WOMEN LEADING IN AI

10 PRINCIPLES OF RESPONSIBLE AI

Contents

Briefing note

Artificial Intelligence (AI) is changing the way we live and work – and it is mostly for the good.

Algorithms are at the heart of AI and are very useful tools to automate decisions and free up humans to do work needing our creativity and discretion. But we are still seeing many algorithms that discriminate against women and ethnic minorities. Meanwhile, subservient female virtual assistants are the default interface for consumers' interactions with machines.

How did we get to this backwards future? And how do we alter our course so that AI helps us build a better society?

In this report, we set out 10 recommendations for government to regulate Artificial Intelligence and drive its development.

They have been developed by the Women Leading in AI Network, whose members are women from all walks of life, including leading AI scientists, algorithm coders, privacy experts, politicians, charity sector leaders and academics. The aim of the Network is to **mobilise politics**, so we can build an AI that supports our human goals and is constrained by our human values.

Our ten recommendations are:

1. Introduce a **regulatory approach governing the deployment of AI** which mirrors that used for the pharmaceutical sector.

2. Establish an **AI regulatory function** working alongside the Information Commissioner's Office and Centre for Data Ethics – to audit algorithms, investigate complaints by individuals, issue notices and fines for breaches of GDPR and equality and human rights law, give wider guidance, spread best practice and ensure algorithms must be fully explained to users and open to public scrutiny.

3. Introduce a new '**Certificate of Fairness for AI systems'** alongside a 'kite mark' type scheme to display it. Criteria to be defined at industry level, similarly to food labelling regulations.

4. Introduce mandatory **AIAs (Algorithm Impact Assessments)** for organisations employing AI systems that have a significant effect on individuals.

5. Introduce a mandatory requirement for **public sector organisations** using AI for particular purposes to **inform citizens** that decisions are made by machines, explain how the decision is reached and what would need to change for individuals to get a different outcome.

6. Introduce a '***reduced liability***' **incentive for companies** that have obtained a Certificate of Fairness to foster innovation and competitiveness.

7. To compel companies and other organisations to **bring their workforce with them** – by publishing the impact of AI on their workforce and offering retraining programmes for employees whose jobs are being automated.

8. Where no redeployment is possible, to compel companies to make a contribution towards a **digital skills fund** for those employees.

9. To carry out a **skills audit** to identify the wide range of skills required to embrace the AI revolution.

10. To establish an **education and training programme** to meet the needs identified by the skills audit, including content on data ethics and social responsibility. As part of that, we recommend the set up of a solid, courageous and rigorous programme to **encourage young women and other underrepresented groups into technology**.

Introduction

Our premise is that innovation and AI are primarily a force for good.

AI holds huge potential for all sectors of the economy and we are increasingly reliant on it.

From improving distribution and logistics to supporting national defence, AI is already delivering huge benefits to society by helping reduce costs, increase efficiencies and improve reliability. For example, AI systems can lower diagnostic errors by 85% in breast cancer and AI cybersecurity can reduce the average time to neutralise attacks from 101 days to a few hours.

But we are also acutely aware that delegating tasks and choices to AI can and does go wrong. Amazon deployed a software that discriminates against women[1]; Google searches showed black women when looking for 'unprofessional hairstyles'[2], Facebook showed certain job advertisements only to men[3]; facial recognition software was unable to recognise black women's faces with the same accuracy as any man's.[4] Citizens are losing jobs and access to vital services such as loans, mortgages and insurance based on murky and unaccountable criteria. Now more than ever, ethics and fairness have become key for policy makers and citizens.

In 2018, a network of women from different backgrounds identified a policy vacuum which needed filling and created *Women Leading in AI (WLinAI)*, meeting regularly to identify key issues within the tech industry and the actions needed to mitigate potentially negative impacts. Our members feel empowered by technology and believe in its new benefits; we want to see these benefits distributed equally, creating a fair system that embodies gender rights and equality. With recent research showing a worrying lack of diversity at senior levels of the technology sector, the voices of women in this debate are especially necessary and urgently required.[5]

---

[1] https://www.reuters.com/article/us-ama5zon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

[2] https://www.theguardian.com/technology/2016/apr/08/does-google-unprofessional-hair-results-prove-algorithms-racist-

[3] https://www.theguardian.com/technology/2018/oct/28/how-target-ads-threaten-the-internet-giants-facebook?CMP=Share_AndroidApp_Email

[4] https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html

[5] https://www.theguardian.com/technology/2018/nov/14/worrying-lack-of-diversity-in-britains-technology-sector-race-gender-report-finds

In this paper, we set out ten recommendations that legislators should embrace to make AI work for everyone, and to ensure that it promotes equality rather than amplifying embedded and often ancient stereotypes that hold back both society and the economy.

The Women Leading in AI Network

This paper stems from research carried out by the WLinAI Network, with insights from technical experts, as well as contributions to the debate from a variety of women representing relevant organisations, think tanks, policy groups and forums.

We bring together the wealth of expertise in the wider AI field, recognising that amidst the heat surrounding ethics, practical, concrete steps must be undertaken now. As new examples of bias continue to emerge, the fast pace of tech innovation means governance is required urgently.

The WLinAI Network was established in May 2018 to achieve the following objectives:

Equality

- Bring more women into the tech field by providing role models and champions;
- Foster a space for women to proactively share ideas;
- Encourage women in tech to grow both professionally and personally;
- Create alliances between BAME and other minority tech groups and forward thinking leaders to ensure AI works for all;

Policy

- Create cutting edge policy proposals regarding the increased use of AI in society;
- To move Ethics well beyond just fixing algorithms, to define what AI should and should not be used for in the bigger picture;
- Investigate governance models for the deployment of AI

Fairness

- Ensure AI does not amplify stereotypes and reinforce prejudices;
- Define design values to avoid AI mirroring existing power imbalances;
- Evaluate and develop policies to mitigate the impact of AI on jobs, especially in areas that harm women the most.

The Need for a Regulatory Approach

Given the increasing pervasiveness of artificial intelligence in our society, it is legitimate and necessary to debate how new technologies ought to be shaped and implemented to best support our democratic values and improve our working life, health and wellbeing.

AI has developed as an extension to the digital Internet economy. As such, its development reflects the extent to which the Internet economy is shaped by the dominance of trans-global corporations. It is therefore unsurprising that the development of AI is dominated by precisely the same ecosystems surrounding the six mega corporations, namely Apple, Google, Facebook, Amazon, Alibaba and Baidu.

We agree with Joanna Bryson that the greatest challenges of appropriately regulating artificial intelligence (AI) are social rather than technical[6].

As a first step, it is essential to bring the governance, future and shaping of AI into the public realm where it should belong.

We welcome the proliferation of stated self-regulation by companies including Microsoft, Google, IBM and others. Such initiatives are worthy of mention and valuable. However, the manner by which AI shapes the world must become a matter of public governance in a new partnership between governments, public and private sectors, and academia.

Undoubtedly, those who own the AI have power, be it at a nation state or commercial level. From a geopolitical standpoint, the race between countries to be leaders in AI technology demonstrates this. China and the US are leading the way, and the consequences of their actions will shape the future geopolitical hierarchy. National AI programmes boost scientific and technological research, potentially leading to supremacy in areas such as energy production and military weaponry.[7]

Technological ownership will also reshape global economic leadership, as countries are affected by job automation to differing degrees. If alternative systems of corporate taxation are not implemented, then the ability to mitigate against the likely social turmoil caused by large scale unemployment and decreased social services greatly decreases.

---

[6] How Society Can Maintain Human-Centric Artificial Intelligence, Joanna J. Bryson and Andreas Theodorou, http://www.cs.bath.ac.uk/~jjb/ftp/BrysonTheodorou-HumanDraft18.pdf

[7] https://www.ianhogarth.com/blog/2018/6/13/ai-nationalism

A regulatory approach to AI is not a battlefield for regulators to stay the pace of innovation; on the contrary, regulation must not stifle innovation, it must foster it. As the applications of machine learning (the basis of much AI technology today) increase, the interaction between private companies and government will be transformed. Autonomous vehicles are an example of this: what will happen to urban infrastructure when buses face competition from shared autonomous Ubers and, who will manage these new interactions when they inevitably arise?

Ensuring AI is for the common good (as stated in most corporate manifestos globally) is paramount; however, relying on self-regulation is not enough to achieve such an aim.

If we do not ensure a robust framework is in place now, we run the risk of turning algorithms into policy makers, thus allowing them to dictate our life, work and interactions.

With the increased pervasiveness of AI, regulation will ensure AI strengthens the values of equality, democracy, as well as human rights. The case for pragmatic, effective and proactive regulation is overwhelming[8] .

We therefore call for the regulation of AI comprising of, but not limited to:

- The establishment of a **regulatory function to support and work alongside the Information Commissioner Officer (ICO) and the Centre of Data Ethics** to
    - oversee complaints around significant effect of algorithms on individuals;
    - perform ethics audits on companies using algorithms for their decision making as well as digital advertising and any process which has significant effect on citizens, including price discrimination;
- The establishment of '**certificates of fairness**'[9] to be issued to companies that undertake an audit and follow the processes set up at industry level.
- In view of the need to grow AI and invest in its development, we recommend that the certificate of fairness grants companies a **'*reduced liability*' incentive** in relation to liability for inadvertent errors within the system. This is in recognition of the fact that it is necessary to foster innovation and increase its speed whilst providing a safe regulatory framework.
- The introduction of **mandatory Algorithmic Impact Assessment (AIA)** for all algorithms significantly impacting the data subject, available for scrutiny by the

[8] https://ainowinstitute.org/AI_Now_2018_Report.pdf

[9] Equality and Privacy by Design: ensuring artificial intelligence is properly trained and fed: a new model of AI Transparency & Certification as Safe Harbour Procedures, Shlomit Yaniski-Ravid and Sean K. Hallisey, Fordham Law CLIP, AI-IP Project

public and the regulator; we recommend that recommendations for AIAs should be issued at industry and sector level at the earliest convenience. This is in recognition of the fact that sectors may be profoundly different: health, for example, presents significantly different challenges from digital advertising.

- When defining **the significance of the impact on individuals**, we recommend the **broadest possible approach** thus incorporating digital advertising, price tailoring and similar activities that although may not qualify as a yes/no decision still have an impact on the individual. In particular, this would help address issues related to the governance of the use of inferred data for the purpose of digital advertising. We are, in fact, extremely concerned by the fact that unaccountable algorithms in digital advertising mean that there is an increasing use of personal and inferred data, 'which also creates greater opportunity to manipulate and control'[10].
- The establishment of an **official procedure for individuals to challenge the outcomes or decisions devised by an AI system**, including a detailed list of what different information would have triggered a different outcome.

[10] https://medium.com/s/2069/a-vision-of-the-dark-future-of-advertising-40347c6ed448

Right to Algorithmic Explainability and a Duty of Transparency

As discussed in the previous recommendation, governance of AI applications by due process[11] will reflect our common good and societal values, and, to this end, we need to be able to understand the systems utilising AI and their impact.

Algorithmic systems are promoted by advocates as logical and objective. However, far from neutral, these systems contain bias, prejudice and opacity where logic and objectivity are hidden behind the machine learning. As AI is becoming ubiquitous, it is important that safeguards are in place to ensure fairness, transparency and the ability to challenge a machine decision.

Different countries are approaching the risks of automated decision-making in a variety of ways, for example the European Union's GDPR, has begun to provide a solution via a variety of tools. GDPR provides for a right not to be subject to automated decision making, as well as to access and receive meaningful information about the logic, significance and envisaged effects of the automated decision- making processes. It also contains several safeguards and restraints for limited cases in which an automated decision is allowed.

The key outstanding issue however is whether the focus on the right to challenge an automated decision once its already been made (ex post) is correct, or whether we should hold the right to be given information in advance of an automated decision being made (ex ante).

In our view, it's crucial that information is made available in advance of an automated decision being made.

In itself, the concept of 'explainability' is somewhat opaque. Whilst a large number of policy makers and advocates agree the importance of the explainability of AI, in practice it means different things to different people. Furthermore, the likelihood of explainability is intertwined with the context in which the algorithm is operating.

We recommend that explainability is defined as **the right for the individual to understand the implications of the system** (some scholars refer to this as a right to 'legibility'[12]) which, in turn, places an obligation on organisations to make their 'workings out' transparent.

---

[11] https://doteveryone.org.uk/wp-content/uploads/2018/10/Doteveryone-Regulating-for-Responsible-Tech-Report.pdf

[12] Why a Right to Legibility of Automated-Decision Making Exists in the General Data Protection Regulation, International Data Privacy Law, Volume 7, Issue, November 2017

We recommend that individuals should be able to understand the importance and implications of algorithmic data processing, and that ensuring explainability should be made a requirement for all organisations.

This would entail:

- A duty placed on organisations to inform citizens when an algorithm is being deployed, alongside information related to what would need to change in order for a different outcome to be achieved;
- A duty to inform citizens of their right of erasure, access, portability and rectification in order to mitigate the adverse risks of automated decision models;
- A duty to inform citizens of the measures that have been implemented to promote equality and human rights and avoid bias.

The key problem with GDPR is that a 'new right to explanation' is mentioned in a Recital[13], not an Article, of GDPR and, thus, it is not binding[14]. While the main body of the law pinpoints only the right to contest a decision, the European Data Protection Board suggests that the data subject 'will only be able to challenge a decision or express their view if they fully understand how it has been made and on what basis'. In line with this, we recommend a duty on organisations using algorithms to:

- Carry out regular quality assurance checks against discrimination and unfair treatment;
- Carry out algorithmic auditing (performed by third parties and randomly by the new regulatory function working within and alongside the ICO and the Centre for data ethics);
- Ensure contractual assurance is in place for third party algorithms;
- Establish a structured mechanism for human intervention;
- Establish a duty on organisations to sign up to a code of conduct and ethical review boards. We envisage these to be best placed at industry / sector level due to the differences across areas.

We recommend that individuals have a right to understand the algorithms whilst organisations have a duty to be transparent.

[13] A recital is a text that sets out reasons for the provisions of an act, while avoiding normative language and political argumentation

[14] Sandra Watcher, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469

This combination of right and duty is what we define as explainability, and encompasses the measures described above. We call for the setting up of a new regulatory function to ensure the enforceability of the explainability requirement.

Explainability in Practice

Legislators have adopted different ways to intervene in automated decision making, and to interpret the relevant GDPR provisions. Germany, for example, emphasises the insurance sector to make sure the data subject has the right to not be subject to a decision based solely on automated processing except for when the outcome of the decision is positive. If the outcome is negative, the data subject has the right to contest, express their view and request human intervention. That is an interesting approach, as it provides for safeguards only if the customer's request is not fulfilled.

The UK approach is the opposite: not a sectorial exception with specific safeguards but a generalised exception with specific safeguards. Namely that the controller must notify the data subject that the decision has been made by automated processes and that the data subject may request that the decision is reconsidered. The UK approach is prescriptive and includes detailed procedures governing processes around automated decisions but does not include the reference to meaningful information as per Recital 71 of the GDPR[15]. This is disappointing in our view.[16]

Therefore, we recommend the UK move to:

- Introduce the right to meaningful information regarding the deployment of algorithms

- Introduce a criteria (in accordance with the European Data Protection Board AI manifesto) which means that forms of deep machine learning without any human control shall not be permitted. Human control is a fundamental safeguard in the design and development of algorithms

- Introduce a mandatory Algorithmic Impact Assessment which takes into account related equality and human rights law with particular regard to discrimination.

[15] www.privacy-regulation.eu/en/recital-71-GDPR.htm
[16] https://rm.coe.int/report-on-artificial-intelligence-artificial-intelligence-and-data-pro/16808e6012

- Introduce the requirement to explainability which should apply to any automated decision-making producing significant effect on the society and business in general and an individual in particular.

In particular, it is important to emphasise that where the GDPR quotes 'legal effects' in relation to the scope of the safeguards around automated processing, those should be interpreted in the widest possible way, i.e. 'all algorithms that are likely to affect citizens to a greater extent'.

As we saw in the aftermath of the Cambridge Analytica scandal and in the wider debate around Facebook which followed, algorithms involved in micro-targeting may arguably not produce a legal or similar effect (as the data subject can just click not to see them) yet their pervasiveness and impact is staggering.

Furthermore, there is a risk that the data industry is increasingly exploiting data to assess individual worthiness. Individuals have a right to know whether their credit or another rating is based on their web browsing activities, social media life and online habits.

Algorithmic legibility and transparency are necessary to curb the effect of the proliferation of adtech companies, who all too often have no direct relationship with the consumer and are operating in ways that are opaque, potentially leading to discriminatory and biased outcomes.

Algorithmic Impact Assessments

Recently a number of organisations and governments are moving towards the development of AIAs. The Canadian Government, for example, are developing an online AIA tool that currently consists of 57 questions[17]. Organisations such as AINow have a recommended early stage AIA system[18] and the researchers who developed the predictive policing HART algorithm have developed ALGOCARE[19]. Guidance for public sector procurement and use of algorithms in the UK have been proposed by NESTA[20]. In a 10-point code of practice, NESTA's suggestions include a points scale for algorithmic risk, requirement for auditable sandbox versions, various

[17] https://canada-ca.github.io/digital-playbook-guide-numerique/views-vues/automated-decision-automatise/en/algorithmic-impact-assessment.html

[18] https://ainowinstitute.org/aiareport2018.pdf

[19] https://www.tandfonline.com/doi/full/10.1080/13600834.2018.1458455

[20] https://www.nesta.org.uk/blog/10-principles-for-public-sector-use-of-algorithmic-decision-making/

transparency and accountability measures and an insurance scheme to provide for people negatively impacted by incorrect algorithmic decisions.

As part of the AIA framework all organisations should be ready for ethics audits performed by the new regulatory function supporting the Centre for Data Ethics and the ICO. These audits cannot be a one- off event but must be conducted periodically. An additional tailored AIA should be performed if the algorithm is used in a manner or to an end for which it was not specifically designed and trained.

These two requirements are important in order to avoid "drift" in use that could reduce the accuracy and fairness of the algorithm. Hence, an AIA is not a static document but one that travels with and develops throughout the full life-cycle of the algorithm.

Public sector organisations such as HMRC, the Department of Work and Pensions (DWP) and the Ministry of Defence (MOD), as well as local authorities, need to be subject to a greater degree of scrutiny as their decisions have far reaching consequences, especially if related to welfare provision, crime or autonomous weapons[21].

Therefore, we recommend:

- Approval must be sought from a regulator prior to deployment should the AIA identify an area of risk;
- Risk to individuals or groups should be determined within the UN Universal Declaration of **Human Rights** (UDHR) to balance any variance in cultural norms with regards to fairness and bias;
- Involvement of an automated system in the decision-making process should be clearly highlighted to the user;
- Data subjects must be informed about what would need to change to obtain a different outcome or a different decision;
- Guidance and specific criteria must be developed at sector level as we recognise that a one size fits all approach will be detrimental to innovation.

We recommend that the use of artificial intelligence applications within the Health sector involved in diagnosis, monitoring and treatment of patients be given special consideration. It is already the case that mechanical medical devices are brought to market without the expected

[21] https://www.theguardian.com/world/2018/nov/10/autonomous-drones-that-decide-who-they-kill-britain-funds-research

clinical trials[22], and it is imperative that devices and applications recommended, controlled and monitored by AI systems are not implemented without undergoing the same rigorous regulatory and legal compliance as clinically trialled pharmaceuticals and medical devices. This is particularly the case with applications involving mental health.

Legislation and guidance requiring these devices and applications to submit to the same clinical testing, peer-review and public body standards approval as other medical devices must also be implemented. This recommendation is an extension of the initial code of conduct for data-driven health and care technology[23] recently published by the Department of Health and Social Care.

Especially within the health the social care sectors, we recommend that AI systems are designed with the intention of aiding skilled workers.

AI systems must not be designed to replace skilled employees with lower skilled workers or machines, thereby removing the position entirely.

The PARETS framework

We strongly recommend a global framework which emphasises ethical design throughout all stages of application and programme development and includes a focus on societal and organisational impact.

Our recommended framework, PARETS, is built around agreed terms that have gained traction within the AI Ethics sphere and incorporates the flexibility to dovetail pre-existing legislation, such as GDPR, as well as other forms of impact assessments. This is a stable framework upon which a practical system of regulation and auditing can be built.

It should be emphasised that the PARETS framework and its Algorithmic Impact Assessment, see below, should be initiated at the commencement of any project to guide ethical development, procurement, utilisation and evaluation. Neither tool must be used as a tick box exercise for end stage acceptance.

The PARETS framework is briefly structured as follows:

[22] https://www.theguardian.com/news/audio/2018/nov/27/untested-and-unsafe-the-medical-implants-scandal

[23] https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

- **Privacy and Data Protection:** ensures the appropriate and secure sourcing, handling and use of data on the correct legal basis, including any required communications with the data subject.
- **Accountability:** defines who is responsible at every step of the process, for example the enterprise architect should ensure that the use of an algorithm is appropriate and that data security, data protection and the application are fit for purpose. Definitions of accountability must be the first step of the design and development process so that information is not siloed, laws are obeyed, and platforms and procurement are aligned with company policy and strategy.
- **Responsibility and Fairness**: are values which may need to be embedded in the machine, depending on the degree of machine autonomy. These values must be based on legal requirements (equality and international human rights conventions amongst others) but the process may also require involvement with the data subjects involved. This should include citizens/stakeholder panels at all stages of the development process and the inclusion of an ethics policy.
- **Explainability:** refers to a description and explanation of the key decision-making processes throughout development and implementation of the software. We strongly recommend that AI systems used within the public sector are fully open source. Where this is not possible, descriptions required include those around data sourcing, data cleaning, feature selection, weightings (if known), algorithm type, the extent to which the algorithm is opaque (a black box), performance metrics (e.g. numbers of false negatives/ false positives and cut off points) and the validation and ongoing training of the algorithm. Including a component of explainability will not infringe on the intellectual property (IP) of the developer.
- **Transparency**: requires that the information listed above be easily accessible to any person subject to the algorithm. An independent quality mark is a simple and effective method to alert subjects to the involvement of a certified automated process. A standardised approach has the additional benefit of indicating where to look for further information, such as a user's rights. Performance metrics and accountability must be prominent in any information provided to the individual data subject.
- **Societal and Organisational Impact:** the AIA needs to highlight the impact on the workforce as well as society / community as a whole. For example, it needs to demonstrate how the system augments human capabilities and how the

algorithm does not become policy, thus removing human autonomy in wider decision making.

Embracing innovation ethically: beyond fixing algorithms

AI presents numerous challenges and, as it becomes increasingly used, it blends into our life, changing our interactions in a way that at times is barely noticeable.

Let us consider for example the way that algorithms act as hidden influences in the digital world we live in, and drive what many now call 'online manipulation' which is, essentially, deception and the ability of these systems to alienate us from our own decision- making powers.

The effects of this on our democratic systems are already in full display as the Information Commissioner pointed out in her 'Democracy Disrupted' report[24] where she examined the big data driven technique and their hyper-nudging effect.

Fixing algorithms is an essential component of good governance of AI, and regulation surrounding AIAs must be put in place at the earliest convenience as we have outlined in the previous pages. This is why we are recommending the issuing of certificates of fairness for organisations to adopt and display to increase transparency and reliability.

However, the AI NOW institute is correct in stating that we may be able to fix an algorithm[25] but we also need to ensure that what we are using it for is ethical, and that the whole trajectory of the technological innovation delivers value for all.

This is a debate that people from every background need to get involved in.

It is now time for society to embrace this discussion. If not, the consequences will be catastrophic and could lead to disempowerment and lack of trust in a technology which, if properly used, could bring fantastic benefit to our world and the common good.

In order to ensure AI is for all and for the common good, we need to address two main aspects:

- Mitigating the impact on jobs losses
- Providing skills for everyone

Mitigating the impact on jobs losses

[24] https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2018/07/findings-recommendations-and-actions-from-ico-investigation-into-data-analytics-in-political-campaigns/
[25] https://hub.packtpub.com/ai-now-institute-releases-current-state-of-ai-2018-report/

There is an international consensus that AI will create new jobs. However, it is also agreed that autonomous systems will certainly replace aspects of existing jobs if not, in some cases, replace certain jobs entirely in the relatively near future. According to a recent report by PwC, up to 30% of UK jobs will be at a high risk of being automated by AI and related technologies[26] over the next 20 years. The report estimates that these losses will be balanced by the additional jobs created to support AI implementations; however, this will require significant retraining or hiring of new employees.

The impact of job losses on individuals, communities and society are many and varied; from a diminished sense of identity and the deterioration of whole regions to increasing social turmoil and the often consequential rise in nationalism, both borne of basic insecurity and fractured communities. These influences must not be underestimated, and governments must prepare for them.

Regions more reliant on the types of jobs that automation is likely to replace will suffer disproportionately, whilst certain sectors will see more significant job losses than others. Both circumstances require workers to reconsider their chosen industry, retrain for a new role, reskill regularly to maintain par with technological developments and, in some cases, relocate to another part of the country.

Commercially speaking, there can be no doubt that AI solutions will lead to significant cost savings for many companies, and the concern is that this will come at the expense of high unemployment. Companies must be compelled to, first and foremost, fulfil their responsibilities and duties of care to their existing employees through re-training and reskilling, in order that they can enter new roles created alongside the new AI framework.

We therefore urge the UK government to:

- Perform a **full and proper impact assessment** on which functions, jobs, workforces and industries are likely to be most affected, and tailor an adequate response;
- Introduce a **requirement for companies deploying AI to evaluate their impact on the workforce and society**.
- Introduce a requirement on larger companies to contribute to an **upskilling fund** should they not be able to retain their employees.
- Evaluate the impact on employment at **regional level** and continue to monitor the evolution of the regional economies. New research reveals that the majority of UK-based

[26] https://www.pwc.co.uk/economic-services/ukeo/ukeo-july18-net-impact-ai-uk-jobs.pdf

AI and machine learning vacancies are based outside of the capital, which is a very important phenomenon[27].

Providing skills for everyone

Furthermore, it is necessary to review our education system. Whilst the emphasis on STEM and the Computer Science GCSE and A Levels are a good step forward, that may not be sufficient.

There is a compelling need for ALL students to fully comprehend the digital world to understand, for example, how their data is harvested and used in the world today and the potential impact algorithms may have on their lives in the future. Digital Literacy is now as essential to modern society as being able to read and write.

This applies to adults, too. Finland for example, has decided to train its population in algorithms and they have started with an initial 1%, with the support of private companies and the government[28].

As the Bank of England's Chief Economist Andy Haldane recently said, the Fourth Industrial Revolution will be on a much greater scale than the previous three, and the UK is in dire need of a skills revolution to avoid mass unemployment in the future. There is a compelling need for defining the multidisciplinary skills needed. Within this, we recommend:

- **The setting up of a task force to tackle the low number of women in STEM. We urge the government to bring together the fantastic initiatives in this area, and establish a national strategy involving parents, teachers and children of every age.**

[27] https://www.businesscloud.co.uk/news/majority-of-uk-ai-jobs-now-outside-of-london
[28] https://www.politico.eu/article/finland-one-percent-ai-artificial-intelligence-courses-learning-training/

Conclusion: challenging the unaccountable and the 'inevitable'

New applications of AI emerge on a daily basis and the big data world is in many ways a new phenomenon. Google is only 20 years old.

Most AI driven innovation is for the good but we do not have to be resigned to the negative uses that we are seeing, or to the gendering of virtual personal assistants like Alexa, Siri and Cortana or to the discriminatory algorithms.

Over the years we have passed successful laws defining the governance of many sectors, from food labelling to environmental protections. Now we have the opportunity to decide how we want to govern Artificial Intelligence, how we want to shape it and what we want to use it for. There is nothing inevitable about how we choose to use this disruptive technology. There is no good reason to neglect to prepare and inform the workforce for it. And there is no excuse for failing to set clear rules so that it remains accountable, fosters our civic values and allows humanity to be stronger and better.

Summary of our proposed Recommendations

This first set of recommendations are aimed at national and international policy makers to ensure that AI benefits all and drives us towards a more equitable future.

1. Introduce a **regulatory approach** governing the deployment of AI which mirrors the one deployed in the pharmaceutical or similar sectors

2. Establish an **AI regulatory function** working alongside the ICO and the Centre for Data Ethics, and responsible for:
    a. auditing algorithms deployed by businesses, organisations, and public sector bodies;
    b. acting as a repository of knowledge and setting best practice for organisations;
    c. investigating complaints made by individuals;
    d. issuing notices and fines to organisations in breach of the EU General Data Protection Regulation (GDPR), equality and human rights law;
    e. establishing and issuing wider guidance around the deployment of algorithms;
    f. ensuring the functions of the algorithms are explained and available for public scrutiny

3. Introduce a '**certificate of fairness**' for AI systems that are audited for risks concerning discrimination, unfairness and privacy. The criteria for these certificates should be defined at industry level and mirror the requirements of food labelling regulations.

4. The introduction of **mandatory Algorithmic Impact Assessments (AIAs)** for organisations deploying AI systems where these have a 'significant effect' on individuals. We recommend that '**significant effect'** in relation to algorithms is not limited to automated decisions but encompasses digital advertising and other effects which influence individuals.

5. Introduce a **mandatory requirement for public sector organisations** using AI for decision making, profiling and allocation of public resources to inform citizens at all times that decisions are made by machines, explain how decisions have been reached, and what would need to change for individuals to obtain a different outcome.

6. Introduce a **'*reduced liability*' incentive** for companies that have obtained a Certificate of Fairness. Such companies may also publish a kitemark to showcase their commitment to fairness and equality. We recommend this trade-off as a strategy to ensure growth and development of the AI industry whilst fostering reliability and trust.

7. To compel companies, businesses, organisations and public sector bodies to **bring their workforce with them** as they embark on their innovation journey. This can be achieved by: publishing the impact on their workforce; offering retraining programmes for employees whose jobs are being automated;

8. Where no redeployment is possible, to compel companies to make a contribution towards a **digital skills fund** for employees who are not reemployed.

9. To carry out a **skills audit** to identify the wide range of skills required to embrace the AI revolution. Skills need to go beyond technology and encompass wider needs.

10. To establish an **education and training programme** to meet those needs identified in the skills audit. Education and training must encompass data ethics in order to foster moral responsibility.

Acknowledgements and Contributors

The Women Leading in AI Team

Ivana Bartoletti

Allison Gardner

Reema Patel

Emma Gibson

Liz Stocks

Samara Banno

Sanya Sheikh

Rebecca Geach

Holly Rafique

For more information about Women Leading in AI, or to get involved, please contact us on Twitter, through our website or by email

@WLinAI

www.womenleadinginai.org

admin@womenleadinginai.org

@WLinAI

www.womenleadinginai.org

admin@womenleadinginai.org