which are serious about achieving inclusive AI will also embed:

- Consistent dialogue between technical and non-technical roles.
- A supportive ecosystem that allows employees to experiment with new inclusive AI approaches in a 'sandbox' environment.
- Full transparency throughout the organisation about systems' purpose and function so that internal teams can evaluate potential biases with reference to the full range of uses of the technology they are developing.
- A culture which values explainability so that clients and users can see why certain decisions are being made and what goes on behind the algorithm.

Externally, organisations can contribute to the pipeline problem by supporting inclusive initiatives that involve minorities in AI & tech roles.

Cultural change must start with leadership. Leadership must be clear that inclusive AI matters and be engaged with achieving it, such as by setting out an Inclusive AI Strategy, targets with suitable incentives, and defining metrics and milestones along the way.

Technical Teams

Technical teams have a central role to play in building inclusive AI. The teams should understand what representative data should look like, which involves gaining a nuanced understanding of the population to be modelled. They should also assess their data's classes and labels for new biases and ensure that annotation is checked by experts with specific sensitivities and cultural awareness of biases.

Models should be exposed to as wide a variety of real-life environments, contexts, and users as possible before release, and methods for selecting and cleaning data should be documented so that the techniques can be improved.

It is important to foster a flexible environment in which team members can ask questions, share uncertainties and alter practices in light of new information.

Technical teams must employ procedures such as penetration testing, iterative data labelling, and mechanisms to collect and process feedback on their AI applications.

This toolkit includes a checklist of questions to help technical teams consider the variety of issues which could arise at each stage of the process.

The question of how to achieve Inclusive AI is constantly evolving. Every organisation which strives to learn, test, and improve its techniques for avoiding unintended bias would do well to share what it learns with the growing community, so tools, practices, and processes can be replicated and refined.

Achieving Inclusive AI is a journey, not a destination. The Women4AI Daring Circle of the Women's Forum for the Economy and Society proposes this toolkit as a start. It is our hope that as difficult as it may be, the incredible power of technology to improve life will continue to accelerate in the coming years in a way that leaves no one behind, no matter their identity.

DEFINING INCLUSIVE AI

What do we mean by ‘inclusive’ AI?

The first step towards creating 'inclusive' AI is to create a common list of definitions that will become industry-wide benchmarks for this and future discussions. In this spirit, the Women4AI Daring Circle offers the following definitions:

We define ‘inclusive AI,’ along with ‘responsible AI,’ as subfields of ‘ethical AI’.

Ethical AI: As a subfield of applied ethics, Ethical AI is the moral guardrail within which the development and use of AI technologies is to be conducted.

Responsible AI: This subfield seeks to translate the vision of ethical AI into concrete actions which guide the development and use of AI. Responsible AI ensures that a fair and trustworthy approach is embedded within the AI lifecycle.

Inclusive AI: This subfield of ethical AI, accounts for needs of different groups, including minorities, marginalised and underrepresented groups. The application of Inclusive AI addresses the issue of bias and discrimination with the aim of reducing inequalities, including representation, accessibility and interpretability. Inclusive AI is non-discriminatory in its production, unbiased in its consequences, and accessible to all.

“Inclusive AI to me is a product that works for everybody. It needs loads of components. It starts in the researcher's head and takes a long time to get to the product. It's about tools, systems, services, products and applications that work for everybody equally and fairly.”

Tabitha Goldstaub, Co-founder, CogX

THE CASE FOR INCLUSIVE AI

This section offers a guide to making the case for inclusive AI within your organisation. It may be useful to use the following framework.

Start by demonstrating that AI is increasingly ubiquitous, its use is growing and this growth is likely to continue.

Next, present the opportunities associated with ethical and inclusive use of AI.

Further, articulate the risks associated with inaction.

Finally, consider possible reasons why there may be resistance to action on this issue, and plan a reply to each objection.

Also, think about how to make the case that action on this issue is time critical. The section below looks at each of these steps:

- AI is ubiquitous, growing, and growth is likely to continue

Show that the growth of AI represents a wave of inevitable technological change likely to have a direct or indirect impact on your organisation. Naturally, examples which are relevant to your firm and sector in terms of hiring practices, customer experience, sourcing and more will be most useful.

In 2021, for example, one study found that 91.9% of companies surveyed said that the pace of investment in data/AI was accelerating.[1] In a small survey of 821 decision makers at companies and government agencies in the US, over 80% said they are currently using AI-enabled tools such as machine learning, computer vision, and natural language processing.[2]

The AI market is likely to continue to grow rapidly. In the year-to-date 2021, the sector had a compound annual growth rate (CAGR) of 16.4%.[3] The market is expected to achieve a CAGR of 17.5% by 2024, by which time the market is expected to be worth $500 billion.

- Present the opportunities associated with ethical and inclusive use of AI.

Continue your case for action by articulating the opportunities from using AI inclusively. We know that there is already a strong case for diversity and inclusion within organisations. Companies in the top 25% for racial and ethnic diversity are 35% more likely to have financial returns above national industry medians, according to McKinsey.[4] Its research also shows that the more women as executives a company had, the higher its outperformance.[5]

Customers trust companies who they think use AI ethically. A Capgemini survey of consumers found that 62% would place higher trust in a company whose AI interactions are deemed ethical.[6]

Customers increasingly expect companies to have an explicit AI code of ethics. Two-thirds of internet users surveyed by the Brookings Institution felt that companies should have an AI code of ethics and review board.[7]

AI can make decision-making processes fairer. For example, there is evidence that algorithms could help reduce racial disparities in the criminal justice system.[8] Anywhere that there is a decision to be made based on data and good quality data available, AI processes can make faster and better-informed decisions.

- Articulate the risks associated with inaction

Using AI without attention to inclusiveness risks reputation. The path towards effective use of AI shows several examples of companies whose reputations have suffered when their use of AI went wrong. In 2014, for example, one large online retailer started working on a project to pick out the best job applicants using AI software. It developed a system to use AI to give job candidates scores ranging from one to five stars. It used data from previously successful applicants. Unfortunately, this

[1] Businesswire. 2021. The Journey to Becoming Data- Driven: A Progress Report in the State of Corporate Data Initiatives. [Accessed here]

[2] Modzy. 2021. Race towards Artificial Intelligence Adoption. [Accessed here]

[3] IDC. 2021. IDC Forecasts Improved Growth for Global AI Market in 2021. [Accessed here]

[4] McKinsey & Company. 2015. Why Diversity Matters. [Accessed here]

[5] McKinsey & Company. 2021. Diversity wins: How Inclusion matters. [Accessed here]

[6] Capgemini Research Institute. 2020. The AI-powered enterprise: Unlocking the potential of AI at scale. [Accessed here]

[7] West, D. 2018. Brookings's survey finds divided views on artificial intelligence for warfare, but support rises if adversaries are developing it. [Accessed here]

[8] Kleinberg, J. et al., (2017). Human Decisions and Machine Predictions. [Accessed here]

replicated existing biases, such as penalising resumés which included words such as 'women's' (example, 'women's chess club captain,') downgrading graduates of two all-women's colleges, and favouring words such as 'executed' or 'captured' which were more often found on men's resumés. In 2018, the company scrapped the project.

There are similar examples from public policy. In Florida, an algorithm designed to help a court decide on which defendants were most likely to reoffend returned racially skewed results: it labelled African-American defendants as 'high risk' at nearly twice the rate of white defendants.[9] A large online company was also found to accidentally discriminate against non-white speakers in the way it captioned its videos. In short, when AI makes bad decisions, it poses reputational risks. This is one reason to avoid the risks associated with a non-inclusive AI policy.

The 2020s will see legal requirements to address algorithmic bias. Legislation currently being developed will increase pressure on companies to act against biased algorithmic decision making. The European Union recently published a proposal for a regulation which requires companies to undertake a risk assessment on AI applications.[10] It may also introduce a code of conduct. In the US, the Federal Trade Commission has recently warned that companies may face enforcement actions and penalties if their AI reflects gender or racial biases.[11]

Companies which adopt measures early can do so in a considered and timely way, rather than playing catch up after laws are passed.

Consider possible reasons why there may be resistance to action on this issue in your organisation. For each objection, plan a reply. Think also about how to make the case that action on this issue is time critical.

To make the case in your organisation, consider how to reply to the following likely objections.

a) Denial ("Our data does not contain biases") – One research participant made the point that "a lot of technical people aren't aware of the fact that the way we collect data is already a bias, even before starting to code."[12] This is one example of a cognitive blind spot, and a reason others may not be able to see the problem. In order to make the case that your organisation is in danger, you may have to build consensus in the organisation. Are you able to find out how your organisation uses data, where it comes from, how it is collected, and where it uses automated decision making? Are you able to find out where it is considering using automated decisions in the short, medium, and long-term? If so, you stand a better chance of making the case internally that there is a real risk of biased AI use.

[9] Angwin, J. et al., (2016). Machine Bias. [Accessed here]

[10] European Union. 2021. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. [Accessed here]

[11] Jillson, E (Federal Trade Commission). 2021. Aiming for truth, fairness, and equity in your company's use of AI. [Accessed here]

[12] This statement was made by an anonymous research participant during focus group sessions that were held in 2020. These as focus groups that were conducted by the Women's Forum for the Economy and Society in collaboration with Professor Gina Neff.

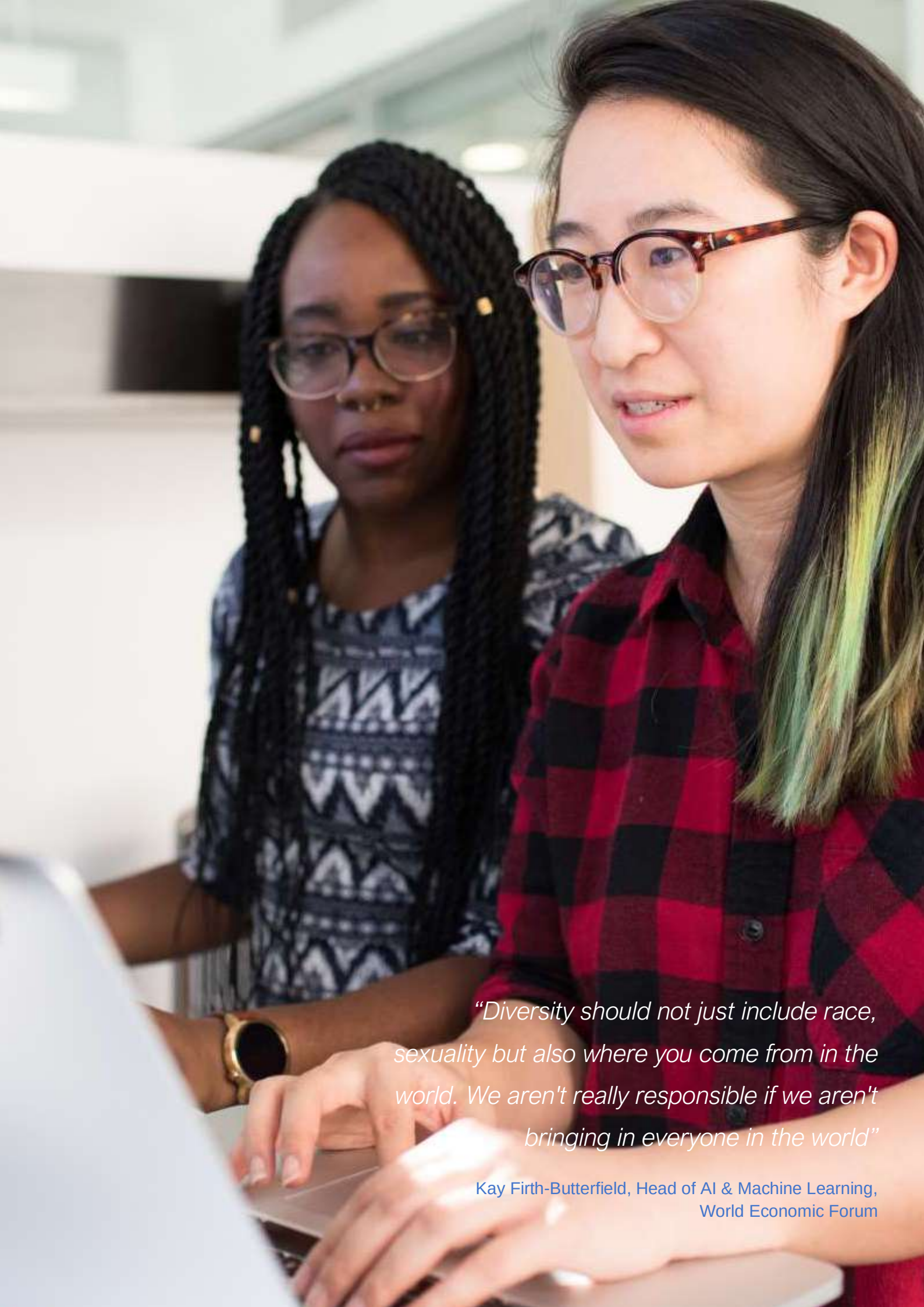
b) Resources ("We don't want to spend time and resource on this.") - While not every decision maker is likely to agree to the costs of addressing this issue in time, it may be possible to persuade others in the company to join you in assessing the risks of inaction, articulating the rewards of action or expressing the strength of feeling inside the company from employees.

c) Mismatched priorities ("The technical team don't see the need for it.") - One research participant said that "technical teams may… fail to imagine ways in which the system they are creating could be improved, or how other systems, tools, or methodologies could be applied to better safeguard and improve human well-being." The solution is greater communication between this team and teams dealing with corporate social responsibility to explain why the issue is important.

d) Timing ("Why now?") - Recent years have exposed the deep rifts in our societies and an increasing consciousness with movements like Black Lives Matter. It is likely that in the coming years, applications will be too challenging to retrofit. Given the pressure for regulation it is likely that legislation will compel action. Over two in five Women4AI 2020 survey respondents believe that AI that is not inclusive will exacerbate inequalities.

e) Uncertainties around implementation ("We don't know how.") – Few companies will oppose more inclusive AI in principle, but most don't know how to achieve it in practice. How can you bridge this gap? Could you, for example, suggest an internal audit of datasets or research how to build capacity within the company on spotting and alleviating biased data, processes, or outcomes? Could you approach a budget holder to make the case to fund a pilot of a programme to acquire these skills?

The case for inclusive AI is ultimately about helping your organisation avoid the risks from biased outcomes and reap the rewards from economies and societies which increasingly expect inclusion as standard.



“Diversity should not just include race, sexuality but also where you come from in the world. We aren't really responsible if we aren't bringing in everyone in the world”

Kay Firth-Butterfield, Head of AI & Machine Learning, World Economic Forum

ORGANISATIONAL CHANGE FOR INCLUSIVE AI

This section of the toolkit offers advice on turning aspirations for inclusive AI into practical policies and lasting change. Most organisations find this a challenge. Our research, in partnership with the Oxford Internet Institute and PwC, finds that despite good intentions, at present most companies fall short of their own aspirations in three distinct ways. These three implementation gaps are elaborated upon more in our joint forthcoming academic journal article publication (Collett, Neff, et al., forthcoming).[13]

The Engagement Gap – Organisations often find it hard to effectively engage leaders and employees on issues of bias and discrimination. This lack of engagement at an organisational level becomes a barrier to effective implementation of inclusive AI.

The Translation Gap – While organisations usually have a set of principles and policies related to Ethical and Responsible AI, they often find it difficult to translate them into practice.

The Dialogue Gap – Organisations struggle to navigate and set up effective systems for communication, feedback, and idea sharing. This is crucial because it leads to a lack of communication between those developing and implementing AI.

This section first addresses systems and practices which enable inclusive AI, and then aspects of organisational culture. These recommendations allow users to look at various aspects of Inclusive AI, measure their progress and determine their best next steps. Please bear in mind the sector of the organisation while developing the strategy as there are sectoral differences that play a major role in AI development and deployment. Organisations are responsible for their own choices with respect to usage of the recommendations.

[13] The Women's Forum for Economy and Society focus groups, which took place with members of the Women4AI Daring Circle, were facilitated by Professor Gina Neff in October 2020. Analysis of the focus group material by Clementine Collett and Professor Neff revealed these three implementation gaps. The ways in which these implementation gaps intersect with inclusive AI have then been explored through the Women's Forum for Economy and Society's partnership with the Oxford Internet Institute and PwC. More about the three implementation gaps to inclusive AI will be published as a joint academic journal publication between the Oxford Internet Institute, The Women's Forum for Economy and Society, and PwC (forthcoming)

Systems and Practices

Teams

- Diversity - Design and development teams should contain a mix of disciplines and identities (gender, racial, religious, sexuality, nationality, disability, neurodiversity, generational, and others). The teams designing the AI should be as close as possible to the populations likely to use it or be affected by it.
- Measurement - Measure these groups. It is critical not just to have gender-disaggregated data but also quantitative measurement of minorities involved in development and deployment processes to managing figures and moving towards diversity.
- Quotas - If the teams are not yet representative, develop quotas for all characteristics your organisation needs to increase. These quotas should be used until they are no longer necessary because representation has been achieved. Although it may not be comfortable, it is effective. Quotas can also be applied in recruitment, training and upskilling programmes. Quotas are obviously not appropriate in jurisdictions where their use is proscribed by law.

Datasets

- Representation - Use datasets that are representative of the population/geography where the AI is used.
- Testing - Establish mechanisms to test for bias and other errors when selling AI to international users. This is key because international users might not resemble the dataset on which the AI was originally based so the chance of non-inclusive results is high. For more advice on using and testing data, see 'how to implement inclusive AI (for technical audiences)' section.

External

- Pipeline - When it comes to ensuring teams include diverse members, the industry faces a pipeline problem. Be part of the solution by supporting inclusive initiatives such as Women in AI, Black in AI.
- Collaboration - Encourage collaboration between companies, customers and regulators to accelerate the deployment of Inclusive AI strategies.

Training

- Programmes - Set up training programmes on how to detect, avoid, and mitigate non-inclusive AI.
- Reskilling - Build and scale programmes to redeploy and reskill the workforce, particularly women and older people. This increases the pool of qualified employees and so makes it easier to build diverse teams, which increases the chance of catching non-inclusive AI in time to avoid it.

Checks and balances

- Checks - Design systematic checks to avoid biased data, processes and outcomes. For example, set up processes to check for bias at every stage of algorithm development, from putting in place ethical frameworks to clear rules for data labelling. A useful resource for organisations involved in creating AI algorithms is UC Berkeley's 'Mitigating Bias in AI' toolkit. For more on this, refer to the chapter on 'how to implement inclusive AI (for technical audiences)'. When procuring AI from partners and vendors, checks must be in place to ensure the absence of bias.
- Balances - Design balances: centres of power within the organisation which are empowered to push back on decisions likely to lead to bias.
- End-to-end - These checks and balances must comprise established standardised processes and power centres covering every stage of the AI lifecycle.

Organisational Culture

Leadership

- It must be clear throughout the organisation that Inclusive AI is important to its leadership. The leadership must understand it, want it, and be engaged with progress to achieve it. This can be evidenced in:
    - The existence of an Inclusive AI Strategy.
    - A C-suite individual incentivised to make progress on inclusive AI against defined metrics (e.g. diverse teams, gender pay differentials, data discrimination tests run rather than skipped).
    - Participation of all stakeholders in the AI lifecycle in terms of process and training.
    - Embedding AI Inclusivity into procurement conditions.

Continuous improvement

- Eliminating non-inclusive AI is not the kind of problem which can be achieved once, then forgotten. Algorithms naturally reflect the world around them, and social understandings of fairness and bias continually evolve, so an inclusive strategy requires continuous improvement in system, culture, mindset and skills to ensure vigilance, and sensitivity. This is not a contingent part of an inclusive AI strategy; it is the foundation on which the strategy is built.

Dialogue between technical and non-technical roles

- Technical teams should be in constant dialogue with 'non-technical' roles about the impacts of what they are building, corporate responsibility, social and ethical impacts, and how to continually embed these concepts into the design process.

Experimentation

- Ensure a supportive ecosystem that allows employees to experiment with new inclusive AI approaches in a 'sandbox' environment.

Transparency

- Full transparency throughout the organisation about systems' purpose and function is essential. Internal teams can only evaluate potential biases if they understand the full range of uses of the project on which they are working.

Explainability

- The technology must be understandable to end-users and all organisations who will deploy it. Clients and users need to be able to see why certain decisions are being made and what goes on behind the algorithm. Inclusive technology requires an organisation which embeds the skill of explaining the technology, the culture of doing so, and processes that allow users to send practical feedback to systems and acts upon received feedback.

HOW TO IMPLEMENT INCLUSIVE AI (FOR TECHNICAL AUDIENCES)

This section offers tools and suggestions for technical teams to avoid developing systems which are not inclusive. It is primarily written for a technical audience but potentially useful for a non-technical audience too - as they build recommendations into internal processes for inclusive AI and set appropriate incentives.

The first part contains a list of steps technical teams can take to implement Inclusive AI, and the second is a questionnaire to help catch and mitigate potential mistakes which can lead to non-inclusive outcomes.

Steps Towards Inclusive AI

There is no fixed recipe for avoiding errors, biases, and biased outcomes. Rather, there are steps which offer a better chance of identifying them. The following is a non-exhaustive list of such steps. If you think there are other important steps which should be included in a future iteration of this document, we would be happy to hear from you. Write to daringcircles@womens-forum.com with your ideas and suggestions.

Data

Understand what representative data should look like. That means looking critically at your understanding of the population that is to be modelled. Ask the following questions about the variety of subgroups within it: Does your data represent that variety? If not, can it be adjusted accordingly?

Assess your data's classes and labels. Datasets can include classes and labels that introduce new biases. As Appen's guide to reducing bias explains, "one option to reduce this is to source from a global crowd of annotators, who can not only provide a difference of perspectives, but also support a variety of languages, dialects, and geographically specific content.[14]"

Another option is to ensure that annotation is checked by experts or campaigners with specific sensitivities and a nuanced cultural awareness of biases, such as gender liaison experts. At minimum, follow the team guidance below. You can then run a penetration test on your data to test for bias. This involves trying to reconstruct protected characteristics which have been removed from training data. You can find some additional guidance here.[15]

[14] Appen. 2021. How to Reduce Bias in AI. [Accessed here]

[15] Thomas. 2019. How can we eliminate bias from AI algorithms? The pen-testing manifesto. [Accessed here]

Team

Diverse teams bring wider awareness to the task of data annotation. They are more likely to perceive and flag problems. To maximise the chance of uncovering biased annotation, diversity should represent as wide as possible a range of characteristics: gender, racial, religious, sexuality, nationality, disability, neurodiversity, generational, and others. If the annotation involves language, the team will need to consider various languages, cultural knowledge, dialects, and geographically specific content. Certain members of a team may know that words are used in multiple ways in different contexts. For example, the symbol 'X' might be seen to mean either 'X' or '10'. But if it results in a map which mislabels 'Malcolm X' as 'Malcolm 10,' the lack of context will cause insensitive results. While having a diverse team is no guarantee that biased datasets of annotation will be caught, it is a good first step.

Assume Iterative Data Labelling

However good your data labelling, blind spots are inevitable, and they will be uncovered eventually by users. It is therefore good practice to assume that data labelling will evolve as you test, validate, and learn from outcomes. Data labelling is iterative and being open and public about the need to refine it over time – and perhaps even asking for help – is advisable.

Flexibility and Ownership

Because the quality and sensitivity of annotation will directly affect the outcome of your project, it will benefit from being done by a team who are invested in your project and how it is perceived, so that they can take time and care to think about how various labels are likely to be perceived by diverse groups. Similarly, a team that has the flexibility to alter their practice in the light of new information or context are more likely to produce inclusive applications and outcomes.

Communication Processes

A diverse team will only be useful if the benefits of their common wider-cultural knowledge are shared within the team. Therefore, it is essential to foster an environment in which team members can ask questions and share uncertainties within the team and receive timely and sensitive responses. This requires some process so that a culture of questions and deliberation can be maintained.

Create a Culture of Reporting and Accountability

Linked to the need for communication processes is the need to foster a culture in your organisation of reporting and accountability, in which everyone is encouraged to report issues and help each other solve them. This increases the chance of finding and addressing biased data, algorithms, systems or outcomes.

Testing

Most examples of biased outcomes are uncovered by diverse users finding that in practice, an application worked differently from how it was intended to work. You can avoid this by maximizing the chance that these biases are discovered before the product is released or used in the real world. Expose your model to as wide a variety of environments, contexts, and users as possible before releasing it. What steps can you take to make the testing environment more similar to a wide variety of real-world scenarios and environments? Can you do more to test it with users with different identities, in different places? Run your statistical methods against real data wherever you can.

Document and Share How Data is Selected and Cleaned

It is not always possible to ensure that your team contains people with a wide variety of characteristics. Even when it does, a diverse team will not always notice all potential biases. The solution is to make your processes transparent so that interested parties beyond your company or organisation can offer feedback and help iterate. Document how you select and clean data, publish your results, and invite others to respond. Transparency enables a wider pool of people to help your team find issues in your data or process, and shows that your organisation is keen to learn, improve, and take steps to achieve inclusive AI, even when doing so may cause internal discomfort. Discussion forums and social media may be useful here to amplify your reach and seek a wider variety of responses.

Inclusive AI Impact Assessment Questionnaire

This questionnaire provides a checklist of issues to consider before 'signing off' on internal processes. It is designed to encourage reflection. After all, a wide variety of issues could arise at every stage of the process.It also offers a series of topics for dialogue between internal and external stakeholders on inclusive AI. There is no need for all questions or areas to be addressed all at once; rather, this is designed to offer a prompt for questions to build into internal processes.

It is necessarily non-exhaustive; no questionnaire could provide a complete measurement of Inclusive AI practices, given the wide sectoral differences in development and deployment. Developers are of course responsible for their own choices. We hope, however, that this questionnaire is a useful tool and prompt.

Datasets and Representation

1. Does your AI product or service (APS) use training data or datasets represent a diverse set of users?
2. Does your dataset truly represent the actual environment or data that the APS will be deployed in?
3. With respect to bias in datasets, does your team:
    1) Know about the different types of biases (such as stereotype bias, omission bias, reinforcement bias)?
    2) Have mechanisms to mitigate these biases?
    3) Have a process to avoid biases resulting from datasets from sensors and cameras? (These are notoriously prone to bias by virtue of capturing data from certain spaces or areas but not others).
    4) If your team uses machine learning and deep data sets, does it mitigate historical biases or gaps?
    5) Does your team work on periodically auditing and updating its datasets and APS to ensure reliability and accuracy?
    6) Does your team have measures in place to evaluate the need for additional data (for improvement of accuracy or elimination of bias)?
    7) Does your team create subsets for a large dataset to mitigate the risk of systematic bias during validation of the APS model?
4. Does your team test the APS with a diverse set of users representative of demographic factors pertaining to gender, age, race, geography, socioeconomic status, language, sexual orientation and religion?
5. Can your APS be used by those with special needs, disabilities or risk of exclusion? Will it work differently?
6. Does your team have mechanisms in place to make APS and algorithms transparent to the people that they assess and/or impact? Are there mechanisms for feedback and recourse for people who feel they have been unfairly or incorrectly assessed or impacted?

7. Are there mechanisms in place to allow users, suppliers and vendors to report issues related to discrimination and bias?
8. Are there mechanisms to audit APS that have been procured externally for bias and discrimination-based vulnerabilities?

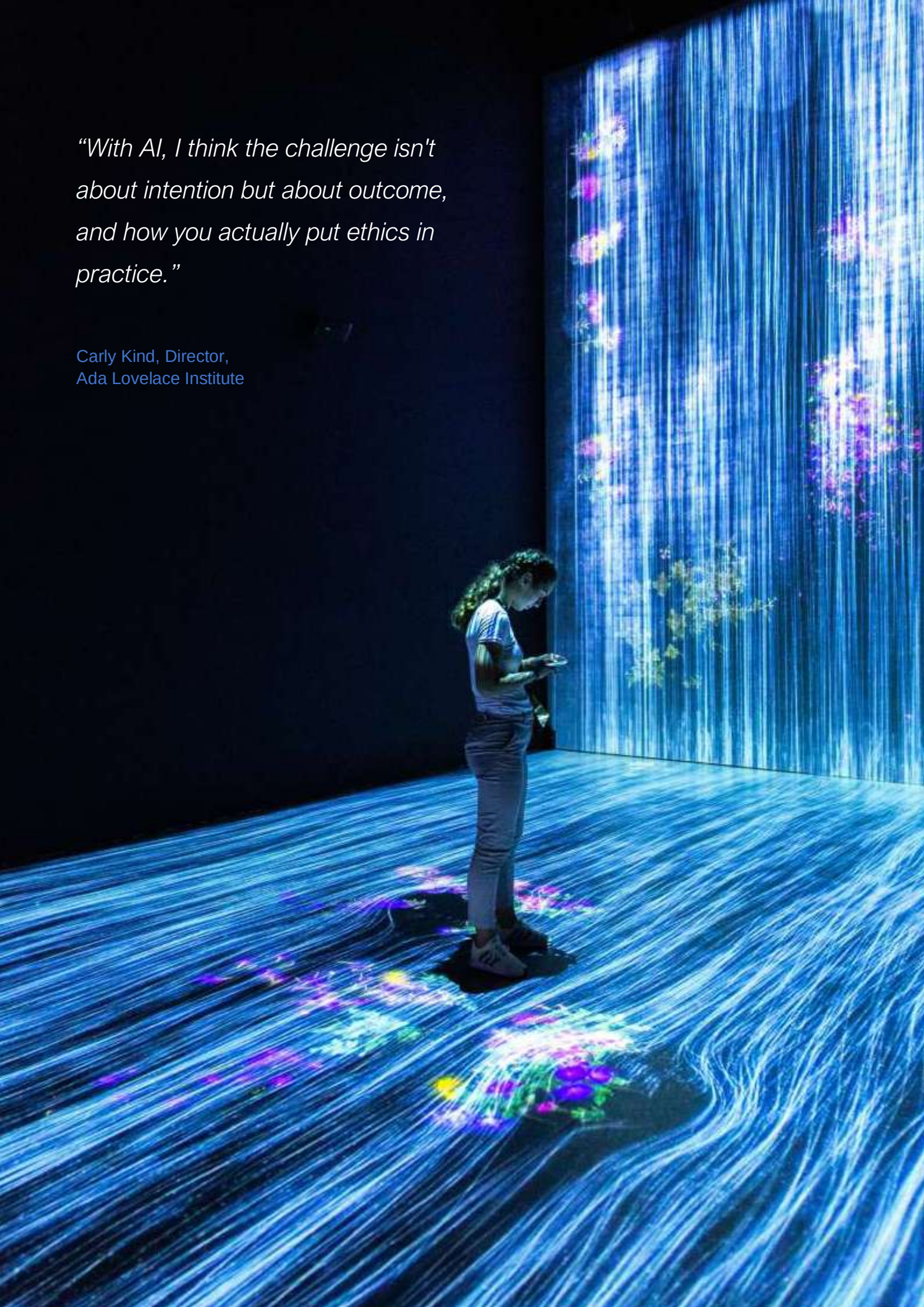
Use

1. Can a bad actor use your APS to spread misinformation, and/or create political distrust or social unrest?
2. Can your APS harm someone directly or indirectly?
3. Could APS enable a malevolent actor to commit a crime? Is there more you can do to create stronger barriers to your APS being used to commit a crime, or conversely, to increase the chance of detection? Have you considered a wide range of crimes, such as fraud, money-laundering, and enabling terrorist financing?
4. Can groups use your APS to spread hate, recruit members, or discriminate against others?

If you have further suggestions for useful questions which could be added to this list, please feel free to suggest them at daringcircles@womens-forum.com.

We recommend the following non-exhaustive list of tools for technical audiences that seek to improve the ethical and inclusive practices of AI systems:

1. Audit AI
2. IBM AI Fairness 360
3. Microsoft's Fairlearn
4. Microsoft's Responsible Innovation Toolkit
5. PwC's Responsible AI toolkit
6. Google's Model Card Toolkit



“With AI, I think the challenge isn't about intention but about outcome, and how you actually put ethics in practice.”

Carly Kind, Director,
Ada Lovelace Institute

CONCLUSION

Each year, awareness of the need for Inclusive AI rises. With each news story about a tech product which unintentionally discriminates against a group because of the colour of their skin, gender, accent, culture, or any other aspect of their identity, promoting inclusivity in AI moves up the social, political, and business agenda. Alongside the growing awareness, we should aim to bring inclusivity to the way we use AI to create a positive impact within and beyond our organisation.

AI has tremendous potential to improve our society and to remedy the shortcomings in our current ways of functioning. However, we can only unleash this potential with intention and investment. To do so, we need to embed inclusivity into all aspects of digital transformation, right from the outset. Companies should measure where they are on their AI journey, define their targets and go through changes in organisational structure to develop an effective strategy for AI.

This toolkit has aims to provide both organisational leadership and technical teams with a clear approach to developing principles, priorities, processes and policies that bring more inclusivity into the applications of AI. The Women4AI Daring Circle will be releasing a Call to Action, a playbook of case-studies and other publications regarding the usage of AI for societal benefit in the coming year to complement the work done so far.

"In the AI industry, there is a need for diversity and women at the table. Having girls, quotas, affirmative actions and no male-only teams is a must."

Gabriela Ramos, Assistant Director-General for the Social and Human Sciences, UNESCO [1]

[1] 'Quote taken from the Women's Forum virtual session "What can inclusive AI do for our economies and societies?" that occurred on the 18th of March 2021.'

ACKNOWLEDGEMENTS

This report was authored by Ela Stopford Sackville, Anusha Raghav and Sophie Lambin, Editorial Partners of the Women's Forum and Matteo Cadenazzi of the Women's Forum, under the oversight of Audrey Tcherkoff, Managing Director.

The research underpinning this work and production of this report was also greatly supported by the Daring Circle's Knowledge Partner PwC and its Academic Partner Oxford Internet Institute. Special thanks to Maria Axente (Responsible AI and AI for Good lead, PwC), Professor Gina Neff (Professor of Technology & Society, Oxford Internet Institute), Clementine Collett (PhD student, Oxford Internet Institute) and Frankie Howland (Senior Associate in Sustainability & Climate Change, PwC).

The Women's Forum would also like to extend its thanks to the following partner representatives and experts for providing their insights to this research:

- **Kay Firth-Butterfield,** Head of AI & Machine Learning and Member of the Executive Committee, World Economic Forum
- **Carly Kind,** Director, Ada Lovelace Institute
- **Tabitha Golstaub**, Co-founder, Cog X
- **Lily Xu,** PhD student, Harvard University
- **Clara Clemente Langevin,** Project Manager, Artificial Intelligence and Machine Learning Platform at the Center for the Fourth Industrial Revolution Brazil
- **Alice Knowles,** Associate, Clifford Chance
- **Ana Trevoux, Legal Officer,** Conseil National des Barreaux
- **Angela Zutavern,** Managing Director, AlixPartners
- **Ann Massal,** Senior Advisor, L'Oréal
- **Claire Calmejane,** Chief Innovation Officer, Société Générale
- **Dr. Daniela Victoria Grohmann,** Head Inclusion & Diversity Bayer Pharma, Executive Board Support, Bayer
- **Dr. Lobna Karoui,** Founder and Chief Empathy Officer, AI Exponential Thinker
- **Eloise Lehujeur,** Education and Edtech Officer, Numeum
- **Elspeth McOmish,** Programme Specialist, Gender Equality Division, UNESCO
- **Faidra Stavropoulou,** Senior Data Scientist, Bayer
- **Genevieve Smith,** Associate Director, Center for Equity, Gender and Leadership, University of California Berkeley, Haas School of Business
- **Georgina Court,** Global Inclusion Specialist, Clifford Chance
- **Helena Yin Koeppl,** Global Head of Data Science and Analytics, Bayer
- **Joanne Pupo,** Head of Digital Engagement, AXA
- **Katya Lainé,** Board Member and President AI Commission, Numeum
- **Lauranne Poulain,** Déléguée au Numérique Responsable / Relations Institutionnelles, Numeum
- **Lea Deleris,** Head of RISK Artificial Intelligence Research, BNP Paribas
- **Malgorzata Zielinska,** Diversity & Inclusion, Wellbeing Manager, AXA
- **Marine Rabeyrin,** EMEA Business Development Manager, Lenovo
- **Mohamed- Ramzi Temanni,** Scientific Director, Head of France AI & Genomics , The Janssen Pharmaceutical Companies of Johnson & Johnson
- **Nancy Rowe,** Head of Inclusion and Diversity, Publicis Sapient
- **Natalia Yakimiuk,** Senior Consultant, FTI Consulting
- **Olga Granaturova,** Co-Founder, Brighter Futures
- **Olivia Jane Baxter,** Data Science Officer, BNP Paribas
- **Rim Tehraoui,** Chief Data Officer, BNP Paribas
- **Sasha Rubel,** Programme Specialist, Digital Innovation and Transformation, UNESCO
- **Sue Glueck,** Senior Director of Academic Relations, Microsoft
- **Virginie Lasserre,** External Affairs Leaders, The Janssen Pharmaceutical Companies of Johnson & Johnson
- **Zaïnaba el Khald,** Data & AI Culture Leader, SociétéGénérale

The Women4AI Daring Circle

The Women4AI Daring Circle focuses on the design of Artificial Intelligence (AI) that drives inclusion, both in its development and its application. We need to create an environment where women are empowering AI, AI is empowering women, and where inclusive AI is implemented into organisations' AI practices. Our work in the Daring Circle is focused on facilitating this ecosystem transformation; through inspiring companies to act and pledge their commitment to inclusive AI whilst adopting effective policies to drive these commitments forward.

We aim to create an ecosystem where AI is designed by diverse teams and governed appropriately, so that AI can become a powerful asset to right existing disparities and shape a more inclusive world. Since Artificial intelligence has a growing influence on the way we work and live, AI will only fulfil this potential if everyone has an equal share in its development and adoption.

The Daring Circle is led by Microsoft in collaboration with L'Oréal, BNP Paribas, Publicis Groupe, AXA, Bayer, Lenovo, Société Générale and ExxonMobil. UNESCO participates as an Institutional Partner, Oxford Internet Institute as an Academic Partner, PwC and Shearman & Sterling as Knowledge & Insight Partners.