

"WOMEN & AI" PLEDGE

FOR AN ACCOUNTABLE
AND GENDER FAIR AI

MACHINE
LEARNING

DATA

NEURAL
NETWORKS

CHATBOT

DEEP LEARNING

ALGORITHM

COGNITIVE
BIAS

PLEDGE VERSION - APRIL 2022

This practical guide is proposed by the Cercle Interelles



FOREWORD

AURÉLIE JEAN



*Doctor in Sciences and
Entrepreneur*

WITH AI COMES RESPONSIBILITY



Artificial intelligence (AI) is everywhere. Embedded in applications, devices, and other technologies, AI assists us in moving, working, communicating, teaching, or even medically treating patients. AI can do amazing things.

That being said, when developing those AI-based technologies, all of us - scientists, engineers, citizens, state and economic leaders, marketers, or any business-related individuals - need to embrace an ethical position in order to produce, regulate, sell, and use inclusive technologies.

AI is not sexist, racist, or discriminatory in general, but we are. Nowadays, Technology by everyone for everyone is the message that should drive every single person, whatever their position on the technological, economical, and societal chess board.

To reach that goal, we need to build a well-defined and motivated governance that includes best practices from a technical, business, and human perspective. The "Women & IA" Pledge from InterElles is a first step towards such a mission and vision.



WHY IS IT IMPORTANT?

Year on year the digital economy has now become firmly established at the center of our lives at work and at home. With the technical revolution that is being brought through Artificial Intelligence (AI)* everything is accelerating. The automated processing of big data*, and the capacity for certain computer programs to learn through data (machine learning*) is generating new use cases in strategic domains such as energy, transportation, healthcare, security and even law and order.

There is often hope that progress in technology will reduce social inequality and discrimination*, and in particular the inequalities between women and men, but unfortunately this hasn't proven to be the case so far for the following reasons:

- At each stage of the implementation of an algorithm*, the underlying motor of AI programmed by humans, seems to integrate and propagate sexist stereotypes. For example, several cases were detected in algorithms used for recruitment process in the past year.
- Diversity* in the AI industry and research is poor with a small and decreasing percentage of woman (22% of research publications in 2017 was written by women- Element AI Lab, 2018). This increases the risk that AI solutions implicitly integrate a masculine view of the world through cognitive bias*.
- Data that is the fuel powering the algorithms, represents actual situations, past and present, that already integrate inequality and gender stereotypes*.



AI is a sounding board for the actual situation in society. It simplifies reality, applying rules and repeatedly taking actions that can amplify inequality and discrimination.

However, we believe it is possible to use this extraordinary technology to leverage a culture shift in society.

For this to happen, we must rally together and take proactive action to force progress in the reduction of inequality.

By the "Women & AI workgroup", Cercle InterElles

WHAT TO DO?

Cercle InterElles is contributing to this movement by launching an action to support and accompany companies who are actively seeking to innovate on this subject through this Women & AI Pledge.

We now know that IT solutions that include AI can reproduce and amplify all forms of discrimination (ethnic origin, social class, sexual orientation...). The core mission of Cercle InterElles is to implement professional equality* between women and men, therefore our pledge focuses on the risk of gender bias* in the AI industry. However, we have chosen an approach based on AI governance (cf charter: the role of the AI ethics committee), that incorporates all the risks of discrimination linked to the processing of large amounts of data.

By including all types of systemic discrimination, our approach can help any company looking for ways to develop AI systems that are globally ethical and accountable. This pledge includes definitions of the key concepts and illustrates the steps necessary for a company to design and develop accountable AI systems that do not discriminate.

Our approach of developing and training employees in a specific "AI culture" provides the means to succeed in the digital transformation of our companies. Through bringing together AI systems and humans we can have an ethical, accountable and inclusive response to the transformative challenges ahead of us.

Making AI a force for social justice is not just a dream for some unwavering optimists. Cercle InterElles is confident that our member companies can act to this end.

Any company, either creating or using AI-based solutions can benefit from this pledge through a set of practical tools enabling them to take action and contribute to the creation of an accountable and gender fair AI.

This pledge is made up of 4 pillars:

COMMIT

By signing the charter for accountable and gender fair AI systems

ASSESS

By using the evaluation grid composed of a set of practical criteria

ACT

By applying a toolbox that is regularly extended through the best practices and innovation within the companies

BECOME AN INDUSTRY MODEL

By sharing the progress made and the best practices

CERCLE INTERELLES



Cercle InterElles is a professional society that groups the diversity networks of 16 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.

Most of the companies who are members of the society are users and producers of AI systems and solutions. As a consequence, Cercle InterElles is able to build on its different company cultures and skills, to better understand women and their specific situation in AI development, both internationally and focusing on the situation in France.

From the start of our research, we were aware of the risks that we are taking by entrusting the development of tomorrow's world to digital systems and without any thought or discussion on the ethics and the governance of such systems. This discussion and any decisions or choices that are made must include a greater involvement of women and men. Diversity must be propagated in all areas of life: education, professions, decision making bodies, economics and politics.

"The Women & AI workgroup" has developed an action plan based on the research carried out by Cercle InterElles. This plan has two main objectives:

- **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.
- 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.

We are convinced that **including more women in the design and governance of digital systems is a must to prevent discrimination and build an ethical and accountable society.**

DIVERSITY NETWORKS OF 16 LARGE COMPANIES IN THE TECHNOLOGY INDUSTRY



CERCLE
INTERELLES

PATRICIA
LECOCQ



Presidente
du Cercle InterElles



For 20 years, the Cercle InterElles has been committed to gender diversity and professional equity in the scientific and technological sectors, with the ambition to create favourable conditions, gender balance and performance.

As a reference business network, we naturally seized this major topic, of responsible Artificial Intelligence. It is a critical, ethical and societal challenge, which aims to take advantage of this technological revolution in order to contribute to the reduction of inequalities and discrimination and more specifically those between women and men.

The Women& AI Pledge for an accountable and gender fair AI is the collective work of the Cercle InterElles companies. This Pledge is a valuable and actionable asset for any company willing to take on this challenge.



DIVERSITY NETWORKS OF 16 LARGE COMPANIES IN THE TECHNOLOGY INDUSTRY



«WOMEN & AI» CHARTER

TO BUILD AN ACCOUNTABLE AND GENDER FAIR AI

FOR COMPANIES

This charter consists of 7 fundamental principles that enable companies to address the risks of discriminatory cognitive bias, during development or while using AI-based solutions or devices.

7 FUNDAMENTAL PRINCIPLES



AI ETHICS COMMITTEE AND GOVERNANCE
PAGE 6



COMPLIANCE BY DESIGN
PAGE 6



DATA SELECTION AND PROCESSING
PAGE 7



THE ETHICS OF ALGORITHMS
PAGE 7



EVALUATION AND MONITORING
PAGE 8



AI TEAM DIVERSITY
PAGE 8



AWARENESS AND ACCOUNTABILITY
PAGE 8

1 AI ETHICS COMMITTEE AND GOVERNANCE

A non-discriminatory process should be part of a broader ethical business approach that goes beyond the legal aspects. The legitimacy of initiatives to create non-discriminatory AI systems should rely on the firm commitment of management.

The company should implement a governance process that ensures an operational action plan and direct access to company Executives for AI and diversity questions.

More specifically, a multidisciplinary AI Ethics Committee, that reports directly to the Executive Management, and where the scope of responsibility extends to the whole company, ensures the detection of gender bias, and rapid corrective action in all AI systems used or produced.

The mission of the AI Ethics Committee is explicitly specified, documented and distributed across the company, specifically the handling of gender bias. The committee has the means to perform their assigned mission efficiently.

IDENTIFIED RISKS

- Lack of support from senior management
- Lack of adoption of the pledge by employees and management
- Failure to understand gender bias in the use of AI-based tools

EXTRACT FROM THE TOOLBOX

- Executive awareness raising and Interview guide
- Appointment of a delegate from the general management within the AI Committee
- Mission Statement and role of the AI Ethics Committee
- Job description of in-house AI champion
- Activity reports of the AI Committee
- Monitoring dashboard of the AI Committee recommendations

2 COMPLIANCE BY DESIGN

Companies are not always aware of the impact of AI solutions on gender diversity*.

From the design of an AI product or project to the moment it is delivered into production, the teams involved make sure that the principles of non-discrimination of gender are integrated at each stage.

Futhermore, the compliance of AI solutions produced or used is analysed by a relevant internal authority, the AI Ethics Committee, taking into account applicable laws on equality in digital economy.

Companies should integrate in its values, agreements and internal procedures an AI and gender bias component.

IDENTIFIED RISKS

- Risk of legal non-compliance of the solution
- Additional costs to comply after AI project ends

EXTRACT FROM THE TOOLBOX

- Applicable equality & digital Laws glossary
- Standards on bias management in algorithms
- Mapping model of all AI-based projects and software developed and/or in use
- Evaluation of the impacts of AI solutions on humans and gender, and risk analysis

3

DATA SELECTION AND PROCESSING

Being aware of gender bias in an AI project starts with data selection.

A team of industry experts and developers analyses and classifies data and identifies any imbalance that exists in the data set.

If necessary, biased data are corrected or deleted.

IDENTIFIED RISKS

- **Risk to replicate and amplify existing gender bias through the use of already biased datasets**

EXTRACT FROM THE TOOLBOX

- **Data mining tools list**
(e.g. software)
- **Tools & techniques list to unbiased dataset**
- **Elimination or neutralisation of the "gender" variable** *(direct and induced)* **if it is not needed**

4

THE ETHICS OF ALGORITHMS

A company that develops algorithms for AI projects should use a tool that can detect gender bias. The results are reviewed by an in-house champion and/or the AI Ethics Committee.

A document that guarantees the transparency, traceability, and the explainability of the algorithms used and the results produced, should be kept up to date.

Companies that are users of AI-based systems should engage with suppliers who develop AI solutions that do not propagate gender bias in their solutions.

IDENTIFIED RISKS

- **Algorithms reproduce and amplify gender bias**

EXTRACT FROM THE TOOLBOX

- **Methods for bias management**
- **Document model updated throughout the product life time**
- **Implementation of a review and validation of the analyses by the Champion/AI Committee.**
- **Tools list of bias management**
(measurement tool of equity)

5

EVALUATION AND MONITORING

The company should establish a procedure for reporting and correcting deviations and possible discriminations perpetrated by AI solutions throughout their lifecycles (collecting data, cleaning data, training Machine Learning models, and so on). This procedure has to be compliant with the regulatory context.

The AI Committee supervise the detection and correction of deviations and discriminations, handled by AI development team.

IDENTIFIED RISKS

- Without a methodology of control and evaluation before and after the fact, the absence of deviations cannot be guaranteed over time

EXTRACT FROM THE TOOLBOX

- Tools for reporting deviation and discrimination.
- List of relevant disciplines and persons responsible for control points in the company producing or using AI

6

AI TEAM DIVERSITY

The company should set global or team-specific objectives that will result in the diversification of profiles in AI teams. In line with global policy, gender diversity should be an objective in the recruitment, retention, and promotion of employees.

AI teams seek to diversify talent by hiring employees possessing hard and soft skills, integrating different profiles in teams, making them more mixed and thus contributing to the elimination of gender discrimination.

To increase the proportion of women in AI teams, the company agrees to develop professional training programs. These programs focus on training or re-training women for jobs in AI.

IDENTIFIED RISKS

- The low percentage of women in AI teams increases the risk of producing discriminating AI

EXTRACT FROM THE TOOLBOX

- Quantified indicators of team gender balance in managers' objectives
- Dashboard for monitoring the gender ratio
- Update of the charter against sexism* to mention AI
- Programs of reconversion towards the new AI professions
- Sponsorship and Mentorship programs for Women Teams members
- Valuation of role models

7

AWARENESS AND ACCOUNTABILITY

The company agrees to raise awareness among employees about issues related to gender bias, especially among those who work in the field of AI.

For example, the company can raise awareness about gender bias by developing a range of audience-specific content, from simple communications to training adapted to employee roles.

The company is encouraged to raise awareness within its wider ecosystem, including schools, universities, and the general public.

IDENTIFIED RISKS

- **Lack of knowledge and awareness of gender bias and its reproduction and amplification via AI can lead to an acceleration of gender bias**

EXTRACT FROM THE TOOLBOX

- **Actions for raising awareness and communicating on the subject to employees.**
- **Gender bias training modules for AI teams**
(potentially including the signing of an oath or code of ethics in conclusion)
- **Bias detection tool ("bias hunter") applicable to company's software programs**

FOR AN ACCOUNTABLE
AND GENDER FAIR AI

THIS PLEDGE IS MADE UP OF 4 PILLARS

COMMIT

By signing the charter for accountable and gender fair AI systems



As for the first pillar, we invite companies to sign this charter to indicate their will to take action for an accountable and gender fair AI by working on the 7 fundamental principles of the pledge.

The signature gives access to the tools, including the evaluation grid and the toolbox.

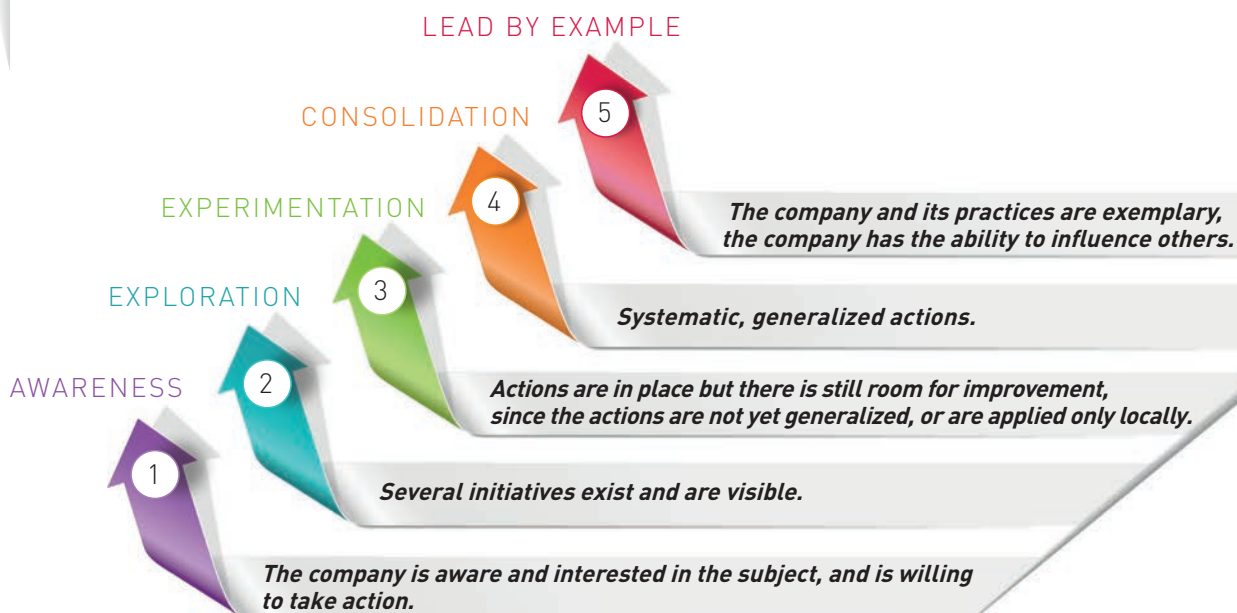
ASSESS

By using the evaluation grid composed of a set of practical criteria

The evaluation grid is part of a global approach that makes it possible for every company not only to measure its maturity level, but also to set a course and strive towards exemplarity. It is an essential tool both at the beginning of the process and to evaluate progress in the long term.

This grid is intended for internal usage within the company but can also be a solid foundation to communicate on the efforts made, leading to exemplarity.

The grid takes up the 7 fundamental principles of the InterElles «Women & AI» Charter to build an accountable and gender fair AI, and describes the different stages/levels of maturity (milestones) from 1 to 5 :



Example for the « governance » principle:

AI Ethics Committee and Governance	0	Remarks
Milestone 1: Our company is considering the creation of an ethics committee	<input type="checkbox"/>	
Milestone 2: There is an ethics committee in our company, but it does not cover AI matter	<input type="checkbox"/>	
Milestone 3: There is an ethics committee in our company which covers AI matter	<input type="checkbox"/>	
Milestone 4: There is a multi-disciplinary AI ethics committee which covers gender issues, with a dedicated AI champion. This AI committee has a strong support of the company's direction (i.e. an Executive Sponsor that is both a member of the Executive Committee and the AI ethics committee, or an AI ethics committee that reports to the Executive Committee (as does the internal audit))	<input type="checkbox"/>	
Milestone 5: The missions of the AI ethics committee are clear, documented and disseminated, especially regarding gender bias treatment. The AI committee possesses the means to carry out the assigned missions efficiently (consulting, project impact, veto right, local correspondents...)	<input type="checkbox"/>	



Evaluation Grid © 2021 by Cercle InterElles is licensed under Attribution-NonCommercial-NoDerivatives 4.0 International. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>



This document allows companies to get a maturity assessment, orients them towards the adequate tools which can help them get to the next level (based on our toolkit), and helps measure progress.

ACT

By applying a toolbox that is regularly extended through the best practices and innovation within the companies

Cercle InterElles supports in different ways companies to take action. A toolbox gathers a collection of techniques, methods, advices to identify and mitigate gender bias. These tools are grouped according to the 7 fundamental principles and available online in French and in English on the platform **SOFIA (Solutions et Outils Femmes & IA)**. They are either created by the Cercle InterElles and partners or selected among the tools available in our ecosystem. Signatories companies are also invited to join Women & AI Action Circle to share experiences and best practices on the implementation of the pledge.



BECOME AN INDUSTRY MODEL

By sharing the progress made and the best practices

The last pillar of our pledge contributes to a productive exchange between companies to accelerate the adoption and the creation of an ethical AI. Using the evaluation grid at different moments in time allows companies to see their advancement through the milestones and shows the global improvement in their maturity level. By sharing the progress made and the best practices, companies lead by example, facilitate the outreach of the pledge and enhance it. In this way, they strengthen their role in the journey to an accountable and gender-fair society.

A GLOSSARY OF TERMS FOR WORKPLACE EQUALITY BETWEEN MEN AND WOMEN

by Cercle InterElles

AFFIRMATIVE ACTION

The policies and actions taken to increase representation of women as a remedy to the effects of long-standing discrimination against them.

COGNITIVE BIAS

A personal interpretation of events or a situation that is based on anecdotal evidence rather than facts or research. Cognitive bias is a mental shortcut that our brain uses to quickly understand a situation, but this shortcut can lead to poor decision making.

DISCRIMINATION

The act of unfairly treating one or more people differently from other people. More precisely, discrimination means that distinctions are made between one social group and others based on extrinsic characteristics (wealth, education, place of residence) that results in a treatment that is often negative.

DIVERSITY

The state of being varied, diverse. A group of people can be said to be “diverse” when it contains many distinct profiles in terms of geographic origins, socio-professional categories, cultures, ages, genders, education levels, sexual orientations, physical appearances, and so on.

EQUALITY

A situation in which the treatment of men and women is identical, which results from collective, non-discriminatory processes.

FEMINISM

The movement that promotes social and political equality between women and men.

GENDER BIAS

A cognitive bias that leads to preferential or discriminatory treatment based on gender.

GENDER DIVERSITY

The presence of men and women in a group, without necessarily being in equal numbers.

GLASS CEILING

A metaphor that refers to an unseen barrier that prevents women from being promoted to managerial- and executive-level positions, regardless of their qualifications.

PARITY

A metric that measures the gaps between men and women in areas such as pay and recruitment.

QUOTA

A fixed minimum percentage of women in a group needed to meet the requirements of affirmative action.

SEXISM

The discrimination or prejudice manifest in words and behaviors that destabilise, denigrate, and delegitimise women in the workforce, in society, and in politics.

STEREOTYPE

A prejudice or mistaken idea about a group of people that leads to a fixed, oversimplified opinion about them.

GLOSSARY

A GLOSSARY OF TERMS FOR ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

ALGORITHM

A set of operative rules or instructions that resolve a problem or perform a task.

ARTIFICIAL INTELLIGENCE (AI)

A system that can analyse its environment and reproduce certain human behaviors to achieve a goal with some degree of autonomy. Research in AI is a cross-disciplinary approach based in cognitive and computer sciences, mathematics, linguistics, psychology, and other disciplines.

DATA

A collection of information that is represented numerically. Large amounts of data ("big data") are often collected and processed to train machine learning models.

DEEP LEARNING

A sub-field of machine learning that produces algorithms that are inspired by the human brain. Deep learning models apply artificial neural networks to various fields to teach systems to recognize a face, detect objects, help doctors diagnose patients more accurately, translate texts, and more.

MACHINE LEARNING (ML)

A sub-field of artificial intelligence that builds computational models that are endowed with the ability to «learn» from historical data without being explicitly programmed.

RAW DATA

A set of data collected from a source. Typically, raw data has not undergone data cleaning, which is a process by which inaccurate or corrupt data is removed from data records.

THANKS TO CERCLE INTERELLES "WOMEN & AI" WORKGROUP MEMBERS

CEA Anne-Marie Jonquière - Ilona Leroy - **Dassault Systèmes** - Meriem Ben Abdallah Bernstein - Laurence Deparday - **EDF** Laurene Vaugrante - **Engie** Philippe Calvez - **GE** Melodie Sperandio - Marianne Gaulhet - **IBM** Sonia Abecassis-Le Lan - Sylviane Bordas - Hélène Fourot Quillaud - Muriel Gardes - **Intel** Florence Peutin - **Lenovo** Christine Dehnel - Sandra Lautric - Marine Rabeyrin - Pernelle Steiner - **NGE** - Pierre Gardeux - Frederique Gouy - Manon Knapp-Ziller - **Orange** Emmanuelle Bravin - Julie Marques - Emilie Sirvent-Hien - **Orano** Kenza Hammou - Tea Rukavina - **Qualcomm** Estelle Jezequel - Xuecan Yang - **Salesforce** Rebecca Bory - **SAP** Laure Le Bars - Stéphanie Le Bris - Assia Mouloudi - Veronique Venditti - **Schlumberger** Jean-Michel Diosse - Stephanie Robisson Montillet - **SNCF Tech** Christelle Wozniak - **Associated members** Sandrine Bortone - Chantho Creze - Bernadette Andrietti



"WOMEN & AI" PLEDGE

FOR AN ACCOUNTABLE
AND GENDER FAIR AI

OUR PARTNERS



THEIR WORKS AND THE INTERACTIONS WITH THEM INSPIRE US

AFNUM • Arborus • Women in AI • Numeum • Social Builder • Politiqu'elles •
Conseil supérieur à l'égalité professionnelle • Ensemble contre le sexisme • Institut Montaigne

TO CONTACT US

www.interelles.com/femmes-et-ia • femmes-et-ia@interelles.com



DIVERSITY NETWORKS OF 16 LARGE COMPANIES IN THE TECHNOLOGY INDUSTRY

